Graph Metrics Definitions

Let $G = (V, E)$ be an undirected graph representing a street network, where V is the set of nodes and E the set of edges. Each edge $e \in E$ has a positive length $\ell(e) > 0$. Let $n = |V|$ and $m = |E|$.

For shortest-path based metrics, let $d(u, v)$ denote the weighted shortest-path distance between nodes u and v using edge lengths $\ell(\cdot)$. For metrics that require connectivity, computations are typically performed on the largest connected component of G.

0.1 Size and scale

- **Number of nodes (n_nodes):** $n = |V|$.
- **Number of edges (n_edges):** $m = |E|$.
- **Total length (total_length):**
$$L_{\text{total}} = \sum_{e \in E} \ell(e).$$

0.2 Degree-based connectivity

Let $\deg(v)$ be the degree of node $v \in V$.

- **Average degree (avg_degree):**
$$\bar{k} = \frac{1}{n} \sum_{v \in V} \deg(v).$$
- **Number of degree-1 nodes (num_deg1):**
$$n_{\deg=1} = \left| \{ v \in V : \deg(v) = 1 \} \right|.$$
- **Proportion of degree-1 nodes (prop_deg1):**
$$p_{\deg=1} = \frac{n_{\deg=1}}{n}.$$
- **Number of nodes with degree ≥ 3 (num_deg_ge3):**
$$n_{\deg \geq 3} = \left| \{ v \in V : \deg(v) \geq 3 \} \right|.$$
- **Proportion of nodes with degree ≥ 3 (prop_deg_ge3):**
$$p_{\deg \geq 3} = \frac{n_{\deg \geq 3}}{n}.$$
- **Degree entropy (degree_entropy):** Let $p(k)$ be the empirical probability that a node has degree k.
$$H_{\deg} = -\sum_{k} p(k) \log p(k).$$

0.3 Density and mesh measures

- **Graph density (density):**

$$\rho = \frac{2m}{n(n-1)}.$$

This measures edge density relative to a complete graph.

- **Mesh density (mesh_density):**

$$\rho_{\text{mesh}} = \frac{m-n+1}{2n-5} \quad (n \geq 3).$$

This measures cyclic redundancy normalized by the planar upper bound; values near 0 indicate tree-like structure and larger values indicate more meshed networks.

0.4 Shortest-path extension metrics

Define the eccentricity of node u as $\epsilon(u) = \max_{v \in V} d(u, v)$ (within the component considered).

- **Average shortest path length (avg_shortest_path_length):**

$$\overline{d} = \frac{1}{n(n-1)} \sum_{u \neq v} d(u, v).$$

- **Radius (radius):**

$$r = \min_{u \in V} \epsilon(u).$$

- **Diameter (diameter):**

$$D = \max_{u \in V} \epsilon(u).$$

- **Average eccentricity (avg_eccentricity):**

$$\overline{\epsilon} = \frac{1}{n} \sum_{u \in V} \epsilon(u).$$

0.5 Centrality metrics

- **Closeness centrality (CC):**

$$CC(u) = \frac{1}{\sum_{v \neq u} d(u, v)}.$$

Average node closeness (avg_node_closeness) is $\frac{1}{n} \sum_{u \in V} CC(u)$.

- **Betweenness centrality (BC):** Let σ_{st} be the number of shortest paths between s and t, and $\sigma_{st}(u)$ the number of those paths that pass through u.

$$BC(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}.$$

(Weighted shortest paths use $\ell(\cdot)$.)

- **Maximum node betweenness (max_node_betweenness):**

$$\max_{u \in V} BC(u).$$

- **Average node betweenness (avg_node_betweenness):**

$$\overline{BC} = \frac{1}{n} \sum_{u \in V} BC(u).$$

- **Coefficient of variation of node betweenness (cv_node_betweenness):**

$$CV_{BC} = \frac{\mathrm{std}(\{BC(u)\})}{\mathrm{mean}(\{BC(u)\})},$$

when the mean is nonzero.

0.6 Branching, redundancy, and spanning tree

- **Branching index (branching_index):**

$$B = \frac{n_{\deg=1}}{n_{\deg \geq 3}},$$

with a large sentinel value used when $n_{\deg \geq 3} = 0$.

- **Cyclomatic number (cyclomatic_number):**

$$\mu = m - n + c,$$

where c is the number of connected components. It equals the number of "extra" edges beyond a spanning forest, i.e., the number of independent cycles.

- **Minimum spanning tree length (mst_length):** Let T be a minimum spanning tree of the (connected) component considered, using $\ell(\cdot)$ as weights:

$$L_{\mathrm{MST}} = \sum_{e \in T} \ell(e).$$

- **MST ratio (mst_ratio):**

$$R_{\mathrm{MST}} = \frac{L_{\mathrm{total}}}{L_{\mathrm{MST}}},$$

when $L_{\mathrm{MST}} > 0$.

0.7 Edge length statistics

Let $\mathcal{L} = \{\ell(e) : e \in E\}$ denote the multiset of edge lengths.

- **Average edge length (avg_edge_length):** $\mathrm{mean}(\mathcal{L})$.
- **Median edge length (median_edge_length):** $\mathrm{median}(\mathcal{L})$.
- **Minimum edge length (min_edge_length):** $\min(\mathcal{L})$.
- **Maximum edge length (max_edge_length):** $\max(\mathcal{L})$.
- **Standard deviation of edge length (std_edge_length):** $\mathrm{std}(\mathcal{L})$.
- **Coefficient of variation of edge length (cv_edge_length):**

$$CV_{\ell} = \frac{\mathrm{std}(\mathcal{L})}{\mathrm{mean}(\mathcal{L})},$$

when the mean is nonzero.

- **25th percentile of edge length (edge_length_p25):** $\mathrm{quantile}_{0.25}(\mathcal{L})$.
- **75th percentile of edge length (edge_length_p75):** $\mathrm{quantile}_{0.75}(\mathcal{L})$.

0.8 Community structure

- **Modularity (modularity):** Q, the modularity value of a community partition (e.g., from greedy modularity maximization) comparing within-community edge density against a null model.
- **Number of communities (n_communities):** the number of detected communities in the chosen partition.

0.9 Straightness and tortuosity

Let $d_G(u, v) = d(u, v)$ be the network shortest-path distance. If node coordinates are available, let $d_E(u, v)$ be the Euclidean distance between u and v.

- **Average straightness (avg_straightness_length):**

$$\overline{S} = \frac{1}{|P|} \sum_{(u,v)\in P} \frac{d_E(u, v)}{d_G(u, v)},$$

  measuring how direct routes are compared to straight-line distances.

- **Average tortuosity (avg_tortuosity_length):**

$$\overline{T} = \frac{1}{|P|} \sum_{(u,v)\in P} \frac{d_G(u, v)}{d_E(u, v)},$$

  measuring detour relative to straight-line distance.