Structural Validation Of Synthetic Power Distribution Networks Using The Multiscale Flat Norm*

Rounak Meyur[1], Kostiantyn Lyman[2], Bala Krishnamoorthy[2], and Mahantesh Halappanavar[1]

[1] Pacific Northwest National Lab, USA {rounak.meyur,hala}@pnnl.gov
[2] Department of Mathematics and Statistics, Washington State University, USA {kostiantyn.lyman,kbala}@wsu.edu

Abstract. We study the problem of comparing a pair of geometric networks that may not be similarly defined, i.e., when they do not have one-to-one correspondences between their nodes and edges. Our motivating application is to compare power distribution networks of a region. Due to the lack of openly available power network datasets, researchers synthesize realistic networks resembling their actual counterparts. But the synthetic digital twins may vary significantly from one another and from actual networks due to varying underlying assumptions and approaches. Hence the user wants to evaluate the quality of networks in terms of their structural similarity to actual power networks. But the lack of correspondence between the networks renders most standard approaches, e.g., subgraph isomorphism and edit distance, unsuitable.
We propose an approach based on the *multiscale flat norm*, a notion of distance between objects defined in the field of geometric measure theory, to compute the distance between a pair of planar geometric networks. Using a triangulation of the domain containing the input networks, the flat norm distance between two networks at a given scale can be computed by solving a linear program. In addition, this computation automatically identifies the 2D regions (patches) that capture where the two networks are different. We demonstrate our approach on a set of actual power networks from a county in the USA. Our approach can be extended to validate synthetic networks created for multiple infrastructures such as transportation, communication, water, and gas networks.

Keywords: synthetic networks · multiscale flat norm · network validation

1 Introduction

The power grid is the most vital infrastructure that provides crucial support for the delivery of basic services to most segments of society. Once considered

* The research is supported in part by the U.S. DOE Exascale Computing Project's (ECP) (17-SC-20-SC) ExaGraph codesign center and Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL).

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

a passive entity in power grid planning and operation, the power distribution system poses significant challenges in the present day. The increased adoption of rooftop solar photovoltaics (PVs) and electric vehicles (EVs) augmented with residential charging units has altered the energy consumption profile of an average consumer. Access to extensive datasets pertaining to power distribution networks and residential consumer demand is vital for public policy researchers and power system engineers alike. However, the proprietary nature of power distribution system data hinders their public availability. This has led researchers to develop frameworks that synthesize realistic datasets pertaining to the power distribution system [4,13,16,17,27,28]. These frameworks create digital replicates similar to the actual power distribution networks in terms of their structure and function. Hence the created networks can be used as *digital duplicates* in simulation studies of policies and methods before implementation in real systems.

The algorithms associated with these frameworks vary widely—ranging from first principles based approaches [17,27] to learning statistical distributions of network attributes [28] to using deep learning models such as generative adversarial neural networks [14]. Validating the synthetic power distribution networks with respect to their physical counterpart is vital for assessing the suitability of their use as effective digital duplicates. Since the underlying assumptions and algorithms of each framework are distinct from each other, some of them may excel compared to others in reproducing digital replicates with better precision for selective regions. To this end, we require well-defined metrics to rank the frameworks and judge their strengths and weaknesses in generating digital duplicates of power distribution networks for a particular geographic region.

The literature pertaining to frameworks for synthetic distribution network creation include certain validation results that compare the generated networks to the actual counterpart [4,12,28]. But the validation results are mostly limited to comparing the statistical network attributes such as degree and hop distributions and power engineering operational attributes such as node voltages and edge power flows. Since power distribution networks represent real physical systems, the created digital replicates have associated geographic embedding. Therefore, a structural comparison of synthetic network graphs to their actual counterpart becomes pertinent for power distribution networks with geographic embedding. Consider an example where a digital twin is used to analyze impact of a weather event [26]. Severe weather events such as hurricanes, earthquakes and wild fires occur in specific geographic trajectories, affecting only portions of societal infrastructures. In order to correctly identify them during simulations, the digital twin should structurally resemble the actual infrastructure.

Problem Statement. In recent years, the problem of evaluating quality of reconstructed networks has been studied for street maps. Certain metrics were defined to compare outputs of frameworks that use GPS trajectory data to reconstruct street map graphs [1,2]. The abstract problem can be stated as follows: *compute the similarity between a given pair of embedded planar graphs*. This is similar to the well known subgraph isomorphism problem [7] wherein we look for isomorphic subgraphs in a pair of given graphs. A major precursor to this

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

problem is that we require a one-to-one mapping between nodes and edges of the two graphs. While such mappings are well-defined for street networks, the same cannot be inferred for power distribution networks. Since power network datasets are proprietary, the node and edge labels are redacted from the network before it is shared. The actual network is obtained as a set of "drawings" with associated geographic embeddings. Each drawing can be considered as a collection of line segments termed a *geometry*. Hence the problem of comparing a set of power distribution networks with geographic embedding can be stated as the following: *compute the similarity between a given pair of geometries lying on a geographic plane*.

Our Contributions. We propose a new distance measure to compare a pair of geometries using the *flat norm*, a notion of distance between generalized objects studied in geometric measure theory [9,20]. This distance combines the difference in length of the geometries with the area of the patches contained between them. The area of patches in between the pair of geometries accounts for the lateral displacement between them. We employ a *multiscale* version of the flat norm [21] that uses a scale parameter $\lambda \geq 0$ to combine the length and area components (for the sake of brevity, we refer to the multiscale flat norm simply as the flat norm). Intuitively, a smaller value of λ captures larger patches of area between the geometries while a large value of λ captures more of the (differences in) lengths of the geometries. Computing the flat norm over a range of values of λ allows us to compare the geometries at multiple scales. For computation, we use a discretized version of the flat norm defined on simplicial complexes [10], which are triangulations in our case. A lack of one-to-one correspondence between edges and nodes in the pair of networks prevents us from performing one-to-one comparison of edges. Instead we can sample random regions in the area of interest and compare the pair of geometries within each region. For performing such local comparisons, we define a *normalized flat norm* where we normalize the flat norm distance between the parts of the two geometries by the sum of the lengths of the two parts in the region. Such comparison enables us to characterize the quality of the digital duplicate for the sampled region. Further, such comparisons over a sequence of sampled regions allows us to characterize the suitability of using the entire synthetic network as a duplicate of the actual network.

Our main **contributions** are the following: (i) we propose a distance measure for comparing a pair of geometries embedded in the same plane using the flat norm that accounts for deviation in length and lateral displacement between the geometries; and (ii) we perform a region-based characterization of synthetic networks by sampling random regions and comparing the pair of geometries contained within the sampled region. The proposed distance allows us to perform a global as well as local comparison between a pair of network geometries.

Related Work. Several well defined graph structure comparison metrics such as subgraph isomorphism and edit distance have been proposed in the literature along with algorithms to compute them efficiently. Tantardini et al. [30] compare graph network structures for the entire graph (global comparison) as well

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

as for small portions of the graph known as motifs (local comparison). Other researchers have proposed methodologies to identify structural similarities in embedded graphs [3,22]. However, all these methods depend on one-to-one correspondence of graph nodes and edges rather than considering the node and edge geometries of the graphs. The edit distance, i.e., the minimum number of edit operations to transform one network to the other, has been widely used to compare networks having structural properties [24,25,31]. Riba et al. [25] used the Hausdorff distance between nodes in the network to compare network geometries. Majhi et al. [15] modified the traditional definition of graph edit distance to be applicable in the context of "geometric graphs" embedded in a Euclidean space. Along with the usual insertion and deletion operations, the authors have proposed a cost for translation in computing the geometric edit distance between the graphs. However, the authors also show that the problem of computing this metric is $\mathcal{NP}$-hard.

Meyur et al. [18] compared network geometries using the Hausdorff distance after partitioning the geographic region into small rectangular grids and comparing the geometries for each grid. However, the Hausdorff metric is sensitive to outliers as it focuses only on the maximum possible distance between the pair of geometries. When the geometries coincide almost entirely except in a few small portions, the Hausdorff metric still records the discrepancy in those small portions without accounting for the similarity over the majority of portions. The similar approach used by Brovelli et al. [5] to compare a pair of road networks in a geographic region suffers from the same drawback. This necessitates a well-defined distance metric between networks with geographic embedding [2].

Several comparison methods have been proposed in the context of planar graphs embedded in a Euclidean space [6,19]. They include local and global metrics to compare road networks. The local metrics characterize the networks based on cliques and motifs, while the global metrics involve computing the *efficiency* of constructing the infrastructure network. The most efficient network is assumed to be the one with only straight line geometries connecting node pairs. Albeit useful to characterize network structures, these methods are not suitable for a numeric comparison of network geometries.

2 Methods

Following Mahji and Wenk [15], we use the term *geometric graph* to define network graphs embedded in a Euclidean space. Next, we define what we mean by structurally similar geometric graphs.

Definition 1 (Geometric graph). *A graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V}$ and edge set $\mathcal{E}$ is said to be a geometric graph of $\mathbb{R}^d$ if the set of nodes $\mathcal{V} \subset \mathbb{R}^d$ and the edges are Euclidean straight line segments $\{\overline{uv} \mid e := (u, v) \in \mathcal{E}\}$ which intersect (possibly) at their endpoints.*

Definition 2 (Structurally similar geometric graphs). *Two geometric graphs $\mathcal{G}_0(\mathcal{V}_0, \mathcal{E}_0)$ and $\mathcal{G}_1(\mathcal{V}_1, \mathcal{E}_1)$ are said to be* structurally similar *at the level*

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

of $\delta \geq 0$*, termed* δ*-similar, if* $\text{dist}\,(\mathcal{G}_0, \mathcal{G}_1) \leq \delta$ *for the distance function* dist *between the two graphs.*

We could consider a given network as a set of edge geometries. Hence we could consider the problem of comparing geometric graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ as that of comparing the set of edge geometries $\mathcal{E}_0$ and $\mathcal{E}_1$. In this paper, we propose a suitable distance that allows us to compare between a pair of geometric graphs or a pair of geometries. We use the multiscale flat norm, which has been well explored in the field of geometric measure theory, to define such a distance between the geometries.

The other aspect of this paper is to identify a suitable threshold δ for inferring the structural similarity of a pair of geometric graphs. But there is no general method to choose the threshold. Here, we perform a statistical analysis for our particular case of comparing power distribution networks. We validate the comparison results with visual inspection to conclude that the proposed metric serves its purpose to identify structurally similar geometric graphs.

2.1 Multiscale Flat Norm

We use the multiscale simplicial flat norm proposed by Ibrahim et al. [10] to compute the distance between two networks. We now introduce some background for this computation. A d-dimensional *current* T (referred to as a d-current) is a generalized d-dimensional geometric object with orientations (or direction) and multiplicities (or magnitude). An example of a 2-current is a surface with finite area (multiplicity) and a specific orientation (clockwise or counterclockwise). The boundary of T, denoted by ∂T, is a $(d-1)$-current. The *multiscale flat norm* of a d-current T at scale $\lambda \geq 0$ is defined as

$$\mathbb{F}_\lambda\,(T) = \min_S \left\{ V_d\,(T - \partial S) + \lambda V_{d+1}\,(S) \right\}, \tag{1}$$

where the minimum is taken over all $(d+1)$-currents S, and V_d denotes the d-dimensional *volume*, e.g., length in 1D or area in 2D. Computing the flat norm of a 1-current (curve) T identifies the optimal 2-current (area patches) S that minimizes the sum of the length of current $T - \partial S$ and the area of patch(es) S. Fig. 1 shows the flat norm computation for a generic 1D current T (blue). The 2D area patches S (magenta) are computed such that the expression in Eq. (1) is minimized for the chosen value of λ that ends up using most of the patch under the sharper spike on the left but only a small portion of the patch under the wider bump to the right.

The scale parameter λ can be intuitively understood as follows. Rolling a ball of radius $1/\lambda$ on the 1-current T traces the output current $T - \partial S$ and the untraced regions constitute the patches S. Hence we observe that for a large λ, the radius of the ball is very small and hence it traces major features while smoothing out (i.e., missing) only minor features (wiggles) of the input current. But for a small λ, the ball with a large radius smoothes out larger scale features (bumps) in the current. Note that for smaller λ, the cost of area patches is smaller

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

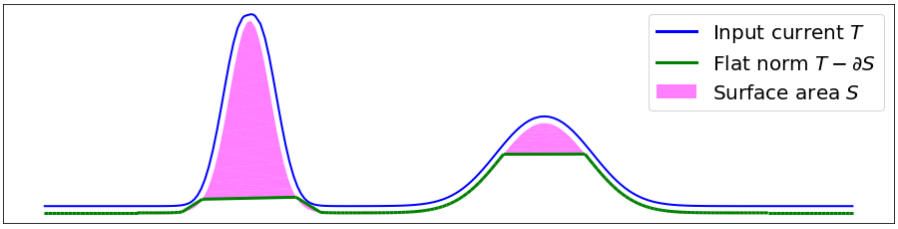

Fig. 1. Multiscale flat norm of a 1D current T (blue). The flat norm is the sum of length of the resulting 1D current $T - \partial S$ (green) and the area of 2D patches S (magenta). We show $T - \partial S$ slightly separated for easy visualization.

in the minimization function and hence more patches are used for computing the flat norm. We can use the flat norm to define a natural distance between a pair of 1-currents T_1 and T_2 as follows [10].

$$\mathbb{F}_\lambda\left(T_1, T_2\right) = \mathbb{F}_\lambda\left(T_1 - T_2\right) \tag{2}$$

We compute the flat norm distance between a pair of input geometries (synthetic and actual) as the flat norm of the current $T = T_1 - T_2$ where T_1 and T_2 are the currents corresponding to individual geometries. Let Σ denote the set of all line segments in the input current T. We perform a constrained triangulation of Σ to obtain a 2-dimensional finite oriented simplicial complex K. A constrained triangulation ensures that each line segment $\sigma_i \in \Sigma$ is an edge in K, and that T is an oriented 1-dimensional subcomplex of K.

Let m and n denote the numbers of edges and triangles in K. We can denote the input current T as a 1-chain $\sum_{i=1}^{m} t_i \sigma_i$ where σ_i denotes an edge in K and t_i is the corresponding multiplicity. Note that $t_i = -1$ indicates that orientation of σ_i and T are opposite, $t_i = 0$ denotes that σ_i is not contained in T, and $t_i = 1$ implies that σ_i is oriented the same way as T. Similarly, we define the set S to be the 2-chain of K and denote it by $\sum_{i=1}^{m} s_i \omega_i$ where ω_i denotes a 2-simplex in K and s_i is the corresponding multiplicity.

The boundary matrix $[\partial] \in \mathbb{Z}^{m \times n}$ captures the intersection of the 1 and 2-simplices of K. The entries of the boundary matrix $[\partial]_{ij} \in \{-1, 0, 1\}$. If edge σ_i is a face of triangle ω_j, then $[\partial]_{ij}$ is nonzero and it is zero otherwise. The entry is -1 if the orientations of σ_i and ω_j are opposite and it is $+1$ if the orientations agree.

We can respectively stack the t_i's and s_i's in m and n-length vectors $\mathbf{t} \in \mathbb{Z}^m$ and $\mathbf{s} \in \mathbb{Z}^n$. The 1-chain representing $T - \partial S$ is denoted by $\mathbf{x} \in \mathbb{Z}^m$ and is given as$\mathbf{x} = \mathbf{t} - [\partial]\,\mathbf{s}$. The multiscale flat norm defined in Eq. (1) can be computed by solving the following optimization problem:

$$\begin{aligned} \mathbb{F}_\lambda\left(T\right) = \min_{\mathbf{s} \in \mathbb{Z}^n} \sum_{i=1}^{m} w_i \left|x_i\right| + \lambda \left(\sum_{j=1}^{n} v_j \left|s_j\right| \right) \\ \text{s.t.} \quad \mathbf{x} = \mathbf{t} - [\partial]\,\mathbf{s}, \quad \mathbf{x} \in \mathbb{Z}^m, \end{aligned} \tag{3}$$

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

where $V_d(\tau)$ in Eq. (1) denotes the volume of the d-dimensional simplex τ. We denote volume of the edge σ_i as $V_1(\sigma_i) = w_i$ and set it to be the Euclidean length, and volume of a triangle τ_j as $V_2(\tau_j) = v_j$ and set it to be the area of the triangle.

In this work, we consider geometric graphs embedded on the geographic plane and are associated with longitude and latitude coordinates. We compute the Euclidean length of edge σ_i as $w_i = R\Delta\phi_i$ where $\Delta\phi_i$ is the Euclidean normed distance between the geographic coordinates of the terminals of σ_i and R is the radius of the earth. Similarly, the area of triangle τ_j is computed as $v_j = R^2\Delta\Omega_j$ where $\Delta\Omega_j$ is the solid angle subtended by the geographic coordinates of the vertices of τ_j.

Using the fact that the objective function is piecewise linear in $\mathbf{x}$ and $\mathbf{s}$, the minimization problem can be reformulated as an integer linear program (ILP) as follows:

$$\mathbb{F}_\lambda(T) = \min \sum_{i=1}^{m} w_i \left(x_i^+ + x_i^-\right) + \lambda \left(\sum_{j=1}^{n} v_j \left(s_j^+ + s_j^-\right)\right) \tag{4a}$$

$$\text{s.t.} \quad \mathbf{x}^+ - \mathbf{x}^- = \mathbf{t} - [\partial]\left(\mathbf{s}^+ - \mathbf{s}^-\right) \tag{4b}$$

$$\mathbf{x}^+, \mathbf{x}^- \geq 0, \quad \mathbf{s}^+, \mathbf{s}^- \geq 0 \tag{4c}$$

$$\mathbf{x}^+, \mathbf{x}^- \in \mathbb{Z}^m, \quad \mathbf{s}^+, \mathbf{s}^- \in \mathbb{Z}^n \tag{4d}$$

The linear programming relaxation of the ILP in Eq. (4) is obtained by ignoring the integer constraints Eq. (4d). We refer to this relaxed linear program (LP) as the ***flat norm LP***. Ibrahim et al. [10] showed that the boundary matrix $[\partial]$ is totally unimodular for our application setting. Hence the flat norm LP will solve the ILP, and hence the flat norm can be computed in polynomial time.

2.2 Proposed Algorithm

Algorithm 1 describes how we compute the distance between a pair of geometries with the associated embedding on a metric space $\mathcal{M}$. We assume that the geometries (networks) $\mathcal{G}_1(\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2(\mathcal{V}_2, \mathcal{E}_2)$ with respective node sets $\mathcal{V}_1, \mathcal{V}_2$ and edge sets $\mathcal{E}_1, \mathcal{E}_2$ have no one-to-one correspondence between the $\mathcal{V}_i$'s or $\mathcal{E}_i$'s. Note that each vertex $v \in \mathcal{V}_1, \mathcal{V}_2$ is a point and each edge $e \in \mathcal{E}_1, \mathcal{E}_2$ is a straight line segment in $\mathcal{M}$. We consider the collection of edges $\mathcal{E}_1, \mathcal{E}_2$ as input to our algorithm. First, we orient the edge geometries in a particular direction (left to right in our case) to define the currents T_1 and T_2, which have both magnitude and direction. Next, we consider the bounding rectangle $\mathcal{E}_{\text{bound}}$ for the edge geometries and define the set Σ to be triangulated as the set of all edges in either geometry and the bounding rectangle. We perform a constrained Delaunay triangulation [29] on the set Σ to construct the 2-dimensional simplicial complex K. The constrained triangulation ensures that the set of edges in Σ is included in the simplicial complex K. Then we define the currents T_1 and T_2 corresponding to the respective edge geometries $\mathcal{E}_1$ and $\mathcal{E}_2$ as 1-chains in K. Finally, the flat norm LP is solved to compute the simplicial flat norm.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

Algorithm 1: Distance between a pair of geometries

Input: Geometries $\mathcal{E}_1, \mathcal{E}_2$

Parameter: Scale λ

1: Orient each edge in the edge sets from left to right:
$\tilde{\mathcal{E}}_1 := \mathsf{Orient}\left(\mathcal{E}_1\right)$; $\tilde{\mathcal{E}}_2 := \mathsf{Orient}\left(\mathcal{E}_2\right)$.

2: Find bounding rectangle for the pair of geometries: $\mathcal{E}_{\text{bound}} = \mathsf{rect}\left(\tilde{\mathcal{E}}_1, \tilde{\mathcal{E}}_2\right)$.

3: Define the set of line segments to be triangulated: $\Sigma = \tilde{\mathcal{E}}_1 \cup \tilde{\mathcal{E}}_2 \cup \mathcal{E}_{\text{bound}}$.

4: Perform constrained triangulation on set Σ to construct 2-dimensional simplicial complex K.

5: Define the currents T_1, T_2 as 1-chains of oriented edges $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$ in K.

6: Solve the flat norm LP to compute flat norm $\mathbb{F}_\lambda\left(T_1 - T_2\right)$.

Output: Flat norm distance $\mathbb{F}_\lambda\left(T_1 - T_2\right)$.

2.3 Normalized Flat Norm

Recall that in our context of synthetic power distribution networks, the primary goal of comparing a synthetic network to its actual counterpart is to infer the quality of the replica or the *digital duplicate* synthesized by the framework. The proposed approach using the flat norm for structural comparison of a pair of geometries provides us a method to perform global as well as local comparison. While we can produce a global comparison by computing the flat norm distance between the two networks, it may not provide us with complete information on the quality of the synthetic replicate. On the other hand, a local comparison can provide us details about the framework generating the synthetic networks. For example, a synthetic network generation framework might produce higher quality digital replicates of actual power distribution networks for urban regions as compared to rural areas. A local comparison highlights this attribute and identifies potential use case scenarios of a given synthetic network generation framework.

Furthermore, availability of actual power distribution network data is sparse due to its proprietary nature. We may not be able to produce a global comparison between two networks due to unavailability of network data from one of the sources. Hence, we want to restrict our comparison to only the portions in the region where data from either network is available, which also necessitates a local comparison between the networks.

For a local comparison, we consider uniform sized regions and compute the flat norm distance between the pair of geometries within the region. However, the computed flat norm is dependent on the length of edges present within the region from either network. Hence we define the *normalized* multiscale flat norm, denoted by $\widetilde{\mathbb{F}}_\lambda$, for a given region as

$$\widetilde{\mathbb{F}}_\lambda\left(T_1 - T_2\right) = \frac{\mathbb{F}_\lambda\left(T_1 - T_2\right)}{|T_1| + |T_2|}\,. \tag{5}$$

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

For a given parameter ϵ, a local region is defined as a square of size $2\epsilon \times 2\epsilon$ steradians. Let $T_{1,\epsilon}$ and $T_{2,\epsilon}$ denote the currents representing the input geometries inside the local region characterized by ϵ. Note that the "amount" or the total length of network geometries within a square region varies depending on the location of the local region. In this case, the lengths of the network geometries are respectively $|T_{1,\epsilon}|$ and $|T_{2,\epsilon}|$. Therefore, we use the ratio of the total length of network geometries inside a square region to the parameter ϵ to characterize this "amount" and denote it by $|T|/\epsilon$ where

$$|T|/\epsilon = \frac{|T_{1,\epsilon}| + |T_{2,\epsilon}|}{\epsilon}\,. \tag{6}$$

Note that while performing a comparison between a pair of network geometries in a local region using the multiscale flat norm, we need to ensure that comparison is performed for similar length of the networks inside similar regions. Therefore, the ratio $|T|/\epsilon$, which indicates the length of networks inside a region scaled to the size of the region, becomes an important aspect of characterization while performing the flat norm based comparison.

3 Results and Discussion

We use the proposed multiscale flat norm to compare a pair of network geometries from power distribution networks for a region in a county in USA. The two networks considered are the actual power distribution network for the region and the synthetic network generated using the methodology proposed by Meyur et al. [17]. We provide a brief overview of these networks.

Actual network. The actual power distribution network was obtained from the power company serving the location. Due to its proprietary nature, node and edge labels were redacted from the shared data. Further, the networks were shared as a set of handmade drawings, many of which had not been drawn to a well-defined scale. We digitized the drawings by overlaying them on OpenStreetMaps [23] and georeferencing to particular points of interest [8]. Geometries corresponding to the actual network edges are obtained as shape files.

Synthetic network. The synthetic power distribution network is generated using a framework with the underlying assumption that the network follows the road network infrastructure to a significant extent [17]. To this end, the residences are connected to local pole top transformers located along the road network to construct the low voltage (LV) secondary distribution network. The local transformers are then connected to the power substation following the road network leading to the medium voltage (MV) primary distribution network. That is, the primary network edges are chosen from the underlying road infrastructure network such that the structural and power engineering constraints are satisfied.

3.1 Comparing Network Geometries

The primary goal of computing the flat norm is to compare the pair of input geometries. As mentioned earlier, the flat norm provides an accurate measure of

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

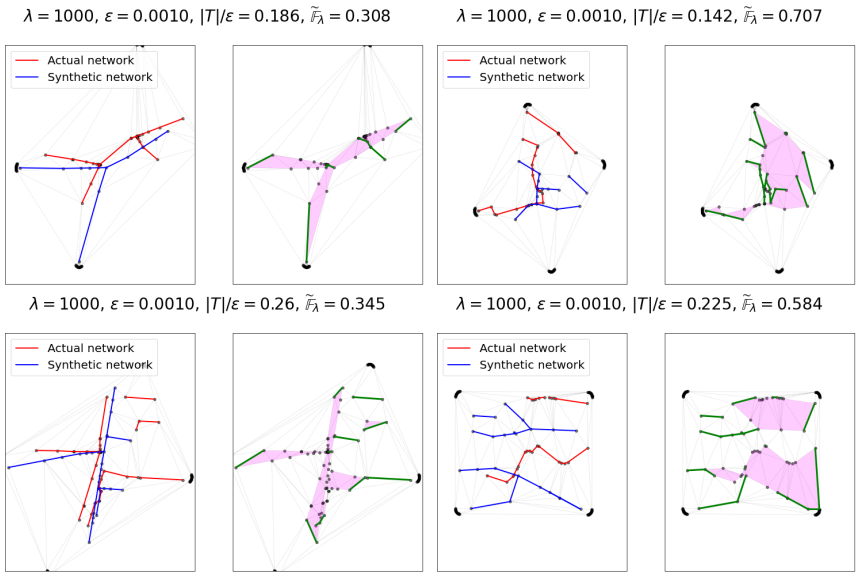

Fig. 2. Normalized flat norm (with scale $\lambda = 1000$) distances for pairs of regions in the network of same size ($\epsilon = 0.001$) with similar $|T|/\epsilon$ ratios (two pairs each in the top and bottom rows). The pairs of geometries for the first plot (on left) are quite similar, which is reflected in the low flat norm distances between them. The network geometries on the right plots are more dissimilar and hence the flat norm distances are high.

difference between the geometries by considering both the length deviation and area patches in between the geometries. Further, we normalize the computed flat norm to the total length of the geometries. In this section, we show examples where we computed the normalized flat norm for the pair of network geometries (actual and synthetic) for a few regions.

The top two plots in Fig. 2 show two regions characterized by $\epsilon = 0.001$ and almost similar $|T|/\epsilon$ ratios. This indicates that the length of network scaled to the region size is almost equal for the two regions. From a mere visual perspective, we can conclude that the first pair of network geometries resemble each other where as the second pair are fairly different. This is further validated from the results of the flat norm distance between the network geometries computed with the scale $\lambda = 1000$, since the first case produces a smaller flat norm distance compared to the latter. The bottom two plots show another example of two regions with almost similar $|T|/\epsilon$ ratios and enable us to infer similar conclusions. The results strengthens our case of using flat norm as an appropriate measure to perform a local comparison of network geometries.

3.2 Comparison of Flat Norm and Hausdorff Distance Metrics

In this section, we compare the proposed flat norm metric for structural comparison with the Hausdorff distance metric which has been extensively used in

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

the literature for similar purposes. The Hausdorff distance is considered to be a *stable* metric since minor perturbations to the geometries do not affect the metric. While this property is advantageous when we are dealing with noisy data, this fails to capture structural differences unless they are significantly large.

The comparison metric is said to be "stable" if the computed normalized flat norm for the pair of perturbed network geometries is close to the normalized flat norm of the unperturbed geometries. A perturbed network is similar to the original network with only the geographic embeddings of the nodes perturbed. To this end, we consider a circular region around each node in the network by defining a perturbation radius (in meters). We then uniformly sample a point in each circular region and use them as the perturbed embeddings of the nodes. We compare the stability of the flat norm metric with the Hausdorff metric—both empirically on our sample networks and using a simple theoretical example.

We consider two simple curves T_1, T_2 in the plane whose end points are the same (see Fig. 3). We perturb T_1 within a small neighborhood of each point on it while keeping T_2 fixed and the end points of both curves also fixed. Hence we consider perturbed versions $\widetilde{T}_1$ that lie within an ε-tube of T_1. We could have cases where $\widetilde{T}_1$ lies mostly at the upper envelope of this ε-tube, or mostly at the lower envelope. In both cases, one would expect the distance between $\widetilde{T}_1$ and T_2 to also change significantly (from that between T_1 and T_2). The flat norm distance accurately captures all such changes (to keep the example simple, we consider the default flat norm distance and not the normalized version). At the same time, both such variations could have the same Hausdorff distance H from T_2 as T_1, which completely misses all the changes to T_1 in either case.

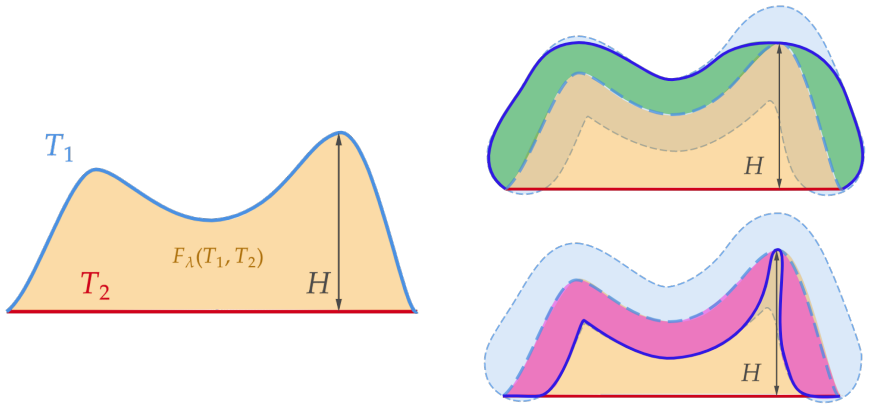

Fig. 3. Left: Curves T_1 and T_2 with shared end points, and their flat norm distance $\mathbb{F}_\lambda(T_1, T_2)$ and Hausdorff distance H. **Right:** Two perturbed versions $\widetilde{T}_1$ (solid blue) of T_1 (now in dashed blue) that lie within an ε-neighborhood of T_1. The Hausdorff distance between $\widetilde{T}_1$ and T_2 remains same, i.e., H. But the flat norm distance increased in the first case (Right, top) as captured by the green patch, and it decreased in the second case (Right, Bottom) by the area shown in pink.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

A modification of this example can illustrate the other extreme case—when Hausdorff distance changes by a lot but the flat norm distance does not change much at all. Consider moving *only* the highest point on T_1 further up so that Hausdorff distance becomes $2H$ (this is the point on T_1 in the left figure in Fig. 3 at the arrow indicating the Hausdorff distance H). We keep T_1 still a connected curve, thus creating a sharp spike in it. While the Hausdorff distance between the curves has doubled, the flat norm distance sees only a minute increase as measured by the tiny area under this spike. Once again, the flat norm distance accurately captures the intuition that the curves have *not* changed much when just a single point moves away while the rest of the curve stays the same. Hence the flat norm is a more robust metric that always captures significant changes while maintaining stability to small perturbations.

We observe similar behavior to those illustrated by the theoretical example (Fig. 3) in our computational experiments. Fig. 4 shows scatter plots denoting empirical distribution of percentage deviation of the two metrics from the original values $\left(\%\Delta\mathbb{D}_{\text{Haus}}, \%\Delta\widetilde{\mathbb{F}}_{\lambda}\right)$ for a local region. The perturbations are considered for three different radii shown in separate plots. We note that the percentage deviations in the two metrics are comparable in most cases. In other words, neither metric behaves abnormally for a small perturbation in one of the networks.

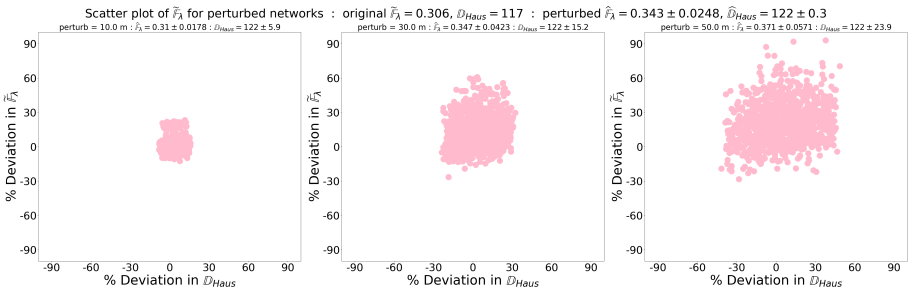

Fig. 4. Scatter plots showing effect of network perturbation on normalized flat norm and Hausdorff distances for a local region. The percentage deviation in the metrics for the perturbations are shown along each axis. We do not observe significantly large deviations in any one metric for a given perturbation.

Next, we compare the sensitivity of the two metrics to outliers. Here, we consider a single random node in one of the network and perturb it. Fig. 5 shows the sensitivity of the metrics to these outliers. The original normalized flat norm and Hausdorff distance metrics are shown by the horizontal and vertical dashed lines respectively. The points along the horizontal dashed line denote the cases where Hausdorff distance metric is more sensitive to the outliers, while the normalized flat norm metric remains the same. These cases occur when the perturbed random node determines the Hausdorff distance, similar to the

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

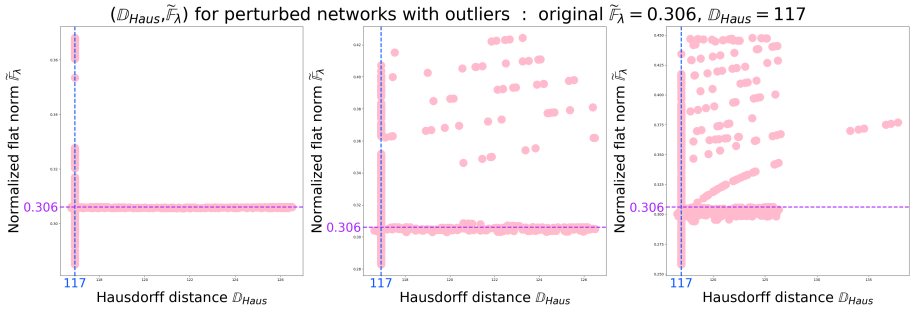

Fig. 5. Scatter plots showing effect of few outliers on normalized flat norm and Hausdorff distance for a local region. The original normalized flat norm and Hausdorff distance are highlighted by the dashed horizontal and vertical lines. We observe multiple cases where the Hausdorff distance is more sensitive to outliers compared to the proposed normalized flat norm metric.

theoretical case (where Hausdorff distance went from H to $2H$). On the flip side, the points along the vertical dashed line denote the Hausdorff distance remaining unchanged while the normalized flat norm metric shows variation. Just as in the theoretical example (Fig. 3), such variation in the normalized flat norm metric implies a variation in the network structure. However, such variation is not captured by the Hausdorff distance metric. Hence, our proposed metric is capable of identifying structural differences due to perturbations, while remaining stable when widely separated nodes (which are involved in Hausdorff distance computation) are perturbed. The other points which are neither on the horizontal nor on the vertical dashed lines indicate that either metric is able to identify the structural variation due to the perturbation.

4 Conclusions

We have proposed a fairly general metric to compare a pair of network geometries embedded on the same plane. Unlike standard approaches that map the geometries to points in a possibly simpler space and then measuring distance between those points [11], or comparing "signatures" for the geometries, our metric works directly in the input space and hence allows us to capture all details in the input. The metric uses the multiscale flat norm from geometric measure theory, and can be used in more general settings as long as we can triangulate the region containing the two geometries. It is impossible to derive *standard* stability results for this distance measure that imply only small changes in the flat norm metric when the inputs change by small amount—there is no alternative metric to measure the *small change in the input*. For instance, our theoretical example (in Fig. 3) shows that the commonly used Hausdorff metric cannot be used for this purpose. At the same time, we do get natural stability results for our distance following the properties of the flat norm—small changes in the input geometries lead to only small changes in the flat norm distance between them [9,20].

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

We use the proposed metric to compare a pair of power distribution networks: (i) actual power distribution networks of two locations in a county of USA obtained from a power company and (ii) synthetically generated digital duplicate of the network created for the same geographic location. The proposed comparison metric is able to perform global as well as local comparison of network geometries for the two locations. We discuss the effect of different parameters used in the metric on the comparison. Further, we validate the suitability of using the flat norm metric for such comparisons using computation as well as theoretical examples.

References

1. Ahmed, M., Fasy, B.T., Hickmann, K.S., Wenk, C.: A path-based distance for street map comparison. ACM Trans. Spatial Algorithms Syst. **1**(1) (2015)
2. Ahmed, M., Fasy, B.T., Wenk, C.: Local persistent homology based distance between maps. In: Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. p. 43–52. SIGSPATIAL '14, Association for Computing Machinery, New York, NY, USA (2014)
3. Bai, Y., Ding, H., Bian, S., Chen, T., Sun, Y., Wang, W.: Simgnn: A neural network approach to fast graph similarity computation. `https://arxiv.org/abs/1808.05689` (2018), last accessed 26 Sept 2022
4. Bidel, A., Schelo, T., Hamacher, T.: Synthetic distribution grid generation based on high resolution spatial data. In: 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe). pp. 1–6. IEEE, Bari, Italy (2021)
5. Brovelli, M.A., Minghini, M., Molinari, M., Mooney, P.: Towards an automated comparison of openstreetmap with authoritative road datasets. Transactions in GIS **21**(2), 191–206 (2017)
6. Cardillo, A., Scellato, S., Latora, V., Porta, S.: Structural properties of planar graphs of urban street patterns. Phys. Rev. E **73**, 066107 (2006)
7. Eppstein, D.: Subgraph isomorphism in planar graphs and related problems. In: Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms. p. 632–640. SODA '95, Society for Industrial and Applied Mathematics, USA (1995)
8. ESRI: Georeferencing a raster to a vector (2022), last accessed 26 Sept 2022
9. Federer, H.: Geometric Measure Theory. Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag, New York (1969)
10. Ibrahim, S., Krishnamoorthy, B., Vixie, K.R.: Simplicial flat norm with scale. Journal of Computational Geometry **4**(1), 133–159 (2013)
11. Kendall, D.G., Barden, D.M., Carne, T., Le, H.: Shape and Shape Theory. Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, USA (2009)
12. Krishnan, V., Bugbee, B., Elgindy, T., Mateo, C., Duenas, P., Postigo, F., Lacroix, J.S., Roman, T.G.S., Palmintier, B.: Validation of synthetic u.s. electric power distribution system data sets. IEEE Transactions on Smart Grid **11**(5), 4477–4489 (2020)
13. Li, H., Wert, J.L., Birchfield, A.B., Overbye, T.J., Roman, T.G.S., Domingo, C.M., Marcos, F.E.P., Martinez, P.D., Elgindy, T., Palmintier, B.: Building highly detailed synthetic electric grid data sets for combined transmission and distribution systems. IEEE Open Access Journal of Power and Energy **7**, 478–488 (2020)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5

14. Liang, M., Meng, Y., Wang, J., Lubkeman, D.L., Lu, N.: Feedergan: Synthetic feeder generation via deep graph adversarial nets. IEEE Transactions on Smart Grid **12**(2), 1163–1173 (2021)
15. Majhi, S., Wenk, C.: Distance measures for geometric graphs. `https://arxiv.org/abs/2209.12869` (2022), last accessed 26 Sept 2022
16. Mateo, C., Postigo, F., de Cuadra, F., Roman, T.G.S., Elgindy, T., Dueñas, P., Hodge, B.M., Krishnan, V., Palmintier, B.: Building large-scale u.s. synthetic electric distribution system models. IEEE Transactions on Smart Grid **11**(6), 5301–5313 (2020)
17. Meyur, R., Marathe, M., Vullikanti, A., Mortveit, H., Swarup, S., Centeno, V., Phadke, A.: Creating realistic power distribution networks using interdependent road infrastructure. In: 2020 IEEE International Conference on Big Data (Big Data). pp. 1226–1235. IEEE, Atlanta, GA, USA (2020)
18. Meyur, R., Vullikanti, A., Swarup, S., Mortveit, H., Centeno, V., Phadke, A., Poor, V., Marathe, M.: Ensembles of realistic power distribution networks. Proceedings of the National Academy of Sciences **119**(26), e2123355119 (2022)
19. Morer, I., Cardillo, A., Díaz-Guilera, A., Prignano, L., Lozano, S.: Comparing spatial networks: A one-size-fits-all efficiency-driven approach. Phys. Rev. E **101**, 042301 (2020)
20. Morgan, F.: Geometric Measure Theory: A Beginner's Guide. Academic Press, Cambridge, MA, USA, fourth edn. (2008)
21. Morgan, S.P., Vixie, K.R.: L^1TV computes the flat norm for boundaries. Abstract and Applied Analysis **2007**, Article ID 45153,14 pages (2007)
22. Ok, S.: A graph similarity for deep learning. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 1–12. Curran Associates, Inc., Vancouver, BC, Canada (2020)
23. Open Street Map Foundation: Open Street Maps (2022), last accessed 20 Aug 2022
24. Paaβen, B.: Revisiting the tree edit distance and its backtracing: A tutorial. `https://arxiv.org/abs/1805.06869` (2022), last accessed 26 Sept 2022
25. Riba, P., Fischer, A., Lladós, J., Fornés, A.: Learning graph edit distance by graph neural networks. Pattern Recognition **120**, 108132 (2021)
26. Roy, K.C., Hasan, S., Culotta, A., Eluru, N.: Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. Transportation Research Part C: Emerging Technologies **131**, 103339 (2021)
27. Saha, S.S., Schweitzer, E., Scaglione, A., Johnson, N.G.: A framework for generating synthetic distribution feeders using openstreetmap. In: 2019 North American Power Symposium (NAPS). pp. 1–6. IEEE, Wichita, KS, USA (2019)
28. Schweitzer, E., Scaglione, A., Monti, A., Pagani, G.A.: 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 (2017)
29. Si, H.: Constrained Delaunay tetrahedral mesh generation and refinement. Finite Elements in Analysis and Design **46**, 33–46 (2010)
30. Tantardini, M., Ieva, F., Tajoli, L., Piccardi, C.: Comparing methods for comparing networks. Scientific Reports **9**(1), 17557 (2019)
31. Xu, H.: An algorithm for comparing similarity between two trees. `https://arxiv.org/abs/1508.03381` (2015), last accessed 26 Sept 2022

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36027-5_5