SCOPE: Bridging Global Form and Local Geometry for Urban Street Network Generation

Wenxuan Guo*†, Yaohui Jin* and Yanyan Xu*
*MoE Key Lab of Artificial Intelligence, AI Institute, School of Computer Science
Shanghai Jiao Tong University
Shanghai, China
†Email: arya_g@sjtu.edu.cn

***Abstract*—Generative modeling of urban street networks poses a distinct challenge in geometric deep learning. It requires synthesizing large-scale graphs that respect physical constraints while capturing statistical patterns shared across real-world cities. Existing approaches often struggle to balance global spatial coherence with precise topological connectivity. Furthermore, many are limited to modifying reference templates rather than generating diverse, novel structures from scratch. In this work, we propose a scalable two-stage framework that effectively decouples these modalities. We employ an autoregressive Transformer to learn diverse node distributions from coordinate sequences, and introduce SCOPE, a specialized edge predictor that conditions on global spatial context while respecting local geometric constraints. Trained on over 35,000 cities, our approach outperforms baselines with 0.91 average precision in edge prediction and faithfully reproduces high-order graph statistics. Qualitatively, it synthesizes physically plausible networks, demonstrating the capacity to hypothesize novel urban forms beyond the training distribution. Additionally, we show that the learned geometric embeddings intrinsically capture spatial laws, offering interpretability alongside generative performance.**

***Index Terms*—Spatial graphs, urban street networks, deep learning, Transformer, graph generation**

I. Introduction

Urban generative modeling has emerged as a critical research direction, enabling autonomous generation of urban data and designs across diverse systems at unprecedented scale [1], [2]. Recent generative approaches span urban planning, network synthesis, and resilience simulation [3]–[5]. These advances demonstrate that urban generative modeling is maturing toward data-driven hypothesis generation across large and diverse urban contexts. Within this landscape, generating realistic street networks remains a distinctive neural learning challenge. Street networks are large-scale spatial graphs where node geometry and edge topology must be jointly consistent with physical constraints and population-level statistical regularities (e.g., natural barriers, local grid continuity, and hierarchical arterial structures) [6], [7]. Unlike abstract graph generation or image synthesis tasks, this problem requires learning from tens of thousands of real cities to capture generalizable urban patterns without relying on template-based or domain-specific procedural rules. The key learning difficulty is to acquire both precise local connectivity constraints and global structural awareness—a dual demand that has proven difficult for existing neural architectures.

This work was jointly supported by the National Natural Science Foundation of China (72432007), the Shanghai Key Laboratory of Urban Design and Urban Science, NYU Shanghai Research Grants (No.2025YYXu_LOUD), and the Fundamental Research Funds for the Central Universities. The computations in this paper were run on the AI for Science Platform, supported by the Artificial Intelligence Institute at Shanghai Jiao Tong University.

Existing street-network generators generally navigate a trade-off between structural validity and spatial coherence. Image-based approaches [8]–[10] capture broad visual patterns but often require non-trivial post-processing to recover functional graphs. Graph-native methods [11], [12] preserve connectivity more naturally, yet many depend on seeds or reference templates and therefore have limited ability to generate diverse cities independently from scratch. Heuristic models [13], [14] provide interpretable geometric priors, but their fixed local rules are often too rigid to represent the full variability observed across real urban systems.

We propose a unified, data-driven framework trained on over 35,000 cities that bridges these paradigms. Our key insight is to decouple generation into two specialized stages: geometry and topology. First, we treat node placement as a sequence modeling task, employing a Transformer to autoregressively learn diverse spatial distributions without requiring geometric templates. Second, we introduce Spatial-Context Optimized Prediction of Edges (SCOPE), a novel edge predictor that fuses the global structural awareness of Convolutional Neural Networks (CNNs) with local proximity constraints. This architecture allows the model to learn statistical regularities from massive datasets while enforcing the topological validity characteristic of graph methods, effectively synthesizing realistic, city-scale street networks from scratch. We validate our framework through extensive experiments with real cities, benchmarking against both classical geometric heuristics and neural graph autoencoder baselines. Our results confirm that the two-stage architecture effectively captures complex urban topologies, significantly outperforming baselines in reproducing statistics like betweenness centrality and clustering coefficients. Ablation studies further verify that the global context mechanism in SCOPE is critical for reconstructing long-range arterial structures. Finally, we analyze the learned latent space, proving that the model spontaneously acquires spatially autocorrelated representations that mirror the First Law of Geography. The key contributions of our work are:

- A scalable graph-native framework that generates diverse

city layouts from scratch by factorizing the problem into autoregressive geometric sequencing and context-aware topological wiring.
- SCOPE, a novel edge predictor that integrates global CNN-based context with local constraints, successfully modeling structural exclusions that standard homophily-based graph autoencoders fail to capture.
- Empirical validation on real cities showing significant superiority over neural and heuristic baselines, particularly in reproducing higher-order statistics.
- Verification of spontaneous emergence of spatially autocorrelated embeddings, revealing an interpretable latent manifold that mirrors physical geographic laws.

II. Related Work

We review generative models for street networks across three families: image-based models, graph-native models, and heuristic approaches.

Image-based generative models use a visual representation of road networks and synthesize street networks as raster images. The preliminary work StreetGAN [8] applies adversarial training to rasterized patches from a single city. However, it exhibits severe artifacts arising from its image-based representation: closely adjacent lanes merge into thick blobs, and complex structures like highway ramps are often lost. Furthermore, limited patch sizes prevent capturing large-scale highway organization. Consequently, recovering valid graphs requires complex non-differentiable post-processing that breaks end-to-end learnability. Similarly, the Variational Autoencoder (VAE)-based method [9] attempts to learn latent representations from city images, but their decoders struggle to reconstruct sharp street details and yield blurred images. CRT-GAN [10] advances the adversarial training paradigm in [8] by introducing controllable generation and a differentiable recovery layer, yet remains restricted to small-scale graphs.

Graph-native generative models preserve topological structure by operating directly on graph representations. Neural Turtle Graphics [12] uses Recurrent Neural Network (RNN) encoder-decoders to incrementally grow graphs from initial seeds, evaluated on 17 cities. RoadNetGAN [11] adapts NetGAN [15] with spatial constraints, generating sequences of displacement vectors through local random walks conditioned on reference backbones. While these models maintain connectivity, they rely on existing reference graphs or manually defined seeds to initiate generation. This dependence restricts them to extending known structures rather than synthesizing novel graphs from scratch. Consequently, they cannot independently generate large and diverse batches of synthetic cities. The most closely related work [16] proposes a two-stage pipeline using a Transformer for node coordinates and a Variational Graph Autoencoder (VGAE) for edge prediction. However, this architecture suffers from limitations of inference-time constraints on graph structure and over-reliance on homophily assumptions.

Heuristic and proximity-based models provide interpretable strong baselines grounded in geometric graph theory. Planar proximity graphs, such as Minimum Spanning Trees (MST) and Relative Neighborhood Graphs (RNG), determine connectivity using local exclusion regions [17]–[19]. Beta-skeleton graphs unify these via a parameter β to match local properties [14], [20]. However, these models impose a rigid planarity constraint which is only an approximation; real-world street networks are not strictly planar due to grade-separated infrastructure (e.g., overpasses, tunnels) [21]. Furthermore, since these definitions rely strictly on local geometric neighborhoods, they fundamentally lack global context. They cannot represent large-scale spatial features such as natural obstacles, arterial backbones, or hierarchical street patterns that transcend immediate neighborhoods. Constructive heuristics like the Universal Model [13] address this by growing edges to match specific distributions. Yet, these methods are explicitly optimized for narrow statistical targets (e.g., degree or betweenness centrality) and fail to capture the broader, data-driven diversity of urban forms and higher-order spatial regularities found in real cities.

Our approach addresses these limitations by establishing a graph-native, data-driven framework. We preserve topological validity without post-processing steps and support the independent generation of large, diverse batches of novel cities. Our model bridges the gap between local heuristics and global reasoning: by factorizing geometry and topology, we combine the global structural awareness of convolutional networks with proximity-aware local constraints. This ensures generated graphs respect both large-scale urban organization and precise local connectivity rules, as demonstrated in Fig. 1.

III. Problem Formulation

We model a street network as an undirected spatial graph $G = (V, E)$ embedded in a two-dimensional Euclidean plane, with $N = |V|$ and $M = |E|$ denoting the number of nodes and edges. The node set $V = \{v_1, \ldots, v_N\}$ corresponds to street intersections and endpoints, where each node $v_i \in V$ is associated with a coordinate vector $\mathbf{x}_i = (x_i, y_i) \in \mathbb{R}^2$. The edge set $E \subseteq V \times V$ represents physical road segments connecting these nodes. An edge $e_{ij} = (v_i, v_j) \in E$ exists if there is a traversable road connection between node v_i and v_j. The graph is assumed to be undirected and simple, containing no self-loops or parallel edges. This simplified and abstract representation facilitates graph generation algorithms without sacrificing the functionality and connectivity of road networks. The goal of street network generation is to learn the underlying data distribution $P(G)$ from a dataset of observed street networks $\mathcal{D}$. A generative model with parameters Θ learns to approximate this distribution with $P_\Theta(G)$, enabling the synthesis of novel graphs that preserve the spatial and topological characteristics of real-world cities.

IV. Methodology

A. Factorizing Street Network Generation

To address the complexity of generating spatial graphs, we follow the factorization approach proposed in [16]. The joint probability of generating a graph G is decomposed into two

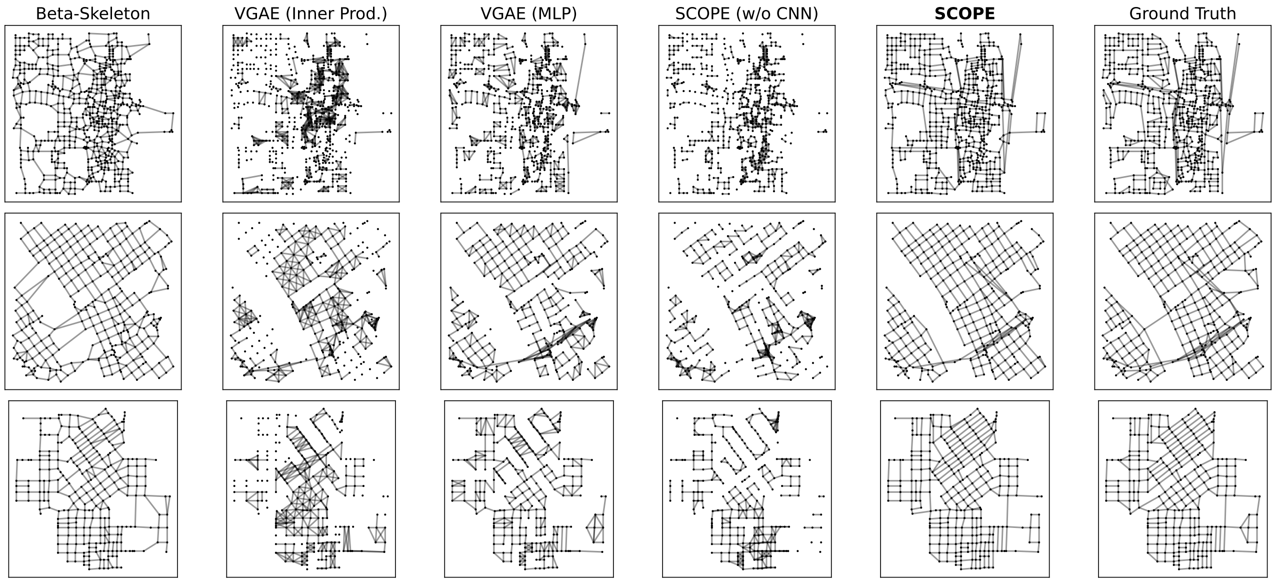

Fig. 1. Qualitative comparison of edge reconstruction on held-out test graphs. Baselines (VGAE variants, SCOPE w/o CNN) struggle with long-range connections and global organization, often over-emphasizing local links. The Beta-Skeleton enforces planarity but misses arterial backbones (top) and gaps (middle). In contrast, SCOPE combines global spatial context with proximity-based candidates, adapting to heterogeneous hierarchies (top) and lattice-like layouts with multiple orientations (middle and bottom). KNN is omitted for clarity.

sequential processes: generating the nodes (geometry) and then generating the edges (topology) conditioned on the nodes. Mathematically, this is expressed as:

$$P_{\Theta}(G) = P_{\Theta}(V, E) = P_{\theta}(V) P_{\phi}(E|V), \tag{1}$$

where $P_\theta(V)$ models the probability of the node spatial distribution and $P_\phi(E|V)$ models the wiring mechanism given the fixed node geometry, parameterized by θ and ϕ respectively. This separation prevents the entanglement of training objectives, ensuring that the learning of geometric distributions does not interfere with the complexity of topological inference. Fig. 2 illustrates the overall architecture of our two-stage generative model. The following sections detail the design of each component.

B. Learning Spatial Geometry as Sequences

We adopt the autoregressive Transformer model from [16] and [22] to learn the node distribution $P_\theta(V)$. We flatten 2D graphs into sequences, where each token represents a coordinate on the x or y axis. Continuous coordinates are discretized into a $2^q \times 2^q$ grid (q is the quantization level), so each coordinate is represented as an integer in $[0, 2^q - 1]$. To support variable-size graphs, we use standard sequence boundary markers: a Start-of-Sequence (SOS) symbol and an End-of-Sequence (EOS) token. SOS is a learned start embedding prepended to every sequence, while EOS is implemented as an additional discrete token in the output vocabulary that indicates termination. To serialize the set of nodes V, we impose a canonical ordering: nodes are sorted in ascending order first by y-coordinate, then by x-coordinate. The set V becomes a sequence $S(V) = (y_1, x_1, y_2, x_2, \ldots, y_N, x_N)$. In this way, the probability of the node set is modeled as the product of conditional probabilities over this sequence:

$$P_\theta(V) = \prod_{t=1}^{2N} P_\theta(s_t | s_{<t}), \tag{2}$$

where s_t is the t-th token in the flattened sequence and θ denotes the learnable parameters of the node model.

The Transformer processes this sequence using a learned embedding strategy. Each token s_i is mapped to a dense vector via three embedding layers: *Value Embedding* $h_v(s_i)$ represents the integer coordinate value in $[0, 2^q - 1]$; *Coordinate Embedding* $h_c(s_i)$ distinguishes whether the token is an x or y coordinate with binary values; and *Positional Embedding* $h_p(s_i)$ encodes the index of the node in the sorted sequence with values in $[0, N - 1]$. These embeddings are summed to yield the final embedding $h(s_i) = h_v(s_i) + h_c(s_i) + h_p(s_i)$ and are fed into the Transformer decoder, which outputs a probability distribution over the 2^q+1 discrete tokens (2^q token values plus EOS). Generation stops when EOS is produced or when the node budget is reached. The Transformer backbone is an 18-layer decoder-only structure following [22].

C. SCOPE: Spatial-Context Optimized Prediction of Edges

Given the synthesized node set V from the Transformer, the goal of the edge model is to predict a plausible wiring pattern that matches real cities under the given geometry (e.g., density gradients, alignments, and barrier-like voids). In the context of street networks, proximity is a crucial feature [17], [20]. Most intersections connect to geographically nearby neighbors. This property is also leveraged in proximity-based heuristics such as K-Nearest Neighbors (KNN) and proximity graphs [14], [18].

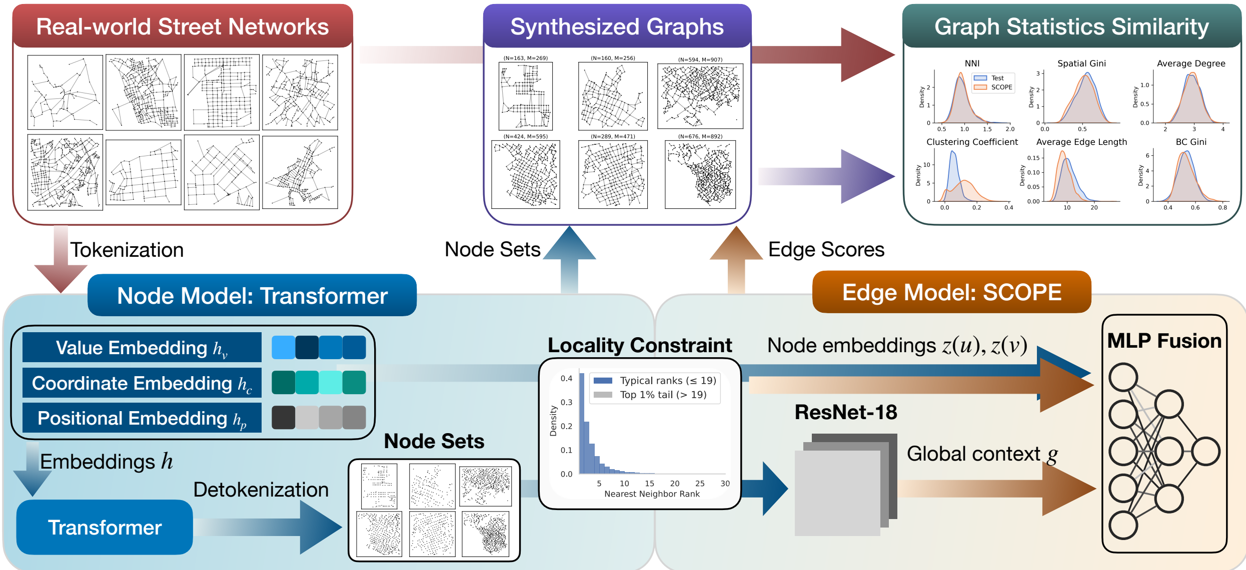

Fig. 2. Overview of the two-stage generative framework. The Transformer autoregressively samples quantized node coordinates as a sequence to model geometry. SCOPE predicts edges on the generated node set by fusing local node embeddings with global spatial context extracted by a CNN and respects a locality constraint. The synthesized graphs are evaluated on distributional statistics and compared with empirical values.

We quantify this locality by analyzing the Nearest Neighbor (NN) rank of connected nodes. For an edge $e = (u, v)$, let $\mathrm{rank}(v|u)$ be the rank of v in the ordered list of all other nodes sorted by Euclidean distance from u. We define the edge rank as $R(e) = \min(\mathrm{rank}(v|u), \mathrm{rank}(u|v))$. We observe that 95% of edges connect nodes within the 9-NN range, and 99% within the 19-NN range, as shown in Fig. 2.

Prior deep baselines also use VGAEs [16] to exploit locality via graph neural networks (GNNs). However, this framework has several limitations for our generation setting: 1. **Inference Constraint:** VGAEs typically rely on GNNs that require an existing graph structure. During generation, however, only the set of generated nodes is available, forcing the encoder to degenerate into an MLP without message passing. This creates a train-test mismatch when using GNN-based encoders for generation. 2. **Homophily Assumption:** The standard inner-product decoder used in VGAEs assumes that nodes with similar embeddings should connect, which is insufficient in modeling the complex connection mechanism of street graphs. 3. **Lack of Global Context:** Predicting an edge (u, v) often requires global geometric context (e.g., barriers and large-scale organization), which independent pairwise scoring without an explicit global conditioning signal fails to capture.

To address these limitations, we propose the Spatial-Context Optimized Prediction of Edges (SCOPE) model. SCOPE is designed to be both locally constrained and globally aware. First, the strong locality in real city networks motivates restricting the candidate edge set to the K nearest neighbors of each node (we use K=20), which substantially reduces inference complexity from $O(N^2)$ to $O(KN)$ while retaining high recall. Second, we utilize a Convolutional Neural Network (CNN) to process the spatial distribution of nodes. The CNN extracts a global context vector from a grid representation of the entire node set V, enabling the predictor to condition on large-scale structure (e.g., grid-like regularity, multi-orientation lattices) and barrier-like empty regions where edges should be discouraged. An ablation in Section V-C shows this global context is essential for reconstructing correct connectivity.

Given the node set V, we score each candidate pair $e(u, v)$ where $R(e) \leq K$. SCOPE outputs a scalar probability $P_\phi(e_{uv} \mid V) \in [0, 1]$. Specifically, we use a ResNet-18 [23] with three channels for the CNN module, operating on binary matrices of shape $2^q \times 2^q$. The first channel forms a global occupancy mask marking all nodes in V, while the remaining two channels mark the positions of u and v respectively. The CNN extracts a context vector g, which is concatenated with the node embeddings $z(u)$ and $z(v)$. The node embedding $z(v_i)$ is the concatenated embeddings of its corresponding coordinate tokens $h(y_i)$ and $h(x_i)$. The fused information is processed by an MLP to yield the final connection probability.

V. Experiments

We first evaluate model performance as a classification problem, and then assess the generative results from both visual and statistical perspectives.

A. Dataset and Setup

Building on the processing pipeline from [16], we filter graphs to 100–800 nodes, normalize them to a unit-diagonal bounding box in $[0, 1]$, and quantize coordinates to 8 bits (q=8) [22]. After removing quantization artifacts, we apply a 92.5/2.5/5 split [22].

The evaluated metrics are extended from [13] and include nearest-neighbor index (NNI), spatial Gini coefficient, average degree, global clustering coefficient, average edge length

TABLE I
EDGE PREDICTION PERFORMANCE ON THE TEST SET.

Methods	AP	ROC-AUC
KNN	0.1215	0.9295
Beta-Skeleton	0.6508	0.9055
VGAE (Inner Prod.)	0.7002	0.9938
VGAE (MLP)	0.8385	0.9972
SCOPE (w/o CNN)	0.6743	0.9874
SCOPE	**0.9138**	**0.9976**

TABLE II
MMD BETWEEN GENERATED AND TEST GRAPH DISTRIBUTIONS

Methods	Deg.	Cluster.	Edge Len.	BC Gini
Universal Model	0.3065	0.6404	0.6391	<u>0.1784</u>
Beta-Skeleton	0.0713	<u>0.4276</u>	**0.0427**	0.4079
VGAE (Inner Prod.)	0.0004	0.6773	0.7458	0.6569
VGAE (MLP)	**0.0001**	0.6885	0.1941	0.6518
SCOPE (w/o CNN)	<u>0.0003</u>	0.6856	0.3451	0.6850
SCOPE	0.0004	**0.2796**	<u>0.0825</u>	**0.0056**

and betweenness centrality (BC) Gini coefficient. Distributional similarity is assessed via Maximum Mean Discrepancy (MMD) [24] between generated and test sets; lower is better. We generate 2000 samples per method for evaluation.

B. Implementation Details

1) Model and Training: Our Transformer implementation follows [22], utilizing 128-dimensional embeddings, 4 attention heads, and 18 decoder layers with a dropout of 0.2. The model supports sequences corresponding to graphs with up to 800 nodes. In SCOPE, we pair two 256-d node embeddings via a three-layer MLP (512, 256, and 1 hidden units; dropout 0.3) and concatenate a global image feature from a ResNet-18 encoder (3 input channels, 256-d output) to score edges. The Transformer is trained with a batch size of 32 using the Adam optimizer (learning rate 3×10^{-4}), cosine annealing with 5k warmup steps, and gradient clipping at 1.0. SCOPE is trained with a batch size of 512, a learning rate of 1×10^{-4}, weight decay of 1×10^{-4}, and a cosine annealing with 10k warmup steps. We employ a 1:1 negative sampling ratio. Both models are trained for 500k steps.

2) Baselines: We use $K = 20$ and $\beta = 1.4$ for KNN and beta-skeleton graphs. The parameters for the Universal Model [13] are set to empirical values on the training set. For autoregressive generation, the probabilistic edge models are discretized with a dynamic threshold by sampling the target degree from the empirical degree distribution $\mathcal{N}(2.918, 0.286)$ determined by the training set.

C. Edge Model Evaluation

We evaluate edge prediction on the held-out test set as a binary classification task. We compare SCOPE against heuristics (KNN and Beta-Skeleton) and two VGAE variants (inner-product vs. MLP decoders) from [16], and include an ablation of SCOPE without the CNN module.

Table I reports Average Precision (AP) and ROC-AUC. Fig. 1 presents qualitative reconstructions; KNN is omitted due to poor visual quality. Overall, SCOPE achieves the highest scores and consistently recovers global street organization, including arterial backbones and barrier-like gaps. Key observations from Table I and Fig. 1 are:

- **Locality alone yields high recall but low precision.** KNN attains a relatively high ROC-AUC but extremely low AP, indicating many false positives at realistic thresholds. Beta-Skeleton improves AP by enforcing planarity, but this assumption ignores real-world non-planar complexities and fails to form long-range arterial connections.
- **Global context is decisive for correct topology.** Removing the CNN causes a large AP drop and degraded reconstruction quality. Qualitatively, both SCOPE w/o CNN and the VGAE variants over-emphasize short links, failing to establish long-range connections. In contrast, SCOPE successfully reconstructs arterial structures and respects sparse regions.
- **Homophily-based decoding is insufficient for street wiring.** The inner-product VGAE underperforms the MLP variant, validating our hypothesis that connectivity is driven by spatial geometry and global constraints rather than simple embedding similarity.

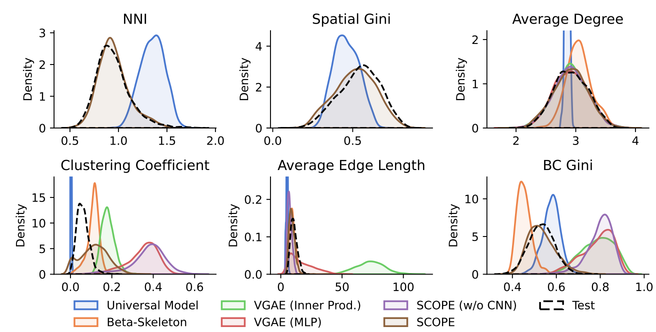

Fig. 3. Distribution of graph statistics for generated vs. real test graphs. SCOPE closely aligns with the empirical distribution across both geometric and topological metrics.

D. Generative Performance

We compose the Transformer node model with each edge model to generate complete graphs, and compare these against the Universal Model [13].

Table II shows the MMD results, with the best method in bold and the second best underscored, while Fig. 3 visualizes the corresponding distributions of graph statistics. Regarding node geometry, the Transformer accurately recovers NNI and Spatial Gini coefficients, successfully capturing the data-driven inhomogeneity of urban point processes that the Universal Model fails to match. For topology, both VGAE and SCOPE employ a dynamic threshold that matches the empirical degree distribution, resulting in closely matching distributions of the average degree. However, SCOPE exhibits stronger alignment with empirical distributions for average edge length and BC Gini, even outperforming the Universal Model, which

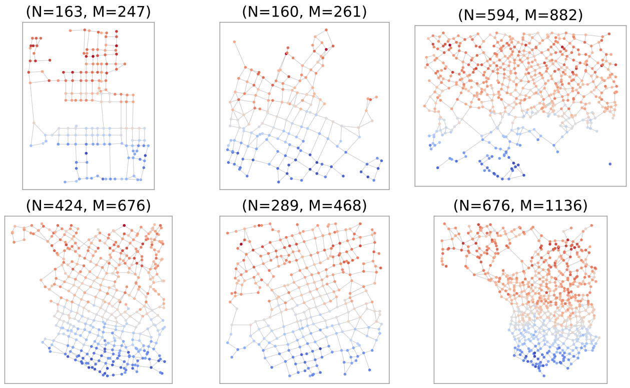

Fig. 4. Diverse synthetic street networks generated by our pipeline. N and M denote node and edge counts. Node colors represent the first principal component of the learned embeddings, showing a smooth pattern spatially.

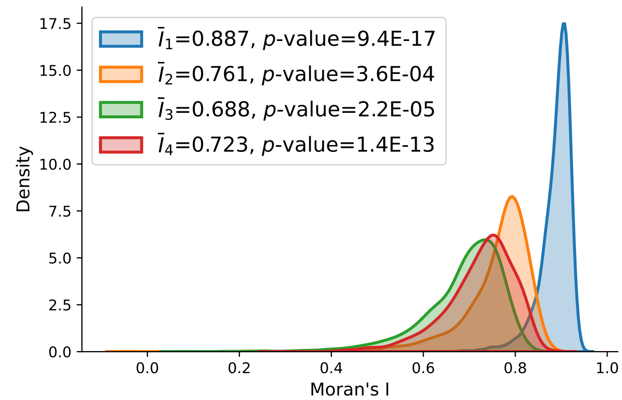

Fig. 5. Spatial autocorrelation (Moran's I) of the principal components of learned node embeddings. High values confirm that the model learns a spatially smooth latent manifold.

is explicitly optimized to match empirical BC heterogeneity. This suggests that SCOPE learns meaningful edge semantics and successfully reconstructs the heterogeneity of road functionality without explicit supervision. Conversely, the Beta-Skeleton baseline performs well on local geometric metrics like clustering and edge length owing to its planar prior, but strictly underestimates BC Gini. This underestimation reflects a fundamental limitation: purely local proximity rules fail to encode global edge semantics such as grade-separated infrastructure, functional hierarchies, and arterial shortcuts, which determine shortest-path behaviors. VGAE variants and SCOPE (w/o CNN) match low-order degree statistics but underperform on higher-order metrics, consistent with the qualitative observation that they lack global coherence.

Fig. 4 illustrates examples of synthetic urban street networks generated by our two-stage pipeline. The model generates diverse graphs varying in scale, node density, orientation, and layout regularity. The synthesized structures are topologically plausible and largely planar. A remaining limitation is the occasional presence of isolated nodes, which could be addressed in future work by incorporating explicit connectivity constraints into the thresholding mechanism.

E. Interpretability of Learned Embeddings

We further investigate the properties of the latent space learned by the Transformer. Moran's I is a standard spatial statistic that quantifies autocorrelation of values over a neighborhood graph (values close to 1 indicate strong positive spatial autocorrelation). We compute Moran's I for the principal components of the node embeddings. Results in Fig. 4 and Fig. 5 show extremely high spatial autocorrelation for the top components. This indicates that the Transformer spontaneously learns a smooth manifold where spatially adjacent nodes share similar latent representations, effectively rediscovering the First Law of Geography purely from data. These spatially coherent embeddings provide a robust local geometric signal to the SCOPE edge predictor, facilitating the learning of distance-dependent connectivity rules even before global context is applied.

VI. Conclusion

We presented a graph-native generative framework for urban street networks that learns from over 35,000 cities. By decoupling generation into an autoregressive geometric stage and a context-aware topological stage, our model captures both local connectivity constraints and global spatial organization, significantly outperforming heuristic and neural baselines on higher-order graph statistics. The learned representations spontaneously encode spatial autocorrelation, confirming that the model discovers meaningful geometric manifolds from data. Future work will extend this probabilistic formulation to support richer attributes and controllable generation, advancing the learnability of complex spatial networks beyond rule-based procedural modeling.

References

[1] H. Xu, F. Omitaomu, S. Sabri, S. Zlatanova, X. Li, and Y. Song, "Leveraging generative ai for urban digital twins: a scoping review on the autonomous generation of urban data, scenarios, designs, and 3d city models for smart city advancement," *Urban Informatics*, vol. 3, no. 1, p. 29, 2024.

[2] M. Moreno-Ibarra, M. Saldaña-Perez, S. Pérez Rodríguez, and E. Juárez Carbajal, "Generative artificial intelligence in the context of urban spaces," in *Telematics and Computing*. Cham: Springer Nature Switzerland, 2024, pp. 209–222.

[3] Q. Wang, Y. Liang, Y. Zheng, K. Xu, J. Zhao, and S. Wang, "Generative ai for urban planning: Synthesizing satellite imagery via diffusion models," *Computers, Environment and Urban Systems*, vol. 122, p. 102339, 2025.

[4] Z. Fang, J. Qi, L. Fan, J. Huang, Y. Jin, and T. Yang, "A framework for human-computer interactive street network design based on a multi-stage deep learning approach," *Computers, Environment and Urban Systems*, vol. 96, p. 101853, 2022.

[5] A. Y. Grinberger, M. Lichter, and D. Felsenstein, *Simulating Urban Resilience: Disasters, Dynamics and (Synthetic) Data*. Cham: Springer International Publishing, 2015, pp. 99–119.

[6] B. Jiang and C. Claramunt, "Topological analysis of urban street networks," *Environment and Planning B: Planning and Design*, vol. 31, no. 1, pp. 151–162, 2004.

[7] S. Porta, P. Crucitti, and V. Latora, "The network analysis of urban streets: A primal approach," *Environment and Planning B: Planning and Design*, vol. 33, no. 5, pp. 705–725, 2006.

[8] S. Hartmann, M. Weinmann, R. Wessel, and R. Klein, "StreetGAN: Towards road network synthesis with generative adversarial networks," in *Proceedings of the International Conference on Computer Graphics, Visualization and Computer Vision*, 2017.

[9] K. Kempinska and R. Murcio, "Modelling urban networks using variational autoencoders," *Appl. Netw. Sci.*, vol. 4, no. 1, p. 114, 2019.

[10] T. Li, R. Li, S. Ye, Z. Zhang, Z. Yin, S. Li, and Z. Pan, "CRTGAN: Controllable road network graphs generation via transformer based gan," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8.

[11] T. Owaki and T. Machida, "RoadNetGAN: Generating road networks in planar graph representation," in *Neural Information Processing*, 2020, pp. 535–543.

[12] H. Chu, D. Li, D. Acuna, A. Kar, M. Shugrina, X. Wei, M.-Y. Liu, A. Torralba, and S. Fidler, "Neural turtle graphics for modeling city road layouts," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[13] M. Barthelemy and G. Boeing, "Universal model of urban street networks," *Phys. Rev. Lett.*, vol. 135, no. 13, p. 137401, 2025.

[14] D. Maniadakis and D. Varoutas, "Fitting planar proximity graphs on real street networks," in *Proceedings of ECCS 2014*, 2016, pp. 11–20.

[15] A. Bojchevski, O. Shchur, D. Zügner, and S. Günnemann, "NetGAN: Generating graphs via random walks," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 2018, pp. 610–619.

[16] M. Neira and R. Murcio, "Graph representation learning on street networks," *ISPRS International Journal of Geo-Information*, vol. 14, no. 8, p. 284, 2025.

[17] A. Cardillo, S. Scellato, V. Latora, and S. Porta, "Structural properties of planar graphs of urban street patterns," *Phys. Rev. E*, vol. 73, no. 6, p. 066107, 2006.

[18] J.-X. Ding, R.-K. Qin, N. Guo, and J.-C. Long, "Urban road network growth model based on rng proximity graph and angle restriction," *Nonlinear Dyn.*, vol. 96, no. 4, pp. 2281–2292, 2019.

[19] A. E. Mireles de Villafranca, R. D. Connors, and R. E. Wilson, "Static traffic assignment on ensembles of synthetic road networks," in *Traffic and Granular Flow '17*, 2019, pp. 29–36.

[20] T. Osaragi and Y. Hiraga, "Street network created by proximity graphs: Its topological structure and travel efficiency," in *Proceedings of the 17th AGILE Conference on Geographic Information Science*, 2014.

[21] G. Boeing, "Planarity and street network representation in urban form analysis," *Environment and Planning B: Urban Analytics and City Science*, vol. 47, no. 5, pp. 855–869, 2020.

[22] C. Nash, Y. Ganin, S. M. A. Eslami, and P. Battaglia, "PolyGen: An autoregressive generative model of 3D meshes," in *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, 2020, pp. 7220–7229.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[24] J. You, R. Ying, X. Ren, W. Hamilton, and J. Leskovec, "GraphRNN: Generating realistic graphs with deep auto-regressive models," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 2018, pp. 5708–5717.