

Creating synthetic distribution networks

August 27, 2019

I Goal

In this work, we try to generate the distribution network for a given county from the available open source data (substation location information, NAVTEQ transportation network data and synthetic population data). The distribution system is assumed to have a radial/tree structure and closely replicate a real distribution network. The goal of this work is to generate the primary and secondary distribution network to connect the substations to activity locations.

II Data set

The NAVTEQ transportation network data is available in the form of an edge-list or a list of links. Each link has associated an integer level from the set $\{1, 2, 3, 4, 5\}$ that describes the link type (e.g., a level-1 link could correspond to an Interstate road segment, while a level-5 link could correspond to a residential road). All links with level ≤ 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. 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 $\mathcal{R} = (\mathcal{V}, \mathcal{L})$, where \mathcal{V} and \mathcal{L} are respectively the sets of nodes and links of the network.
- (ii) The set of substations $\mathcal{S} = \{s_1, s_2, \dots, s_M\}$, where the county consists of M substations.
- (iii) The set of activity locations $\mathcal{H} = \{h_1, h_2, \dots, h_N\}$, where the county consists of N activity locations.

Each node in the road network graph, the substations and activity locations has an associated spatial embedding in the form of longitude and latitude.

III Approach

We want a mapping between the substations and activity locations. This can be achieved by mapping the substations to nodes of the road network and assigning a map between activity locations 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.

IV Algorithm

A Constructing the mapping

- (a) Assign a road node proxy to each substation using the map $g : \mathcal{S} \rightarrow \mathcal{V}$. In this case, the substation $s \in \mathcal{S}$ is assigned to its nearest element in \mathcal{V} . We have

$$\mathcal{N} = \{g(s), \text{ for all } s \in \mathcal{S}\} \quad (1)$$

Special Case: $|g^{-1}(n)| > 1$ for some $n \in \mathcal{N}$ where $|\cdot|$ is cardinality of the set “ \cdot ”. This implies multiple substations are located near to each other and therefore have the same road network node n mapped to them. This problem can be solved in two ways:

- (i) combining the substations to a single substation and
 - (ii) select a different proxy road node for each substation.
- (b) Find connected components $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_L$ in the road network graph. Let the i^{th} connected component be denoted by $\mathcal{C}_i = (\mathcal{V}_i, \mathcal{L}_i)$, where \mathcal{V}_i and \mathcal{L}_i are respectively the sets of nodes and links of the connected component. We remove the connected components \mathcal{C}_i which do not have any common nodes with the set \mathcal{N} .

$$\mathcal{R} = \mathcal{R}[\mathcal{V} \setminus \mathcal{V}_i], \text{ if } \mathcal{V}_i \cap \mathcal{N} = \emptyset, \text{ for } i = 1, 2, \dots, L \quad (2)$$

- (c) Each link $e \in \mathcal{L}$ has a weight associated with it and is given by $l(e)$. Let $d : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^+$ be the graph distance in \mathcal{R} using $l(e)$ as the weight of links. Since $\mathcal{N} \subset \mathcal{V}$, let $d(v, n)$ be the graph distance between the road node $v \in \mathcal{V}$ and road node proxy for a substation $n \in \mathcal{N}$. We define a map $q : \mathcal{V} \rightarrow \mathcal{N}$ which assigns to every road node $v \in \mathcal{V}$ its nearest element in \mathcal{N} .

$$q(v) = \arg \min_{n \in \mathcal{N}} d(v, n) \quad (3)$$

The inverse map $q^{-1}(n)$ is the set of nodes $v \in \mathcal{V}$ which are nearest to the road node proxy n . We identify a mapping between the substations and road network. Let \mathcal{V}_s denotes the road network nodes mapped to the substation s and is given by

$$\mathcal{V}_s = \{q^{-1}(g(s)), \text{ for } s \in \mathcal{S}\} \quad (4)$$

- (d) For each activity location $h \in \mathcal{H}$, there exists a mapping $f : \mathcal{H} \rightarrow \mathcal{L}$ which assigns its nearest road link $e \in \mathcal{L}$. Let \mathcal{H}_e denote the set of activity locations mapped to the road network link $e \in \mathcal{L}$.

$$\mathcal{H}_e = \{f^{-1}(e), \text{ for } e \in \mathcal{L}\} \quad (5)$$

- (e) Let $E(v)$ denote the straight line equation represented by the edge e such that all coordinates along edge e satisfies $E(v) = 0$. Each activity location $h \in \mathcal{H}_e$ is included in either set $\mathcal{H}_{e,A}$ or $\mathcal{H}_{e,B}$ depending on whether $E(h)$ is positive or negative respectively.

$$\begin{aligned} \mathcal{H}_{e,A} &= \{h, \text{ if } E(h) \geq 0, \text{ for } h \in \mathcal{H}_e\} \\ \mathcal{H}_{e,B} &= \{h, \text{ if } E(h) < 0, \text{ for } h \in \mathcal{H}_e\} \end{aligned} \quad (6)$$

- (f) Let an edge $e \in \mathcal{L}$ be given by $e(v_i, v_j)$ where v_i and v_j are the nodes at two ends of edge e . Let $c : \mathcal{V} \times \mathcal{H} \rightarrow \mathbb{R}^+$ be the Euclidean distance between a road node and an activity location. Therefore, $c(v, h)$ is the Euclidean distance between road node v and activity location h . For each activity location $h \in \mathcal{H}_e$, we assign a mapping $p : \mathcal{H} \rightarrow \mathcal{R}$ such that

$$p(h) = \begin{cases} v_i, & \text{if } c(v_i, h) < c(v_j, h) \\ v_j, & \text{otherwise} \end{cases} \quad (7)$$

Let \mathcal{H}_v denote the set of activity locations mapped to the road network node $v \in \mathcal{R}$.

$$\mathcal{H}_v = \{p^{-1}(v), \text{ for } v \in \mathcal{R}\} \quad (8)$$

B Constructing the links

- (a) The primary distribution network is generated following the road network as a proxy. For the substation $s \in \mathcal{S}$, let \mathcal{R}_s denote the induced subgraph consisting of the road nodes mapped to s . Therefore,

$$\mathcal{R}_s = \mathcal{R}[\mathcal{V}_s] \quad (9)$$

and \mathcal{T}_s be the tree rooted at $g(s)$ and constructed by a depth first search (DFS) from $g(s)$. The primary distribution network, therefore, is a forest and is given by

$$\mathcal{F} = \{\mathcal{T}_s | s \in \mathcal{S}\} \quad (10)$$

- (b) For a node $v \in \mathcal{R}$ in the road network, first we identify the set of edges $E_v \subseteq \mathcal{L}$ connected to it. Let $E_v = \{e_1, e_2, \dots, e_m\}$. Thereafter, we identify the sets of activity locations as follows

$$\mathcal{A}_v = \left\{ \mathcal{H}_v \cap \mathcal{H}_{e_1, A}, \mathcal{H}_v \cap \mathcal{H}_{e_1, B}, \mathcal{H}_v \cap \mathcal{H}_{e_2, A}, \mathcal{H}_v \cap \mathcal{H}_{e_2, B}, \dots, \mathcal{H}_v \cap \mathcal{H}_{e_m, A}, \mathcal{H}_v \cap \mathcal{H}_{e_m, B} \right\} \quad (11)$$

We connect nodes in subsets of \mathcal{A}_v in a chain fashion to obtain the secondary network.