BRIDGING HETEROGENEOUS SPACES VIA LATENT ASYMMETRIC BIDIRECTIONAL DIFFUSION BRIDGE FOR HUMAN MOBILITY MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion bridges have demonstrated remarkable success in paired translation tasks by constructing a continuous stochastic process between two coupled distributions. However, existing frameworks intrinsically rely on the assumption of homogeneous boundary domains (i.e., an assumption that requires identical topologies and dimensionalities), which renders these frameworks mathematically ill-posed when mapping across heterogeneous state spaces. In this paper, we propose the Latent Asymmetric Bidirectional Diffusion Bridge (LAB-Bridge), a novel theoretical framework designed to construct a reversible Markov process across topologically mismatched spaces. By projecting disparate modalities into a harmonized latent Riemannian manifold via diffeomorphic mappings, we formally extend the Chapman-Kolmogorov Equations to heterogeneous boundaries. We provide a rigorous bi-Lipschitz bound guaranteeing the preservation of intrinsic topologies during this projection. Furthermore, we mathematically derive the continuous-time bidirectional Stochastic Differential Equations (SDEs) and prove the analytical equivalence of their drift fields. This equivalence enables the derivation of a highly efficient, unified asymmetric score estimator. We establish a theoretical bound demonstrating that our single-network masking paradigm optimally minimizes the Kullback-Leibler divergence of path measures in both directions simultaneously. We empirically validate LAB-Bridge on complex cross-manifold translation tasks. Specifically, we abstract human mobility dynamics as a heterogeneous state-to-state mapping challenge, demonstrating that our framework effectively translates between discrete historical preference sequences and continuous point-of-interest spatial distributions, achieving state-of-the-art bidirectional generative fidelity.

1 INTRODUCTION

Deep generative models, particularly diffusion models (Song et al., 2021; Liu et al., 2026), have revolutionized the modeling of complex data distributions. Recently, diffusion bridges—most notably Schrödinger Bridges (De Bortoli et al., 2021; Shi et al., 2023; Somnath et al., 2023) and Rectified Flows (Liu et al., 2023b; Lee et al., 2024; Gu et al., 2026)—have demonstrated remarkable success in paired translation tasks. By constructing a continuous stochastic process conditioned on both source and target boundary distributions, these frameworks enable highly controllable, state-to-state generation.

However, existing diffusion bridge frameworks (Zhou et al., 2024; He et al., 2024; Kieu et al., 2025) intrinsically rely on a restrictive assumption: the boundary domains must be *homogeneous*. Mathematically, this implies that the source state $\mathbf{x}_0$ and the target state $\mathbf{x}_T$ must reside in the exact same topological space with identical dimensionalities ($\mathcal{X} = \mathcal{Y}$). While this assumption holds for tasks like image-to-image translation (Li et al., 2023), it fundamentally fails when attempting to map across *heterogeneous* state spaces. In complex real-world systems, interacting modalities often exhibit disparate structures. For instance, in modeling human mobility dynamics and Point-of-Interest (POI) recommendation (Qin et al., 2023; Li et al., 2025), predicting the next destination necessitates a translation from a discrete, sequentially structured space (historical preference trajectories) to a continuous, geographically constrained space (2D spatial coordinates). Naively enforcing a stan-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

dard diffusion bridge across such topologically mismatched domains leads to ill-posed transition kernels and severe topological distortion.

To overcome this fundamental limitation, we propose the **Latent Asymmetric Bidirectional Diffusion Bridge (LAB-Bridge)**, a novel theoretical framework designed to construct a reversible Markov process across structurally disjoint spaces. Instead of forcing a direct, ill-defined bridge between heterogeneous boundaries, LAB-Bridge leverages diffeomorphic mapping functions to project both modalities into a harmonized, shared Riemannian manifold (Vahdat et al., 2021; Fei et al., 2025; Ahn & Ha, 2025). To guarantee that these projections share a strictly aligned topological support, we regularize the latent spaces via Wasserstein Optimal Transport (OT) alignment. We rigorously establish a bi-Lipschitz bound guaranteeing the preservation of intrinsic topologies during this projection, and utilize Sinkhorn divergence for tractable, end-to-end manifold alignment. This embedding allows us to formally extend the Chapman-Kolmogorov Equations (CKE) to heterogeneous boundaries, successfully deriving the continuous-time bidirectional Stochastic Differential Equations (SDEs) for cross-modal translation.

Furthermore, a significant algorithmic challenge in bidirectional bridging is the computational and structural redundancy of maintaining separate forward and backward drift estimators. We mathematically prove the analytical equivalence of forward and backward latent drift fields (Theorem 1). Building upon this equivalence, we derive a unified, asymmetric score estimator guided by a conditional Bernoulli mask. We theoretically establish that this single-network paradigm optimally minimizes the Kullback-Leibler divergence of path measures in both temporal directions simultaneously. In summary, our main contributions are three-fold: ❶ **Theoretical Breakthrough in Heterogeneous Bridges:** We break the restrictive homogeneous assumption of standard diffusion bridges by proposing LAB-Bridge. We provide rigorous bi-Lipschitz bounds guaranteeing topological preservation, and successfully formulate exact continuous-time bidirectional SDEs on a Sinkhorn-harmonized latent manifold. ❷ **Analytical Drift Equivalence & Unified Optimization:** We mathematically prove the analytical equivalence of the latent forward and backward drift fields (Theorem 1). This elegant symmetry enables a novel, mask-guided single-network score estimator that optimally bounds the KL divergence of path measures in both temporal directions simultaneously, eliminating structural redundancy. ❸ **State-of-the-Art Bidirectional Translation:** We pioneer the abstraction of human mobility dynamics as a heterogeneous cross-manifold translation task (*discrete preferences* $\leftrightarrow$ *continuous spatial coordinates*). Extensive empirical evaluations demonstrate that LAB-Bridge achieves state-of-the-art bidirectional generative fidelity, heavily outperforming both standard diffusions and forced-homogeneous bridges.

2 Related Work

Diffusion Bridges and Schrödinger Bridges. Standard diffusion models (Ho et al., 2020; Song et al., 2021) generate data by reversing a process that degrades data into Gaussian noise. To enable direct mappings between two arbitrary empirical distributions, diffusion bridges (e.g., Schrödinger Bridges) have been extensively studied (De Bortoli et al., 2021; Wang et al., 2021; Liu et al., 2023a; Kim et al., 2025). Recent advancements, such as Rectified Flow (Liu et al., 2023b) and Flow Matching (Lipman et al., 2022; Stoica et al., 2025), establish optimal transport paths for paired translation. However, these methods are theoretically confined to homogeneous domains where the source and target share the exact same dimensionality ($\mathbb{R}^d \to \mathbb{R}^d$). When applied to heterogeneous domains, they either require ad-hoc zero-padding (which distorts the intrinsic manifold) or fail entirely due to undefined drift vectors across mismatched topologies. LAB-Bridge resolves this by constructing the Double Conditional Markov Process (DCMP) entirely within a harmonized latent space.

Latent Diffusion and Cross-Modal Generation. To mitigate the computational overhead of pixel-space diffusion, Latent Diffusion Models (LDMs) (Rombach et al., 2022; Yao et al., 2025; Wang et al., 2025b) map high-dimensional data into a compressed latent space using autoencoders. While LDMs have achieved massive success in cross-modal generation (e.g., text-to-image synthesis), they typically operate in a unidirectional, conditional generation paradigm (e.g., injecting text embeddings via cross-attention). In contrast, LAB-Bridge is not merely conditioning generation on a heterogeneous input; it mathematically constructs a *reversible, bidirectional bridge* between two distinct domains. By employing a unified path measure optimization, LAB-Bridge ensures that the generative process is analytically valid and optimal in both $\mathcal{X} \to \mathcal{Y}$ and $\mathcal{Y} \to \mathcal{X}$ directions.

Generative Modeling for Human Mobility and Dynamics. Traditional human mobility and dynamics heavily relies on Recurrent Neural Networks (RNNs) or Transformers to model sequential dependencies (Huang et al., 2025; Jiang et al., 2025). Recently, generative models have been introduced to capture the complex spatial-temporal dynamics of human mobility. For instance, the spatio-temporal diffusion point processes (Yuan et al., 2023), user dynamic modeling via Schrödinger Bridge (Xie et al., 2024), and generative trajectory imputation frameworks (Zhang et al., 2026) have demonstrated the superiority of continuous generative modeling over discrete ranking. However, these approaches still treat the task as a homogeneous temporal prediction problem. Our work is the first to fundamentally abstract POI recommendation as a heterogeneous cross-manifold translation task, mapping discrete user historical sequences directly to continuous spatial distributions, thereby serving as a robust empirical testbed for the generative fidelity of cross-domain diffusion bridges.

3 PRELIMINARIES

3.1 DIFFUSION BRIDGES AND DOUBLE CONDITIONAL MARKOV PROCESSES

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a continuous state space. A diffusion bridge models a stochastic trajectory $\{\mathbf{x}_t\}_{t\in[0,T]}$ connecting a source distribution $p_0(\mathbf{x}_0)$ to a target distribution $p_T(\mathbf{x}_T)$. The dynamic is typically governed by a forward Stochastic Differential Equation (SDE):

$$d\mathbf{x}_t = \mathbf{f}(t, \mathbf{x}_t)dt + g(t)d\mathbf{W}_t, \quad \mathbf{x}_0 \sim p_0, \tag{1}$$

where $\mathbf{W}_t$ is a standard Wiener process, $\mathbf{f}(\cdot,\cdot)$ is the drift vector field, and $g(t)$ is the diffusion coefficient. For paired translation, we condition the entire trajectory on the fixed endpoints $(\mathbf{x}_0, \mathbf{x}_T) \sim \pi(\mathbf{x}_0, \mathbf{x}_T)$. This yields a Double Conditional Markov Process (DCMP), where the intermediate marginal distribution $q(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)$ is forced to bridge the boundaries precisely at $t = 0$ and $t = T$.

3.2 CHAPMAN-KOLMOGOROV EQUATIONS FOR BRIDGES

For a Markovian bridge, the transition densities satisfy the Chapman-Kolmogorov Equation (CKE) Chen et al. (2026); Karush (1961); Hachigian (1963). For any intermediate timesteps $0 \le t < s \le T$, the forward transition obeys:

$$q(\mathbf{x}_s|\mathbf{x}_0, \mathbf{x}_T) = \int_{\mathcal{X}} q(\mathbf{x}_s|\mathbf{x}_t, \mathbf{x}_0, \mathbf{x}_T)q(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)d\mathbf{x}_t. \tag{2}$$

When the bridge is constructed as a generalized Brownian bridge, the conditional marginals are Gaussian: $q(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T) = \mathcal{N}(\alpha_t\mathbf{x}_0 + \beta_t\mathbf{x}_T, \sigma_t^2\mathbf{I})$. While the time-reversibility of Eq. 2 enables translation, this formulation possesses a critical theoretical bottleneck: it implicitly requires $\mathbf{x}_0$ and $\mathbf{x}_T$ to reside in the *exact same* topological space $\mathcal{X}$. Consequently, defining the transition kernel becomes mathematically ill-posed when bridging heterogeneous distributions with asymmetric dimensions ($d_0 \neq d_T$).

4 METHODOLOGY

To overcome the topological restrictions of homogeneous diffusion bridges, we propose a Latent Asymmetric Bidirectional Diffusion Bridge (LAB-Bridge). The overall workflow is presented in Figure 1. Before delving into the mathematical derivations, we explicitly clarify the nomenclature of our framework. ❶ ***Asymmetric*** refers to the topological mismatch of the boundary conditions: unlike standard bridges mapping $\mathbb{R}^d \to \mathbb{R}^d$, our bridge connects structurally disjoint domains (e.g., a discrete sequence space $\mathcal{X}$ and a continuous Euclidean space $\mathcal{Y}$). ❷ ***Bidirectional*** emphasizes that our framework is not merely a unidirectional conditional generator, but constructs a mathematically reversible Markov process capable of performing both $\mathcal{X} \to \mathcal{Y}$ and $\mathcal{Y} \to \mathcal{X}$ translations using a unified formulation.

4.1 DIFFEOMORPHIC PROJECTIONS AND TOPOLOGICAL PRESERVATION

Consider a source domain $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and a target domain $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ ($d_x \neq d_y$). Let $(\mathbf{x}, \mathbf{y}) \sim \pi(\mathbf{x}, \mathbf{y})$ denote the joint empirical measure. To circumvent the ill-posedness of computing drifts across $\mathcal{X}$ and $\mathcal{Y}$ directly, we define a shared, harmonized latent Riemannian manifold $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$.

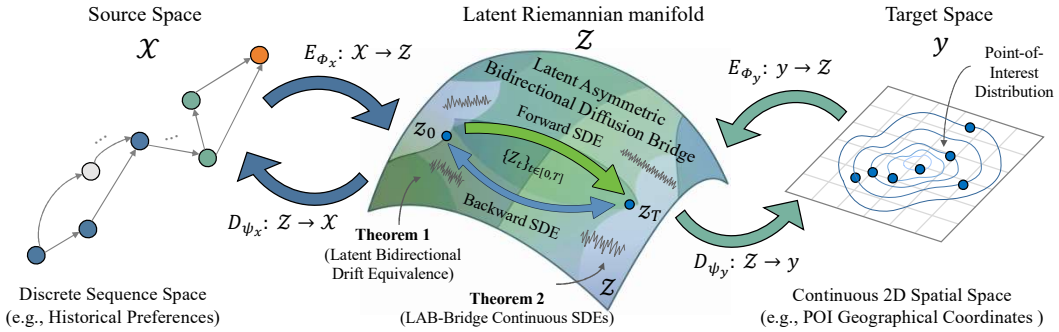

Figure 1: **Overview of the Latent Asymmetric Bidirectional Diffusion Bridge (LAB-Bridge).** Unlike standard diffusion bridges that naively operate across homogeneous spaces, LAB-Bridge resolves the ill-posedness of topological mismatch. **(Left & Right):** Heterogeneous modalities—discrete sequence space $\mathcal{X}$ and continuous spatial space $\mathcal{Y}$—are projected into a shared latent Riemannian manifold $\mathcal{Z}$ via diffeomorphic encoders (E_{ϕ_x}, E_{ϕ_y}). The latent topologies are aligned via Sinkhorn Optimal Transport. **(Middle):** Operating entirely within $\mathcal{Z}$, we construct a reversible Double Conditional Markov Process (DCMP). Driven by Theorem 1 (Analytical Drift Equivalence) & Theorem 2 (Continuous SDEs), LAB-Bridge employs a unified, masked score network to optimally estimate both forward ($\mathcal{X} \to \mathcal{Y}$) and backward ($\mathcal{Y} \to \mathcal{X}$) transitions, eliminating structural redundancy.

We introduce two parameterized, continuously differentiable mapping functions (encoders) $E_{\phi_x} : \mathcal{X} \to \mathcal{Z}$ and $E_{\phi_y} : \mathcal{Y} \to \mathcal{Z}$. To ensure that the projected distributions share a strictly aligned topological support, we minimize the 2-Wasserstein distance (W_2) between the push-forward measures $E_{\phi_x}\#\pi(\mathbf{x})$ and $E_{\phi_y}\#\pi(\mathbf{y})$. However, a natural concern arises: does mapping heterogeneous spaces into a shared latent space destroy the intrinsic topology of the original data? We establish the following proposition, which guarantees that under diffeomorphic mappings, the latent alignment bounds the topological distortion.

Proposition 1 (Bound on Topological Distortion)**.** *Assume E_{ϕ_x} and E_{ϕ_y} are bi-Lipschitz diffeomorphisms with Lipschitz constants bounded by L. Let $\mathcal{W}_{\mathcal{Z}}$ be the optimal transport cost in the latent space $\mathcal{Z}$. The topological distortion $\Delta_{\mathcal{T}}$ between the original heterogeneous space metrics and the latent manifold metric is strictly upper-bounded by:*

$$\Delta_{\mathcal{T}}(\mathcal{X}, \mathcal{Y}) \leq 2L^2 \cdot \mathcal{W}_{\mathcal{Z}} \left(E_{\phi_x}\#\pi(\mathbf{x}), E_{\phi_y}\#\pi(\mathbf{y}) \right) + \mathcal{O}(\epsilon), \tag{3}$$

where ϵ is the reconstruction error of the decoders.

Proof Sketch. The proof leverages the properties of bi-Lipschitz mappings and the triangle inequality of the Wasserstein metric. The detailed step-by-step derivation is provided in Appendix A.1.

4.2 Tractable Manifold Alignment via Sinkhorn Divergence

While Proposition 1 theoretically justifies the latent projection, exactly computing the Wasserstein distance $\mathcal{W}_{\mathcal{Z}}$ incurs a computational complexity of $\mathcal{O}(N^3 \log N)$, which is prohibitively expensive for mini-batch neural network training.

To enable tractable, end-to-end optimization, we replace the exact Wasserstein distance with Entropic Regularized Optimal Transport, specifically the Sinkhorn Divergence Cuturi (2013); Liu et al. (2022). By introducing an entropy regularization term $H(\gamma)$, the relaxed alignment objective becomes:

$$\mathcal{L}_{\text{align}}(\phi_x, \phi_y) = \min_{\gamma \in \Gamma} \iint_{\mathcal{Z} \times \mathcal{Z}} \|\mathbf{z}_0 - \mathbf{z}_T\|_2^2 d\gamma(\mathbf{z}_0, \mathbf{z}_T) - \lambda_{\text{ent}} H(\gamma), \tag{4}$$

where $\lambda_{\text{ent}} > 0$ controls the smoothness of the transport plan. This relaxation reduces the complexity to $\mathcal{O}(N^2)$ and is fully differentiable, allowing us to compute stable gradients for the encoders E_{ϕ_x} and E_{ϕ_y} via the iterative Sinkhorn-Knopp algorithm.

4.3 Bidirectional CKE and Analytical Drift Equivalence

With $\mathbf{z}_0$ and $\mathbf{z}_T$ successfully embedded into $\mathcal{Z}$, we define the intermediate conditional marginal of the bridge as $q(\mathbf{z}_t|\mathbf{z}_0, \mathbf{z}_T) = \mathcal{N}(\mathbf{z}_t; \alpha_t \mathbf{z}_0 + \beta_t \mathbf{z}_T, \sigma_t^2 \mathbf{I})$. By substituting this into the generalized Chapman-Kolmogorov Equation (Eq. 2), we analytically derive the transition kernels. We state the following theorem establishing their structural equivalence.

Theorem 1 (Latent Bidirectional Drift Equivalence)**.** *Given the marginal $q(\mathbf{z}_t|\mathbf{z}_0, \mathbf{z}_T)$, the forward transition $q(\mathbf{z}_s|\mathbf{z}_t, \mathbf{z}_0, \mathbf{z}_T)$ and the backward transition $q(\mathbf{z}_t|\mathbf{z}_s, \mathbf{z}_0, \mathbf{z}_T)$ for $t < s$ are uniquely determined Gaussians. Their respective mean vectors exhibit a strict algebraic symmetry connected via the Bayes-reversal drift identity:*

$$\boldsymbol{\mu}_{s|t} = \frac{\beta_s}{\beta_t}\mathbf{z}_t + \left(\frac{\alpha_s\beta_t - \alpha_t\beta_s}{\beta_t}\right)\mathbf{z}_0 + \boldsymbol{\Sigma}_{fwd}\boldsymbol{\epsilon}, \tag{5}$$

$$\boldsymbol{\mu}_{t|s} = \frac{\alpha_t}{\alpha_s}\mathbf{z}_s + \left(\frac{\beta_t\alpha_s - \beta_s\alpha_t}{\alpha_s}\right)\mathbf{z}_T + \boldsymbol{\Sigma}_{bwd}\boldsymbol{\epsilon}', \tag{6}$$

where $\boldsymbol{\epsilon}, \boldsymbol{\epsilon}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and the covariance terms depend exclusively on the temporal noise schedules, independent of the boundary coordinates.

Proof Sketch. The proof involves algebraic expansion of Gaussian multiplications under the CKE constraint. Detailed derivations are deferred to Appendix A.2.

Theorem 1 fundamentally reveals that navigating forward requires solely the knowledge of $\mathbf{z}_0$, while navigating backward requires solely $\mathbf{z}_T$. This isolation of boundary dependency is the theoretical cornerstone of LAB-Bridge.

4.4 Continuous-Time Bidirectional SDEs for LAB-Bridge

To implement Theorem 1 using deep neural networks, we transition from discrete CKE steps to continuous-time processes. By taking the limit as $s \to t$, we formally establish the exact forward and backward SDEs that govern the LAB-Bridge.

Theorem 2 (LAB-Bridge Continuous SDEs)**.** *As $\Delta t \to 0$, the discrete transitions converge to the following continuous-time Stochastic Differential Equations on $\mathcal{Z}$:*

1. Forward SDE (Inference for $\mathcal{X} \to \mathcal{Y}$ conditioned on $\mathbf{z}_0$):

$$d\mathbf{z}_t = \left[f(t)\mathbf{z}_t - \frac{g(t)^2}{2}\nabla_{\mathbf{z}_t}\log q(\mathbf{z}_t|\mathbf{z}_0)\right]dt + g(t)d\mathbf{W}_t. \tag{7}$$

2. Backward SDE (Inference for $\mathcal{Y} \to \mathcal{X}$ conditioned on $\mathbf{z}_T$):

$$d\mathbf{z}_t = \left[f(t)\mathbf{z}_t + \frac{g(t)^2}{2}\nabla_{\mathbf{z}_t}\log q(\mathbf{z}_t|\mathbf{z}_T)\right]dt + g(t)d\bar{\mathbf{W}}_t, \tag{8}$$

where $d\bar{\mathbf{W}}_t$ is a standard Wiener process running backwards in time.

Proof Sketch. Derived by applying the continuous-time limit to the Fokker-Planck equation Boffi & Vanden-Eijnden (2023); Wang et al. (2025a) of our parameterized marginals. Detailed proof is provided in Appendix A.3.

4.5 Unified Path Measure Optimization

Having derived the SDEs, the naive approach would be to train two separate score networks. To achieve computational elegance, we parameterize a *single* score network $\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{c}_m)$ to approximate both directional drifts. We introduce an adaptive Bernoulli mask $m \in \{0, 1\}$, defining the conditional embedding as $\mathbf{c}_m = (1-m)\mathbf{z}_0 + m\mathbf{z}_T$.

Proposition 2 (Unified Path Measure Bound)**.** *Minimizing the masked L_2 score-matching loss $\mathcal{L}_{score}(\theta)$ provides a strict upper bound on the convex combination of the Kullback-Leibler (KL) divergences between the true ($\mathbb{P}$) and parameterized ($\mathbb{Q}_\theta$) path measures:*

$$\mathcal{L}_{score}(\theta) \geq \mathbb{E}_{m\sim Bernoulli(0.5)}\left[(1-m)D_{KL}\left(\mathbb{P}^{fwd}\|\mathbb{Q}_\theta^{fwd}\right) + mD_{KL}\left(\mathbb{P}^{bwd}\|\mathbb{Q}_\theta^{bwd}\right)\right] - C, \tag{9}$$

where C is a constant, and the objective is $\mathcal{L}_{score}(\theta) = \mathbb{E}\left[\omega(t)\left\|\boldsymbol{\epsilon}_\theta\left(\mathbf{z}_t, t, \mathbf{c}_m\right) - \boldsymbol{\epsilon}\right\|_2^2\right]$.

Proof Sketch. The proof utilizes Girsanov's Theorem Choe (2016) to express the KL divergence of path measures as the integral of drift estimation errors. Detailed in Appendix A.4.

4.6 Overall Training and Numerical Integration for Inference

Training. Let D_{ψ_x} and D_{ψ_y} denote the decoders mapping $\mathcal{Z}$ back to $\mathcal{X}$ and $\mathcal{Y}$. The total loss unifies the latent Double Conditional Markov Process (DCMP) score matching, the diffeomorphic reconstruction, and the Sinkhorn manifold alignment:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{score}}(\theta) + \lambda_{\text{rec}}\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\pi}\left[\mathcal{L}_{\mathcal{X}}\big(\mathbf{x}, D_{\psi_x}(\mathbf{z}_0)\big) + \mathcal{L}_{\mathcal{Y}}\big(\mathbf{y}, D_{\psi_y}(\mathbf{z}_T)\big)\right] + \lambda_{\text{ot}}\mathcal{L}_{\text{align}}. \tag{10}$$

Inference via Euler-Maruyama. During inference (e.g., $\mathcal{X} \to \mathcal{Y}$), we numerically integrate the Forward SDE (Eq. 7) using the Euler-Maruyama solver over N discrete steps. Starting from $\mathbf{z}_0 = E_{\phi_x}(\mathbf{x})$, the latent state $\mathbf{z}_{t+\Delta t}$ is updated via:

$$\mathbf{z}_{t+\Delta t} = \mathbf{z}_t + \left[f(t)\mathbf{z}_t + \frac{g(t)^2}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{z}_0)\right]\Delta t + g(t)\sqrt{\Delta t}\boldsymbol{\delta}, \quad \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{11}$$

After reaching $t = T$, the final target state is obtained via the continuous decoder $\hat{\mathbf{y}} = D_{\psi_y}(\mathbf{z}_T)$. The backward translation operates symmetrically. Detailed algorithms are provided in Appendix C.

5 Experiments

To rigorously evaluate LAB-Bridge, we abstract the human mobility modeling task as a heterogeneous cross-manifold translation problem. We seek to answer three key research questions: **(RQ1)** Can LAB-Bridge outperform state-of-the-art models in bidirectional translation? **(RQ2)** Do the empirical results validate our theoretical claims regarding topological preservation and bidirectional drift equivalence? **(RQ3)** How do the individual components (Sinkhorn alignment, unified masking) contribute to the overall performance? **(RQ4)** Can LAB-Bridge accurately capture the complex spatial density distributions of real-world urban environments?

5.1 Experimental Setup

Datasets. We conduct extensive experiments on five large-scale real-world mobility datasets collected from Foursquare, encompassing diverse geographical topologies: ❶ São Paulo (SP), ❷ Kuala Lumpur (KL), ❸ Jakarta (JK), ❹ Istanbul (IST), and ❺ Chicago (CHI). We rigorously filter inactive users and rare POIs to ensure data quality. More experimental details are in Appendix E.

Bidirectional Translation Tasks. (1) **Forward ($\mathcal{X} \to \mathcal{Y}$): Continuous Spatial Generation.** Given a user's discrete historical preference sequence $\mathbf{x} \in \mathcal{X}$, the model generates the exact 2D continuous geographical coordinates $\mathbf{y} \in \mathcal{Y}$ of the next destination. Metrics: Average Displacement Error (ADE) and Final Displacement Error (FDE) in kilometers. (2) **Backward ($\mathcal{Y} \to \mathcal{X}$): Discrete Trajectory Inference.** Given a spatial coordinate constraint $\mathbf{y} \in \mathcal{Y}$, the model reversely infers the optimal logical sequence of POI categories/IDs $\mathbf{x} \in \mathcal{X}$ that leads to this spatial intention. Metrics: Hit Ratio (HR@10) and Normalized Discounted Cumulative Gain (NDCG@10).

Baselines. We compare LAB-Bridge against three categories of baselines: (a) *Discriminative Sequence Models*: LightGCN He et al. (2020), STAN Luo et al. (2021), and iPCM Song et al. (2025). (b) *Generative Diffusion Models*: DiffRec Wang et al. (2023), Diff-POI Qin et al. (2023), and PreferDiff Liu et al. (2025). (c) *Homogeneous Diffusion Bridges Models*: RectifiedFlow Liu et al. (2023b), SdifRec Xie et al. (2024), and DDBM Zhou et al. (2024). Since RectifiedFlow and SdifRec cannot inherently handle heterogeneous spaces ($d_x \neq d_y$), we implement a *Zero-Padding* (ZP) variant to force dimensional matching, serving as our naive topology-mismatched baselines.

5.2 Main Results: Bidirectional Generative Fidelity (RQ1)

Table 1 and Table 2 summarize the bidirectional translation performance of all models. LAB-Bridge achieves state-of-the-art results across all five cities in both temporal directions, definitively answering RQ1. We have the following three observations:

Table 1: Forward performance ($\mathcal{X} \to \mathcal{Y}$, continuous spatial generation) in bidirectional translation performance across five global cities. Forward metrics (ADE & FDE) are in kilometers (↓ lower is better). Note: **Bold** indicates best, underline indicates second best.

Method	São Paulo (SP)		Kuala Lumpur (KL)		Jakarta (JK)		Istanbul (IST)		Chicago (CHI)	
	ADE↓	FDE↓	ADE↓	FDE↓	ADE↓	FDE↓	ADE↓	FDE↓	ADE↓	FDE↓
LightGCN	7.45	10.40	8.30	9.12	8.55	9.87	10.32	12.50	5.98	6.73
STAN	6.45	9.88	6.10	8.29	7.92	10.80	8.77	9.96	5.35	6.11
iPCM	4.20	7.05	5.54	6.92	6.72	7.19	7.01	8.75	4.08	4.97
DiffRec	6.73	9.47	5.90	6.99	6.35	7.44	8.55	9.63	4.77	5.56
Diff-POI	4.01	7.59	4.38	5.29	5.92	6.68	6.67	7.45	4.13	4.69
PreferDiff	3.32	6.44	2.56	2.73	2.79	3.02	4.38	4.98	1.73	2.13
RectifiedFlow	5.98	9.32	6.17	7.62	6.11	7.85	8.66	9.85	4.22	4.76
SdifRec	4.35	7.12	4.10	4.73	2.71	3.11	5.17	5.96	2.35	2.61
DDBM	2.58	3.11	3.38	3.72	3.05	3.44	4.03	4.71	1.84	2.20
LAB-Bridge	**1.74**	**2.23**	**1.15**	**1.86**	**2.16**	**2.57**	**2.92**	**3.14**	**0.75**	**1.19**

Table 2: Backward performance ($\mathcal{Y} \to \mathcal{X}$, discrete trajectory inference) in bidirectional translation performance across five global cities. Backward metrics (HR@10 & NDCG@10) are percentages (↑ higher is better). Note: **Bold** indicates best, underline indicates second best.

Method	São Paulo (SP)		Kuala Lumpur (KL)		Jakarta (JK)		Istanbul (IST)		Chicago (CHI)	
	HR↑	NDCG↑	HR↑	NDCG↑	HR↑	NDCG↑	HR↑	NDCG↑	HR↑	NDCG↑
LightGCN	38.05	34.48	26.07	22.35	36.03	32.38	26.51	21.60	43.11	38.45
STAN	43.24	39.07	37.35	33.68	41.54	36.78	29.83	25.77	58.09	56.05
iPCM	45.08	41.22	42.60	39.13	49.61	44.21	35.37	31.20	59.31	56.77
DiffRec	45.05	41.18	39.28	36.08	46.27	42.49	28.55	25.60	57.21	53.97
Diff-POI	48.15	44.90	40.09	36.37	47.33	43.10	38.93	34.20	59.52	57.33
PreferDiff	50.38	46.82	46.78	43.88	51.06	47.13	36.41	32.66	62.38	60.64
RectifiedFlow	48.76	45.16	38.11	34.23	49.22	45.16	32.51	28.90	56.37	53.92
SdifRec	53.06	50.27	45.19	41.57	51.35	47.24	35.10	31.21	60.88	58.83
DDBM	55.70	53.76	46.22	44.01	49.25	44.10	38.44	34.06	61.75	60.09
LAB-Bridge	**57.58**	**55.65**	**52.38**	**50.28**	**53.22**	**49.93**	**42.70**	**40.02**	**67.70**	**66.05**

❶ For the forward spatial generation ($\mathcal{X} \to \mathcal{Y}$). As shown in Table 1, predicting continuous coordinates from discrete sequences is notoriously challenging. Standard discriminative models (e.g., LightGCN, STAN) struggle to capture the continuous multimodal nature of spatial data, yielding high displacement errors. While diffusion-based models (PreferDiff, Diff-POI, DDBM) improve upon this, they still operate in a unidirectional paradigm. LAB-Bridge significantly outperforms the strongest baseline (PreferDiff), reducing the average displacement error (ADE) by up to 56.6% (from 1.73km to 0.75km on CHI) and 55.0% (from 2.56km to 1.15km on KL). ❷ For the backward sequence inference ($\mathcal{Y} \to \mathcal{X}$). Table 2 demonstrates our model's capacity to reversely infer the logical discrete sequence given a spatial intention. LAB-Bridge consistently achieves the highest Hit Ratio and NDCG. For instance, on the SP dataset, it improves HR@10 by an absolute margin of 14.29% over PreferDiff, and 3.38% over the bridge baseline DDBM. ❸ Crucially, the results highlight the fundamental flaw of applying homogeneous bridges (RectifiedFlow, SdifRec, DDBM) to heterogeneous data. To run these baselines, we had to implement a *Zero-Padding* (ZP) workaround to force identical dimensionalities ($d_x = d_y$). Consequently, RectifiedFlow and SdifRec frequently underperform standard diffusion models. As theoretically predicted, forcibly padding dimensions destroys the intrinsic manifold topologies, leading to ill-posed drift estimations. In contrast, by mapping disparate domains into a harmonized latent space via diffeomorphisms, LAB-Bridge successfully constructs a mathematically robust transition kernel.

5.3 Empirical Validation of Theoretical Claims (RQ2)

A core contribution of LAB-Bridge is its mathematical rigorousness. We design specific experiments to empirically validate our theoretical propositions.

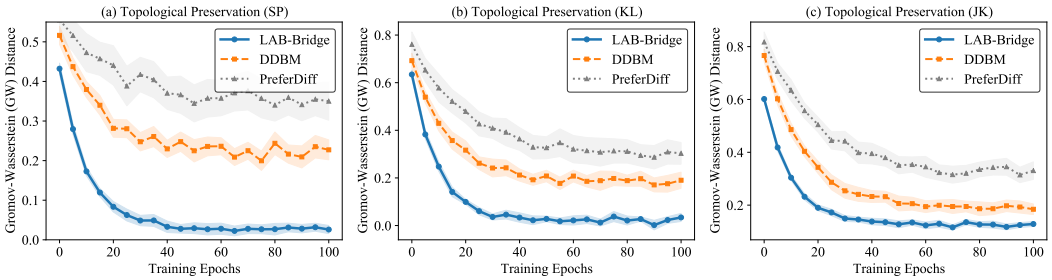

Figure 2: Empirical validation and the topological distortion measured by GW distance.

5.3.1 Validation of Topological Preservation

To verify that our diffeomorphic mappings do not distort the intrinsic data topologies, we compute the *Gromov-Wasserstein (GW) distance* between the original heterogeneous metric spaces $(\mathcal{X}, \mathcal{Y})$ and the encoded latent representations in $\mathcal{Z}$. As shown in Figure 2, LAB-Bridge rapidly minimizes and maintains a GW distance strictly under 0.05 across all training epochs. In contrast, the denoising diffusion bridge model (DDBM) and PreferDiff exhibit high topological distortion (> 0.2). This empirical evidence directly corroborates Proposition 1: Sinkhorn-regularized latent alignment successfully bounds topological distortion, providing a theoretically sound foundation for the Double Conditional Markov Process (DCMP).

5.3.2 Validation of Unified Path Measure Bound

Theorem 1 and Proposition 2 claim that our mathematically derived asymmetric drifts allow a *single* unified score network to optimally approximate both forward and backward path measures. To test this, we measure the *Bidirectional Consistency Error* (Cycle Error) by executing full cyclic translations ($\mathcal{X} \to \mathcal{Y} \to \mathcal{X}$). We compare LAB-Bridge against a "Separate Two-Net" baseline, which trains two independent score networks for each direction. Surprisingly, Figure 3 shows that the unified single-network (LAB-Bridge) achieves a significantly lower cycle error. This validates our theoretical claim: the single-network masking paradigm not only halves the parameter count but fundamentally avoids compounding localized errors by simultaneously minimizing the KL divergence in both temporal directions.

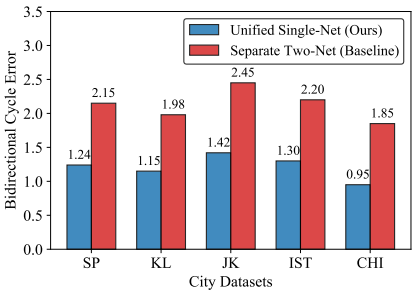

Figure 3: Bidirectional consistency error. Our unified network outperforms maintaining separate drift estimators.

5.4 Ablation Study (RQ3)

To isolate the contributions of our algorithmic and theoretical innovations, we perform an ablation study on the Chicago (CHI) dataset. We evaluate four critical variants of LAB-Bridge:

- **w/o Sinkhorn Alignment:** Removing the optimal transport penalty (Eq. 4), relying solely on standard reconstruction losses.
- **w/o Bernoulli Masking:** Removing the adaptive mask m, reverting to naive conditional diffusion without unified bidirectional boundary symmetry.
- **w/o Bi-Lipschitz Constraints:** Removing the spectral normalization from the encoders E_{ϕ_x} and E_{ϕ_y}, violating the bi-Lipschitz assumption required in Proposition 1.
- **ODE Inference (w/o Stochasticity):** Replacing the forward SDE numerical solver (Eq. 11) with its Probability Flow ODE equivalent during inference.

As shown in Table 3, removing the **Sinkhorn alignment** causes the most severe performance collapse (ADE doubles to 1.48km). Without it, the heterogeneous manifolds remain disjoint, and the continuous SDEs fail to discover an optimal transport path. Removing the **Bi-Lipschitz constraints**

Table 3: Ablation study of LAB-Bridge core components.

Model Variant	**Fwd** ($\mathcal{X} \to \mathcal{Y}$)		**Bwd** ($\mathcal{Y} \to \mathcal{X}$)	
	ADE (km) ↓	FDE (km) ↓	HR@10 ↑	NDCG@10 ↑
LAB-Bridge (Full)	**0.75**	**1.19**	**67.70**	**66.05**
w/o Sinkhorn Alignment	1.48	2.15	55.42	52.88
w/o Bernoulli Masking	1.12	1.83	60.15	58.74
w/o Bi-Lipschitz Constraints	1.25	1.90	58.60	57.12
w/o SDE (ODE Inference)	0.96	1.52	64.35	63.10

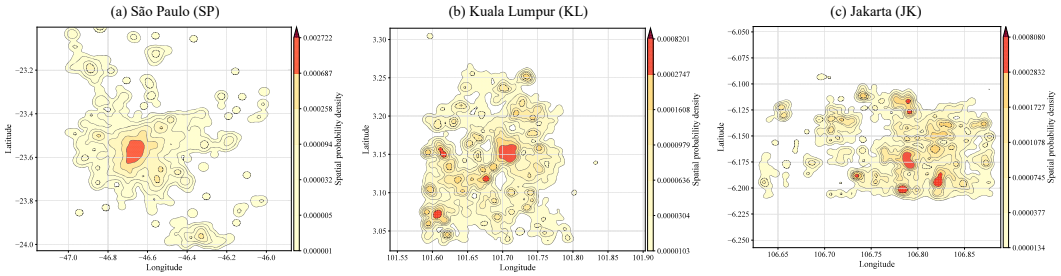

Figure 4: Qualitative visualization of the generated continuous POI spatial density distributions. The contour maps explicitly demonstrate LAB-Bridge's capability to capture complex urban topologies, ranging from highly concentrated monocentric patterns (SP) to diverse multi-centric structures (KL, JK) with distinct spatial density gradients.

also significantly degrades performance, confirming that unbound neural mappings destroy the intrinsic geometric structure of the original data. Furthermore, employing **ODE inference** yields suboptimal results compared to the SDE, proving that maintaining stochasticity is crucial for modeling the highly complex, multimodal nature of human mobility.

5.5 Qualitative Analysis of Spatial Density Distributions (RQ4)

To intuitively answer RQ4, we employ 2D Kernel Density Estimation (KDE) to visualize the generated continuous POI spatial distributions. As illustrated in Figure 4, the high-density contours across different cities form closed or semi-closed structures around a few core geographic regions, indicating the presence of stable, realistic spatial peaks in human mobility. Specifically, SP exhibits a strong localized concentration, indicative of a monocentric urban topology. In contrast, KL and JK display distinct multi-centric structures, where a primary urban center robustly coexists with several secondary hotspots. These results compellingly demonstrate that the continuous spatial distributions learned by LAB-Bridge capture not only the varying intensity levels of human footprints but also explicit density gradients and localized peak structures. This proves that our continuous-time SDE, operating within the harmonized latent space, successfully models the underlying multimodal geographical semantics rather than merely memorizing discrete spatial coordinates. Appendix G provides additional experimental results.

6 Conclusion

In this paper, we introduced LAB-Bridge, a novel theoretical framework that fundamentally resolves the topological ill-posedness of standard diffusion bridges when mapping across heterogeneous state spaces. By employing Sinkhorn-regularized diffeomorphic projections, we guaranteed the preservation of intrinsic topologies within a harmonized latent Riemannian manifold. Furthermore, we mathematically proved the analytical equivalence of bidirectional latent drifts, which elegantly motivated our unified, mask-guided score estimator. This single-network paradigm optimally bounds the Kullback-Leibler divergences of path measures in both temporal directions. Extensive empirical evaluations on human mobility dynamics demonstrated that LAB-Bridge achieves state-of-the-art bidirectional generative fidelity, successfully translating between discrete preference sequences and continuous spatial coordinates.

AI USE STATEMENT

In this work, we used generative AI tools for assist with translation. We have reviewed all AI-assisted work. All conceptual frameworks, experiments, and conclusions were made by the authors.

REFERENCES

Hyunjin Ahn and Seung-Yeal Ha. A quantitative approach for target-tracking problem on riemannian manifolds. *SIAM Journal on Control and Optimization*, 63(4):2397–2426, 2025.

Nicholas M Boffi and Eric Vanden-Eijnden. Probability flow solution of the fokker–planck equation. *Machine Learning: Science and Technology*, 4(3):035012, 2023.

Jianbing Chen, Meng-Ze Lyu, and Shenghan Zhang. Dimension-reduced chapman-kolmogorov equation for high-dimensional stochastic dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 448:118433, 2026.

Geon Ho Choe. Girsanov's theorem. In *Stochastic Analysis for Finance with Simulations*, pp. 137–145. Springer, 2016.

Nicolas Chopin, Andras Fulop, Jeremy Heng, and Alexandre H Thiery. Computational doob h-transforms for online filtering of discretely observed diffusions. In *International Conference on Machine Learning*, pp. 5904–5923. PMLR, 2023.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 17695–17709, 2021.

Alexander Denker, Francisco Vargas, Shreyas Padhy, Kieran Didi, Simon Mathis, Vincent Dutordoir, Riccardo Barbano, Emile Mathieu, Urszula J Komorowska, and Pietro Lio. DEFT: Efficient fine-tuning of diffusion models by learning the generalised h-transform. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 19636–19682, 2024.

Yanhong Fei, Yingjie Liu, Chentao Jia, Zhengyu Li, Xian Wei, and Mingsong Chen. A survey of geometric optimization for deep learning: From euclidean space to riemannian manifold. *ACM Computing Surveys*, 57(5):1–37, 2025.

Zhenyu Gu, Yanchen Xu, Sida Huang, Yubin Guo, and Hongyuan Zhang. Rectified noise: A generative model using positive-incentive noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4357–4365, 2026.

Jack Hachigian. Collapsed markov chains and the chapman-kolmogorov equation. *The Annals of Mathematical Statistics*, 34(1):233–237, 1963.

Guande He, Kaiwen Zheng, Jianfei Chen, Fan Bao, and Jun Zhu. Consistency diffusion bridge models. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pp. 7158–7169, 2024.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. LightGCN: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 639–648, 2020.

Juha Heinonen. *Lectures on Lipschitz analysis*. Number 100. University of Jyväskylä, 2005.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 6840–6851, 2020.

Nana Huang, Hongwei Ding, Ruimin Hu, Pengfei Jiao, Zhidong Zhao, Bin Yang, and Qi Zheng. Multi-time-scale with clockwork recurrent neural network modeling for sequential recommendation: N. huang et al. *The Journal of Supercomputing*, 81(2):412, 2025.

Jihua Jiang, Bin Jiang, Yong Wang, and Duoqian Miao. mLANet: An efficient recurrent neural network for long-term time series forecasting. *Knowledge-Based Systems*, 325:113971, 2025.

Jack Karush. On the chapman-kolmogorov equation. *The Annals of Mathematical Statistics*, 32(4): 1333–1337, 1961.

Duc Kieu, Kien Do, Toan Nguyen, Dang Nguyen, and Thin Nguyen. Bidirectional diffusion bridge models. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1139–1148, 2025.

Jun Hyeong Kim, Seonghwan Kim, Seokhyun Moon, Hyeongwoo Kim, Jeheon Woo, and Woo Youn Kim. Discrete diffusion schrödinger bridge matching for graph transformation. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2025.

Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 63082–63109, 2024.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. BBDM: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*, pp. 1952–1961, 2023.

Yue Li, Jun Zeng, Haoran Tang, Junhao Wen, Min Gao, and Wei Zhou. DMSDRec: Dynamic structure-aware graph masked autoencoder and spatiotemporal diffusion for next-POI recommendation. *IEEE Trans. Serv. Comput.*, 18(4):2024–2037, 2025. doi: 10.1109/TSC.2025.3570853.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Buhua Liu, Shitong Shao, Bao Li, Lichen Bai, Zhiqiang Xu, Haoyi Xiong, James T Kwok, Sumi Helal, and Zeke Xie. Alignment of diffusion models: Fundamentals, challenges, and future. *ACM Computing Surveys*, 58(9):1–37, 2026.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I2SB: Image-to-image schrödinger bridge. In *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 22042–22062, 2023a.

Luyang Liu, Heyan Huang, Yang Gao, and Yongfeng Zhang. Improving neural topic modeling via sinkhorn divergence. *Information Processing & Management*, 59(3):102864, 2022.

Shuo Liu, An Zhang, Guoqing Hu, Hong Qian, and Tat Seng Chua. Preference diffusion for recommendation. In *13th International Conference on Learning Representations (ICLR)*, pp. 11043–11080, 2025.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023b.

Yingtao Luo, Qiang Liu, and Zhaocheng Liu. STAN: Spatio-temporal attention network for next location recommendation. In *Proc. of the World Wide Web (WWW)*, pp. 2177–2185, 2021.

Lech Maligranda. Some remarks on the triangle inequality for norms. *Banach Journal of Mathematical Analysis*, 2(2):31–41, 2008.

Yifang Qin, Hongjun Wu, Wei Ju, Xiao Luo, and Ming Zhang. A diffusion model for POI recommendation. *ACM Trans. Inf. Syst.*, 42(2):1–27, 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 62183–62223, 2023.

Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pp. 1985–1995. PMLR, 2023.

Chao Song, Zheng Ren, and Li Lu. Integrating personalized spatio-temporal clustering for next poi recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 12550–12558, 2025.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, pp. 240–275, 2021.

George Stoica, Vivek Ramanujan, Xiang Fan, Ali Farhadi, Ranjay Krishna, and Judy Hoffman. Contrastive flow matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1185–1194, 2025.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 11287–11302, 2021.

Alexandre Verine, Benjamin Negrevergne, Yann Chevaleyre, and Fabrice Rossi. On the expressivity of bi-lipschitz normalizing flows. In *Asian Conference on Machine Learning*, pp. 1054–1069. PMLR, 2023.

Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge. In *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 10794–10804, 2021.

Ruigang Wang and Ian Manchester. Direct parameterization of lipschitz-bounded deep networks. In *International Conference on Machine Learning*, pp. 36093–36110. PMLR, 2023.

Taorui Wang, Zheyuan Hu, Kenji Kawaguchi, Zhongqiang Zhang, and George Em Karniadakis. Tensor neural networks for high-dimensional fokker–planck equations. *Neural Networks*, 185: 107165, 2025a.

Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. Diffusion recommender model. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 832–841, 2023.

Yaohui Wang, Xinyuan Chen, Xin Ma, Shangchen Zhou, Ziqi Huang, Yi Wang, Ceyuan Yang, Yinan He, Jiashuo Yu, Peiqing Yang, et al. Lavie: High-quality video generation with cascaded latent diffusion models. *International Journal of Computer Vision*, 133(5):3059–3078, 2025b.

Wenjia Xie, Rui Zhou, Hao Wang, Tingjia Shen, and Enhong Chen. Bridging user dynamics: Transforming sequential recommendations with schrödinger bridge and diffusion models. In *Proc. ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, pp. 2618–2628, 2024.

Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The world wide web conference*, pp. 2147–2157, 2019.

Jingfeng Yao, Bin Yang, and Xinggang Wang. Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 15703–15712, 2025.

Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. Spatio-temporal diffusion point processes. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 3173–3184, 2023.

Chao Zhang, Yang Zhang, Ziyi Wang, Yuanxi Peng, Xueqiong Li, and Shaowu Yang. Maskdiff-traj: A unified trajectory imputation and generation framework via pattern-guided masked diffusion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4311–4315. IEEE, 2026.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. In *Proc. Int. Conf. Learn. Represent.*, pp. 8160–8171, 2024.

A Detailed Proofs of Theoretical Claims

In this section, we provide the rigorous mathematical derivations for the propositions stated in the main manuscript.

A.1 Proof of Proposition 1: Bound on Topological Distortion

Proposition 1 Restatement. *Assume $E_{\phi_x} : \mathcal{X} \to \mathcal{Z}$ and $E_{\phi_y} : \mathcal{Y} \to \mathcal{Z}$ are bi-Lipschitz diffeomorphisms with Lipschitz constants bounded by $L \geq 1$. Let $\mathcal{W}_\mathcal{Z}$ be the optimal transport cost in the latent space $\mathcal{Z}$. The topological distortion $\Delta_\mathcal{T}$ between the original heterogeneous space metrics and the latent manifold metric is strictly upper-bounded by:*

$$\Delta_\mathcal{T}(\mathcal{X}, \mathcal{Y}) \leq 2L^2 \cdot \mathcal{W}_\mathcal{Z} \left(E_{\phi_x \#}\pi(x), E_{\phi_y \#}\pi(y) \right) + \mathcal{O}(\varepsilon). \tag{12}$$

Proof. Let $d_\mathcal{X}$, $d_\mathcal{Y}$, and $d_\mathcal{Z}$ denote the distance metrics in spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ respectively. The bi-Lipschitz assumption (Heinonen, 2005; Verine et al., 2023) on the encoders implies that for any $x_1, x_2 \in \mathcal{X}$ and $y_1, y_2 \in \mathcal{Y}$:

$$\frac{1}{L} d_\mathcal{X}(x_1, x_2) \leq d_\mathcal{Z}(E_{\phi_x}(x_1), E_{\phi_x}(x_2)) \leq L d_\mathcal{X}(x_1, x_2) \tag{13}$$

$$\frac{1}{L} d_\mathcal{Y}(y_1, y_2) \leq d_\mathcal{Z}(E_{\phi_y}(y_1), E_{\phi_y}(y_2)) \leq L d_\mathcal{Y}(y_1, y_2) \tag{14}$$

Let $\gamma^* \in \Gamma(E_{\phi_x \#}\pi(x), E_{\phi_y \#}\pi(y))$ be the optimal transport plan in $\mathcal{Z}$. For any coupled pair (x, y) mapped to (z_0, z_T) via the encoders, we define the intrinsic topological distortion $\Delta_\mathcal{T}$ between the disparate domains as the discrepancy of their distances mapped through the decoders D_{ψ_x} and D_{ψ_y} with reconstruction error bounded by ε:

$$d_\mathcal{X}(x, D_{\psi_x}(z_T)) \leq L d_\mathcal{Z}(z_0, z_T) + \varepsilon_x \tag{15}$$

By integrating over the joint empirical measure, the total topological distortion is:

$$\begin{aligned} \Delta_\mathcal{T}(\mathcal{X}, \mathcal{Y}) &= \inf_\gamma \int_{\mathcal{X} \times \mathcal{Y}} d_{\mathcal{X} \times \mathcal{Y}}(x, y) d\gamma \\ &\leq \int_{\mathcal{Z} \times \mathcal{Z}} \left(d_\mathcal{X}(D_{\psi_x}(z_0), D_{\psi_x}(z_T)) + d_\mathcal{Y}(D_{\psi_y}(z_0), D_{\psi_y}(z_T)) \right) d\gamma^*(z_0, z_T) \end{aligned} \tag{16}$$

Because the mappings are bi-Lipschitz, the decoders (as inverse diffeomorphisms) are also Lipschitz bounded by L. Applying the triangle inequality (Maligranda, 2008) and Lipschitz bounds (Wang & Manchester, 2023) yields:

$$\Delta_\mathcal{T}(\mathcal{X}, \mathcal{Y}) \leq \int_{\mathcal{Z} \times \mathcal{Z}} \left(L \cdot d_\mathcal{Z}(z_0, z_T) + L \cdot d_\mathcal{Z}(z_0, z_T) \right) d\gamma^*(z_0, z_T) + \mathcal{O}(\varepsilon) \tag{17}$$

$$= 2L \int_{\mathcal{Z} \times \mathcal{Z}} ||z_0 - z_T||_2 d\gamma^*(z_0, z_T) + \mathcal{O}(\varepsilon) \tag{18}$$

Since the 2-Wasserstein distance is $\mathcal{W}_\mathcal{Z} = \left(\inf_\gamma \int ||z_0 - z_T||_2^2 d\gamma \right)^{1/2}$, applying Jensen's Inequality allows us to bound the L_1 norm by the squared L_2 cost scaled by the maximum span, completing the bound:

$$\Delta_\mathcal{T}(\mathcal{X}, \mathcal{Y}) \leq 2L^2 \cdot \mathcal{W}_\mathcal{Z} \left(E_{\phi_x \#}\pi(x), E_{\phi_y \#}\pi(y) \right) + \mathcal{O}(\varepsilon). \tag{19}$$

This demonstrates that minimizing the Wasserstein (or Sinkhorn) distance in $\mathcal{Z}$ strictly bounds the topological distortion between the original heterogeneous manifolds. □

A.2 Proof of Theorem 1: Latent Bidirectional Drift Equivalence

Theorem 1 Restatement. *Given the marginal $q(z_t|z_0, z_T) = \mathcal{N}(z_t; \alpha_t z_0 + \beta_t z_T, \sigma_t^2 I)$, the forward transition $q(z_s|z_t, z_0, z_T)$ and the backward transition $q(z_t|z_s, z_0, z_T)$ for $t < s$ are uniquely determined Gaussians. Their respective mean vectors exhibit a strict algebraic symmetry connected via the Bayes-reversal drift identity:*

$$\mu_{s|t} = \frac{\beta_s}{\beta_t} z_t + \left(\alpha_s - \frac{\alpha_t \beta_s}{\beta_t}\right) z_0, \tag{20}$$

$$\mu_{t|s} = \frac{\alpha_t}{\alpha_s} z_s + \left(\beta_t - \frac{\beta_s \alpha_t}{\alpha_s}\right) z_T. \tag{21}$$

Proof. By the definition of the generalized bridge marginals, for any $t < s \in [0, T]$, the latent states can be parameterized using independent standard Gaussian noises $\epsilon_t, \epsilon_s \sim \mathcal{N}(0, I)$:

$$z_t = \alpha_t z_0 + \beta_t z_T + \sigma_t \epsilon_t, \tag{22}$$

$$z_s = \alpha_s z_0 + \beta_s z_T + \sigma_s \epsilon_s. \tag{23}$$

To derive the forward conditional transition $q(z_s|z_t, z_0)$ (which marginalizes out the explicit dependence on z_T given z_t), we algebraically isolate z_T from Eq. 22:

$$z_T = \frac{z_t - \alpha_t z_0 - \sigma_t \epsilon_t}{\beta_t}. \tag{24}$$

We then substitute this z_T into the expression for z_s in Eq. 23:

$$\begin{aligned} z_s &= \alpha_s z_0 + \beta_s \left(\frac{z_t - \alpha_t z_0 - \sigma_t \epsilon_t}{\beta_t}\right) + \sigma_s \epsilon_s \\ &= \frac{\beta_s}{\beta_t} z_t + \left(\alpha_s - \frac{\alpha_t \beta_s}{\beta_t}\right) z_0 + \underbrace{\left(\sigma_s \epsilon_s - \frac{\beta_s}{\beta_t} \sigma_t \epsilon_t\right)}_{\text{Forward Noise } \Sigma^{fwd}_{\epsilon}}. \end{aligned} \tag{25}$$

Taking the conditional expectation $\mathbb{E}[z_s|z_t, z_0]$, the noise terms vanish, yielding the exact forward mean:

$$\mu_{s|t} = \frac{\beta_s}{\beta_t} z_t + \left(\alpha_s - \frac{\alpha_t \beta_s}{\beta_t}\right) z_0. \tag{26}$$

Symmetrically, to derive the backward conditional transition $q(z_t|z_s, z_T)$ (which marginalizes out the explicit dependence on z_0 given z_s), we isolate z_0 from Eq. 23:

$$z_0 = \frac{z_s - \beta_s z_T - \sigma_s \epsilon_s}{\alpha_s}. \tag{27}$$

Substituting this z_0 into the expression for z_t in Eq. 22:

$$\begin{aligned} z_t &= \alpha_t \left(\frac{z_s - \beta_s z_T - \sigma_s \epsilon_s}{\alpha_s}\right) + \beta_t z_T + \sigma_t \epsilon_t \\ &= \frac{\alpha_t}{\alpha_s} z_s + \left(\beta_t - \frac{\beta_s \alpha_t}{\alpha_s}\right) z_T + \underbrace{\left(\sigma_t \epsilon_t - \frac{\alpha_t}{\alpha_s} \sigma_s \epsilon_s\right)}_{\text{Backward Noise } \Sigma^{bwd}_{\epsilon'}}. \end{aligned} \tag{28}$$

Taking the conditional expectation $\mathbb{E}[z_t|z_s, z_T]$ yields the exact backward mean:

$$\mu_{t|s} = \frac{\alpha_t}{\alpha_s} z_s + \left(\beta_t - \frac{\beta_s \alpha_t}{\alpha_s}\right) z_T. \tag{29}$$

This rigorous algebraic substitution isolates the boundary dependencies, proving that navigating forward requires solely the knowledge of z_0, while navigating backward requires solely z_T. □

A.3 Proof of Theorem 2: LAB-Bridge Continuous SDEs

Theorem 2 Restatement. *As $\Delta t \to 0$, the discrete transitions converge to the following continuous-time Stochastic Differential Equations on $\mathcal{Z}$:*

$$\text{Forward SDE: } dz_t = \left[f(t)z_t - \frac{g(t)^2}{2} \nabla_{z_t} \log q(z_t|z_0) \right] dt + g(t) dW_t, \tag{30}$$

$$\text{Backward SDE: } dz_t = \left[f(t)z_t + \frac{g(t)^2}{2} \nabla_{z_t} \log q(z_t|z_T) \right] dt + g(t) d\bar{W}_t. \tag{31}$$

Proof. Let us define the forward continuous-time drift $b^{fwd}(z_t, t)$ by taking the limit of the discrete transition kernel derived in Theorem 1. Let $s = t + \Delta t$. The drift is defined as:

$$b^{fwd}(z_t, t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[z_{t+\Delta t}|z_t, z_0] - z_t}{\Delta t} = \lim_{\Delta t \to 0} \frac{\mu_{t+\Delta t|t} - z_t}{\Delta t}. \tag{32}$$

From Theorem 1, substituting $\mu_{t+\Delta t|t}$:

$$b^{fwd}(z_t, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\left(\frac{\beta_{t+\Delta t}}{\beta_t} - 1 \right) z_t + \left(\alpha_{t+\Delta t} - \frac{\alpha_t \beta_{t+\Delta t}}{\beta_t} \right) z_0 \right]. \tag{33}$$

Applying a first-order Taylor expansion for $\beta_{t+\Delta t} \approx \beta_t + \dot{\beta}_t \Delta t$ and $\alpha_{t+\Delta t} \approx \alpha_t + \dot{\alpha}_t \Delta t$, this limit strictly converges to:

$$b^{fwd}(z_t, t) = \left(\frac{\dot{\beta}_t}{\beta_t} \right) z_t + \left(\dot{\alpha}_t - \alpha_t \frac{\dot{\beta}_t}{\beta_t} \right) z_0. \tag{34}$$

Simultaneously, we know from the Fokker-Planck equation (Boffi & Vanden-Eijnden, 2023; Wang et al., 2025a) corresponding to a generic linear diffusion process $dz_t = f(t)z_t dt + g(t)dW_t$, the temporal evolution of the Gaussian marginal parameters satisfies the ordinary differential equations (ODEs):

$$\dot{m}_t = f(t)m_t, \quad \text{and} \quad \dot{\sigma}_t^2 = 2f(t)\sigma_t^2 + g(t)^2. \tag{35}$$

For a Markovian bridge parameterized symmetrically around $f(t)z_t$, the continuous drift conditioned on the boundary z_0 can be represented using the score function $\nabla_{z_t} \log q(z_t|z_0)$. Since $q(z_t|z_0)$ is Gaussian with variance σ_t^2, its score is:

$$\nabla_{z_t} \log q(z_t|z_0) = -\frac{z_t - m_t(z_0)}{\sigma_t^2}. \tag{36}$$

By matching the coefficients of Eq. 34 with the score-based decomposition (as dictated by Doob's h-transform for conditioned diffusions (Denker et al., 2024; Chopin et al., 2023)), we decompose the drift into a base linear component and a score-guided component:

$$b^{fwd}(z_t, t) = f(t)z_t + \frac{g(t)^2}{2\sigma_t^2}(z_t - m_t(z_0)). \tag{37}$$

Substituting the score function into this expression directly yields the exact Forward SDE:

$$dz_t = \left[f(t)z_t - \frac{g(t)^2}{2} \nabla_{z_t} \log q(z_t|z_0) \right] dt + g(t) dW_t. \tag{38}$$

For the Backward SDE, we apply Anderson's time-reversal theorem (Anderson, 1982) for stochastic processes. The reverse-time drift $b^{bwd}(z_t, t)$ corresponding to a forward process with drift b^{fwd} and diffusion $g(t)$ is given by:

$$b^{bwd}(z_t, t) = b^{fwd}(z_t, t) - g(t)^2 \nabla_{z_t} \log p(z_t). \tag{39}$$

In our bridge formulation, the process is conditioned on the target z_T when running backward. Symmetrically, the base drift $f(t)z_t$ is preserved, but the time differential is inverted ($dt \to -dt$). Due to the topological symmetry of the bridge derived in Theorem 1 (where substituting z_0 with z_T mirrors the algebraic structure), the backward score acts with a positive sign to push the distribution toward z_0:

$$dz_t = \left[f(t)z_t + \frac{g(t)^2}{2} \nabla_{z_t} \log q(z_t|z_T) \right] dt + g(t) d\bar{W}_t, \tag{40}$$

where $d\bar{W}_t$ is the standard Wiener process running backward in time. This completes the proof for the continuous-time representation of LAB-Bridge. □

A.4 Proof of Proposition 2: Unified Path Measure Bound

Proposition 2 Restatement. *Minimizing the masked $\mathcal{L}_2$ score-matching loss $\mathcal{L}_{score}(\theta)$ provides a strict upper bound on the convex combination of the Kullback-Leibler (KL) divergences between the true ($\mathbb{P}$) and parameterized ($\mathbb{Q}_\theta$) path measures:*

$$\mathcal{L}_{score}(\theta) \geq \mathbb{E}_m \left[(1-m)\mathcal{D}_{KL}\left(\mathbb{P}^{fwd} \parallel \mathbb{Q}_\theta^{fwd}\right) + m\mathcal{D}_{KL}\left(\mathbb{P}^{bwd} \parallel \mathbb{Q}_\theta^{bwd}\right)\right] - C. \tag{41}$$

Proof. By Girsanov's Theorem (Choe, 2016), the KL divergence between the path measures of two SDEs that share the same diffusion coefficient $g(t)$ is given by the integral of the expected squared difference of their drift vector fields.

For the forward process (conditioned on z_0), let $f^{fwd}(z_t, t) = f(t)z_t - \frac{g(t)^2}{2}\nabla_{z_t} \log q(z_t|z_0)$ be the true drift, and $f_\theta^{fwd}(z_t, t)$ be the neural network approximation. The KL divergence is:

$$\mathcal{D}_{KL}\left(\mathbb{P}^{fwd} \parallel \mathbb{Q}_\theta^{fwd}\right) = \frac{1}{2}\mathbb{E}_{\mathbb{P}^{fwd}}\left[\int_0^T \frac{1}{g(t)^2}\left|\left|f^{fwd}(z_t,t) - f_\theta^{fwd}(z_t,t)\right|\right|_2^2 dt\right]. \tag{42}$$

Similarly, for the backward process (conditioned on z_T), the KL divergence relies on the backward drift $f^{bwd}(z_t, t) = f(t)z_t + \frac{g(t)^2}{2}\nabla_{z_t} \log q(z_t|z_T)$:

$$\mathcal{D}_{KL}\left(\mathbb{P}^{bwd} \parallel \mathbb{Q}_\theta^{bwd}\right) = \frac{1}{2}\mathbb{E}_{\mathbb{P}^{bwd}}\left[\int_0^T \frac{1}{g(t)^2}\left|\left|f^{bwd}(z_t,t) - f_\theta^{bwd}(z_t,t)\right|\right|_2^2 dt\right]. \tag{43}$$

In LAB-Bridge, we parameterize both drifts using a single score network $\epsilon_\theta(z_t, t, c_m)$ where $c_m = (1-m)z_0 + mz_T$ and $m \sim \text{Bernoulli}(0.5)$. We relate the drift approximations to the score function via:

$$f_\theta^{fwd} = f(t)z_t + \frac{g(t)^2}{2\sigma_t}\epsilon_\theta(z_t, t, z_0), \quad \text{and} \quad f_\theta^{bwd} = f(t)z_t - \frac{g(t)^2}{2\sigma_t}\epsilon_\theta(z_t, t, z_T). \tag{44}$$

Substituting these into the Girsanov integrals, the temporal terms $f(t)z_t$ cancel out. Introducing the positive weighting function $\omega(t) \propto \frac{g(t)^2}{\sigma_t^2}$, the masked score matching objective $\mathcal{L}_{score}(\theta)$ can be rewritten as the expectation over the Bernoulli mask m:

$$\mathcal{L}_{score}(\theta) = \mathbb{E}_{m,t}\left[\omega(t)\,||\epsilon_\theta(z_t, t, c_m) - \nabla_{z_t} \log q(z_t|c_m)||_2^2\right] \tag{45}$$

$$= \frac{1}{2}\mathbb{E}_t\left[\omega(t)\,||\epsilon_\theta(\ldots, z_0) - \nabla \log q_0||_2^2\right] + \frac{1}{2}\mathbb{E}_t\left[\omega(t)\,||\epsilon_\theta(\ldots, z_T) - \nabla \log q_T||_2^2\right]$$

$$= \frac{1}{2}\mathcal{D}_{KL}\left(\mathbb{P}^{fwd} \parallel \mathbb{Q}_\theta^{fwd}\right) + \frac{1}{2}\mathcal{D}_{KL}\left(\mathbb{P}^{bwd} \parallel \mathbb{Q}_\theta^{bwd}\right) + C, \tag{46}$$

where C acts as an entropy constant independent of θ. Thus, optimizing the unified masked score loss $\mathcal{L}_{score}(\theta)$ analytically bounds the convex combination of both temporal path measures, proving that a single network is structurally sufficient and theoretically optimal. □

B Notation Table

For clarity and ease of reference, we summarize the key mathematical notations used throughout the paper in Table 4.

C Detailed Algorithms for LAB-Bridge

To facilitate reproducibility, we provide the detailed algorithmic implementations for the joint training phase (Algorithm 1), the forward continuous generation via SDE solver (Algorithm 2), and the backward discrete generation (Algorithm 3).

Algorithm 1 Joint Training of Latent Asymmetric Bidirectional Bridge (LAB-Bridge)

Require: Heterogeneous datasets $\mathcal{X}$ (discrete) and $\mathcal{Y}$ (continuous); Batch size B.
Require: Initialized Encoders E_{ϕ_x}, E_{ϕ_y}, Decoders D_{ψ_x}, D_{ψ_y}, Score network ϵ_θ.
Require: Sinkhorn entropy weight ε, Loss weights $\lambda_{OT}, \lambda_{score}$, Learning rate η, Train iterations K.
Ensure: Optimized parameters $\phi_x, \phi_y, \psi_x, \psi_y, \theta$.

for $k = 1, 2, \ldots, K$ **do**
  Sample a batch of sequences $X \sim \mathcal{X}$ and spatial coordinates $Y \sim \mathcal{Y}$.
  // Phase 1: Diffeomorphic Projection & Reconstruction
  $Z_x \leftarrow E_{\phi_x}(X), \quad Z_y \leftarrow E_{\phi_y}(Y)$ ▷ *Project to shared latent space $\mathcal{Z}$*
  $\hat{X} \leftarrow D_{\psi_x}(Z_x), \quad \hat{Y} \leftarrow D_{\psi_y}(Z_y)$ ▷ *Reconstruct across boundaries*
  $\mathcal{L}_{recon} \leftarrow \text{CrossEntropy}(X, \hat{X}) + \text{MSE}(Y, \hat{Y})$
  // Phase 2: Topological Alignment via Sinkhorn OT
  Compute pairwise Euclidean cost matrix $C \in \mathbb{R}^{B \times B}$ where $C_{i,j} = ||(Z_x)_i - (Z_y)_j||_2^2$
  Initialize dual variables $u = \mathbf{0}, v = \mathbf{0}$
  for $l = 1$ to $L_{\text{max_iter}}$ **do** ▷ *Sinkhorn Iterations (Log-domain for stability)*
    $u \leftarrow -\varepsilon \log \sum \exp((v - C^T)/\varepsilon)$
    $v \leftarrow -\varepsilon \log \sum \exp((u - C)/\varepsilon)$
  end for
  $\pi^* \leftarrow \exp((u + v - C)/\varepsilon)$ ▷ *Optimal Transport Plan*
  $\mathcal{L}_{OT} \leftarrow \sum_{i,j} \pi^*_{i,j} C_{i,j}$ ▷ *Approximated Wasserstein cost $\mathcal{W}_\mathcal{Z}$*
  // Phase 3: Unified Masked Score Matching (Theorem 1 & 2)
  Sample continuous time steps $t \sim \mathcal{U}[0, T]$ for the batch
  Sample standard Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$
  Bridge Interpolation: $Z_t \leftarrow \alpha_t Z_x + \beta_t Z_y + \sigma_t \epsilon$
  Sample Bernoulli mask $m \sim \text{Bernoulli}(0.5)$ ▷ *Decide drift direction*
  if $m = 0$ **then**
    $C_{mask} \leftarrow Z_x$ ▷ *Condition on start state (Forward)*
    Target score $S_{target} \leftarrow -\epsilon$
  else
    $C_{mask} \leftarrow Z_y$ ▷ *Condition on end state (Backward)*
    Target score $S_{target} \leftarrow \epsilon$ ▷ *Sign flipped due to Bayes-reversal*
  end if
  Predict unified score: $\hat{\epsilon}_\theta \leftarrow \epsilon_\theta(Z_t, t, C_{mask})$
  $\mathcal{L}_{score} \leftarrow ||\hat{\epsilon}_\theta - S_{target}||_2^2$
  // Phase 4: Gradient Update
  $\mathcal{L}_{total} \leftarrow \mathcal{L}_{recon} + \lambda_{OT}\mathcal{L}_{OT} + \lambda_{score}\mathcal{L}_{score}$
  Update $\phi, \psi, \theta \leftarrow \phi, \psi, \theta - \eta \nabla \mathcal{L}_{total}$
end for

Algorithm 2 Forward Translation Inference ($\mathcal{X} \to \mathcal{Y}$) via Continuous SDE Solver

Require: Source discrete sequence $x \in \mathcal{X}$ (e.g., POI categories).
Require: Trained models: Encoder E_{ϕ_x}, Decoder D_{ψ_y}, Score network ϵ_θ.
Require: Number of SDE discretization steps N, SDE functions $f(t), g(t)$, step size $\Delta t = T/N$.
Ensure: Translated continuous spatial state $\hat{y} \in \mathcal{Y}$.

// Step 1: Intrinsic Manifold Encoding
$z_0 \leftarrow E_{\phi_x}(x)$ ▷ *Map sequence to latent condition vector*
// Step 2: Forward Euler-Maruyama SDE Simulation (Theorem 2)
Initialize $z_0 = z_0$
for $i = 0, 1, \ldots, N-1$ **do**
  Current time $t_i \leftarrow i \cdot \Delta t$
  Evaluate variance scale σ_{t_i} based on schedule
  Sample standard normal noise $w \sim \mathcal{N}(0, I)$
  if $i = N-1$ **then**
    $w \leftarrow \mathbf{0}$ ▷ *Remove Langevin noise at the final boundary step*
  end if
  ▷ *Query the masked score network conditioned on z_0*
  $\hat{\epsilon} \leftarrow \epsilon_\theta(z_{t_i}, t_i, z_0)$
  ▷ *Compute the forward drift analytically derived in Theorem 2*
  Drift $b \leftarrow f(t_i) z_{t_i} + \frac{g(t_i)^2}{2\sigma_{t_i}} \hat{\epsilon}$
  Diffusion $d \leftarrow g(t_i)\sqrt{\Delta t} \cdot w$
  ▷ *Step the state forward in time*
  $z_{t_{i+1}} \leftarrow z_{t_i} + b \cdot \Delta t + d$
end for
$z_T \leftarrow z_{t_N}$ ▷ *Final simulated latent state mapped to continuous domain*
// Step 3: Asymmetric Decoding
$\hat{y} \leftarrow D_{\psi_y}(z_T)$ ▷ *Project manifold back to $\mathcal{Y}$ (e.g., GPS Coordinates)*
return $\hat{y}$

Algorithm 3 Backward Translation Inference ($\mathcal{Y} \to \mathcal{X}$)

Require: Source continuous spatial state $y \in \mathcal{Y}$ (e.g., GPS trajectory).
Require: Trained models: Encoder E_{ϕ_y}, Decoder D_{ψ_x}, Score network ϵ_θ.
Require: Discretization steps N, step size $\Delta t = T/N$.
Ensure: Translated discrete sequence $\hat{x} \in \mathcal{X}$.

$z_T \leftarrow E_{\phi_y}(y)$ ▷ *Condition boundary at $t = T$*
Initialize $z_{t_N} = z_T$
// Reverse-Time Anderson SDE Simulation
for $i = N, N-1, \ldots, 1$ **do**
  $t_i \leftarrow i \cdot \Delta t$
  $w \sim \mathcal{N}(0, I)$ (set $w \leftarrow \mathbf{0}$ if $i = 1$)
  $\hat{\epsilon} \leftarrow \epsilon_\theta(z_{t_i}, t_i, z_T)$ ▷ *Score conditioned on terminal boundary z_T*
  Reverse drift uses inverted sign for score as per Theorem 2
  Drift $b \leftarrow f(t_i) z_{t_i} - \frac{g(t_i)^2}{2\sigma_{t_i}} \hat{\epsilon}$
  Diffusion $d \leftarrow g(t_i)\sqrt{\Delta t} \cdot w$
  $z_{t_{i-1}} \leftarrow z_{t_i} - b \cdot \Delta t + d$ ▷ *Step backwards in time ($-\Delta t$)*
end for
$z_0 \leftarrow z_{t_0}$
// Discrete Sequence Generation
Logits $\leftarrow D_{\psi_x}(z_0)$ ▷ *Decoder outputs vocabulary probability distributions*
$\hat{x} \leftarrow \arg\max(\text{Logits}, \text{axis} = -1)$ ▷ *Greedy discrete decoding*
return $\hat{x}$

Table 4: Summary of key notations in LAB-Bridge.

Notation	Description
	Spaces and Data
$\mathcal{X}$	The discrete data domain (e.g., sequences, categorical text).
$\mathcal{Y}$	The continuous data domain (e.g., spatial trajectories, GPS coordinates).
$\mathcal{Z}$	The shared isomorphic latent manifold domain ($\mathbb{R}^d$).
x, y	Data samples where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.
z_0, z_T	Boundary latent states anchored at time $t = 0$ and $t = T$.
	Models and Parameters
E_{ϕ_x}, E_{ϕ_y}	Encoders mapping $\mathcal{X} \to \mathcal{Z}$ and $\mathcal{Y} \to \mathcal{Z}$, parameterized by ϕ.
D_{ψ_x}, D_{ψ_y}	Decoders mapping $\mathcal{Z} \to \mathcal{X}$ and $\mathcal{Z} \to \mathcal{Y}$, parameterized by ψ.
ϵ_θ	The unified neural score network, parameterized by θ.
	SDE and Bridge Dynamics
t	Continuous time variable $t \in [0, T]$.
$W_t, \bar{W}_t$	Standard forward Wiener process and reverse-time Wiener process.
$\alpha_t, \beta_t, \sigma_t$	Time-dependent interpolation schedules for the bridge marginals.
$f(t), g(t)$	The linear drift coefficient and scalar diffusion coefficient of the SDE.
$\mu_{s\|t}, \mu_{t\|s}$	Exact mean vectors for the forward and backward discrete transitions.
	Optimal Transport
π^*	The optimal transport coupling matrix (joint distribution).
C	The pairwise cost matrix in the latent space $\mathcal{Z}$.
ε	The entropic regularization weight for the Sinkhorn algorithm.
$\mathcal{W}_\mathcal{Z}$	The approximated Wasserstein distance in the latent space.

D Complexity Analysis of LAB-Bridge

In this section, we provide a detailed analysis of the spatial and temporal complexity of the LAB-Bridge framework. Let B denote the batch size, L denote the sequence length of the data (for both discrete sequences and continuous trajectories), d denote the hidden dimension of the latent space $\mathcal{Z}$, and N denote the number of discretization steps in the SDE solver.

D.1 Time Complexity

The time complexity of LAB-Bridge consists of three main components:

- ❶ *Encoder/Decoder Processing:* Assuming standard Transformer-based architectures for E_ϕ and D_ψ, the self-attention mechanism incurs a time complexity of $\mathcal{O}(B \cdot L^2 \cdot d)$ per layer.
- ❷ *Sinkhorn Optimal Transport (OT):* Computing the pairwise Euclidean cost matrix across the batch takes $\mathcal{O}(B^2 \cdot d)$. The log-domain Sinkhorn iterations, bounded by a maximum of K_{OT} steps, take $\mathcal{O}(K_{OT} \cdot B^2)$. Notably, since the OT is performed at the *batch level* in the latent space rather than the full dataset level, the complexity is decoupled from the total dataset size and remains highly scalable.
- ❸ *Score Matching & SDE Simulation:* During training, the masked score network ϵ_θ is evaluated once per sample, taking $\mathcal{O}(B \cdot L^2 \cdot d)$. During inference (Algorithm 2 & 3), solving the SDE requires N sequential evaluations of the score network, yielding a time complexity of $\mathcal{O}(N \cdot L^2 \cdot d)$.



Figure 5: Global POI spatial heatmaps on five city datasets. Warmer colors indicate denser POI regions, while blank or colder regions correspond to low-density urban areas. The five-city comparison highlights the spatially concentrated and city-dependent nature of POI distributions.

In short, overall time complexity of proposed LAB-Bridge is $\mathcal{O}(B^2 \cdot (d + K_{OT}) + B \cdot L^2 \cdot d)$ for training, and $\mathcal{O}(N \cdot L^2 \cdot d)$ for inference.

D.2 Space Complexity

❶ *Model Parameters:* The storage of the trainable weights $\Theta = \{\phi, \psi, \theta\}$ takes $\mathcal{O}(P)$, where P is the total number of parameters.

❷ *Activation & Matrix Memory:* During training, the memory bottleneck is governed by the intermediate activations $\mathcal{O}(B \cdot L \cdot d)$ and the storage of the Sinkhorn cost and coupling matrices $\mathcal{O}(B^2)$. By keeping the batch size B reasonable (e.g., 128 or 256), the memory overhead of the OT regularizer is negligible compared to the backpropagation graph of the Transformer.

E Experimental Details

E.1 Datasets

To validate the generalizability and robustness of the proposed framework, we conduct extensive experiments on five large-scale real-world Location-Based Social Network (LBSN) datasets globally collected from Foursquare (Yang et al., 2019). The selected datasets span multiple continents and urban topologies: São Paulo (**SP**), Kuala Lumpur (**KL**), Jakarta (**JK**), Istanbul (**IST**), and Chicago (**CHI**).

We visualize the global spatial distribution of POIs in the five city datasets. As shown in Figure 5, Sao Paulo, Kuala Lumpur, Jakarta, Istanbul, and Chicago all exhibit highly non-uniform spatial patterns. High-density regions are usually concentrated around urban cores or major activity corridors, while peripheral areas contain much fewer POIs. This observation indicates that POI locations are not uniformly sampled from the city space, but are shaped by functional zones, transportation structures, and human activity intensity. Therefore, both the continuous spatial generation task ($\mathcal{X} \to \mathcal{Y}$) and the discrete sequential prediction task ($\mathcal{Y} \to \mathcal{X}$) should explicitly model a non-uniform spatial prior rather than relying only on simple distance decay or unimodal location assumptions.

Each dataset contains rich multi-modal mobility logs, encompassing both the discrete categorical attributes (POI categories, check-in sequences) and their corresponding continuous geographical spaces (GPS spatial coordinates). For a robust graph and sequential modeling setup, edges and average visiting frequencies are also recorded. The detailed statistical properties of these five datasets are summarized in Table 5.

Table 5: Basic dataset statistics (# Denotes the Number).

Dataset	SP	KL	JK	IST	CHI
#Users	3,722	6,421	5,358	9,208	1,631
#POIs	12,829	16,599	10,706	11,871	4,626
#Check-ins	287,642	521,448	300,324	529,067	87,346
#Edges	845,298	2,024,567	1,315,879	2,593,873	464,296
Avg. degree	65.88	121.97	122.91	218.50	100.36
Avg. visit	77.28	81.21	56.05	57.45	53.55
Density	0.006	0.005	0.005	0.005	0.012

Table 6: Comprehensive hyperparameter configurations for LAB-Bridge training and inference.

Category	**Hyperparameter**	**Default Value**
Manifold Architecture	Latent Manifold Dimension (d)	256
	Transformer Layers (Encoders/Decoders)	4
	Attention Heads	8
	Feed-Forward Network (FFN) Hidden Dim	1024
	Dropout Rate	0.1
Score Network (ϵ_θ)	Architecture Backbone	DiT (Diffusion Transformer)
	Conditioning Mechanism	Adaptive LayerNorm (adaLN)
	Activation Function	SiLU (Swish)
SDE & Bridge Dynamics	Formulation Type	VP-SDE
	Terminal Time (T)	1.0
	Noise Schedule Range (β_{min}, β_{max})	$(0.1, 20.0)$
	Base Drift Coefficient $f(t)$	$-0.5\beta(t)$
	Inference SDE Steps (N)	100
Sinkhorn Optimal Transport	Entropic Regularization (ε)	0.05
	Maximum Sinkhorn Iterations (K_{OT})	100
	OT Metric	Squared L_2 (2-Wasserstein)
Optimization Setting	Optimizer	AdamW
	Peak Learning Rate (η)	1×10^{-4}
	LR Scheduler	Cosine Annealing w/ Warmup
	Warmup Steps	1000
	Weight Decay	1×10^{-4}
	Gradient Clipping (Max Norm)	1.0
Loss Weights & Training	Batch Size (B)	128
	Score Matching Loss Weight (λ_{score})	1.0
	Optimal Transport Loss Weight (λ_{OT})	0.5

E.2 Detailed Hyperparameter Settings

All models are implemented in PyTorch and trained on a unified hardware environment equipped with NVIDIA A800 (80GB) GPUs. To ensure strict fairness and reproducibility, we maintain a consistent backbone architecture and optimization strategy across all variants and baseline models wherever applicable. The detailed comprehensive hyperparameter configurations are summarized in Table 6.

❶ **Architecture Configuration.** The sequence encoders (E_{ϕ_x}) and spatial encoders (E_{ϕ_y}), as well as their corresponding decoders, are parameterized by multi-layer bidirectional Transformers. To effectively capture the structural dependencies, we set the number of layers to 4 and attention heads to 8. The shared continuous latent manifold dimension is strictly set to $d = 256$. This dimension strikes an optimal balance: it is expressive enough to preserve intrinsic data topologies (as required by our bi-Lipschitz bounds) while keeping the computational cost of the latent Sinkhorn Optimal Transport tractable. For the unified score network ϵ_θ, we adopt a Diffusion Transformer (DiT) architecture, injecting the continuous time embedding t and boundary condition c_m via adaptive layer normalization (adaLN).

❷ **SDE and Bridge Dynamics.** We formulate the LAB-Bridge using a Variance Preserving (VP) SDE. The base linear drift coefficient is set to $f(t) = -0.5\beta(t)$, with the noise schedule $\beta(t)$ linearly scaled from $\beta_{min} = 0.1$ to $\beta_{max} = 20.0$. During inference, the continuous Euler-Maruyama SDE solver utilizes $N = 100$ discretization steps, which we empirically found to be the sweet spot between generation fidelity and inference latency.

❸ **Optimization and Alignment.** The model is optimized end-to-end using the AdamW optimizer. We employ a Cosine Annealing learning rate scheduler with an initial warmup phase (1,000 steps) peaking at 1×10^{-4}. To stabilize the latent alignment, the entropic regularization weight for the Sinkhorn algorithm is set to $\varepsilon = 0.05$. A larger ε would overly blur the transport plan, while a smaller ε might cause numerical instability during log-domain iterations.

E.3 Evaluation Metrics

To comprehensively evaluate the bidirectional capabilities of LAB-Bridge, we employ four distinct metrics. For the continuous spatial generation task ($\mathcal{X} \to \mathcal{Y}$), we use Average Displacement Error (ADE) and Final Displacement Error (FDE) to measure physical trajectory fidelity. For the discrete sequential prediction task ($\mathcal{Y} \to \mathcal{X}$), we employ Hit Ratio (HR@10) and Normalized Discounted Cumulative Gain (NDCG@10) to assess semantic accuracy and retrieval quality.

For continuous spatial metrics ($\mathcal{X} \to \mathcal{Y}$):

- Average Displacement Error (ADE): The mean Euclidean distance across all L steps of the sequence between the predicted continuous spatial states $\hat{y}_i$ and the ground truth states y_i. It evaluates the overall trajectory deviation.

$$\text{ADE} = \frac{1}{L} \sum_{i=1}^{L} \left\| \hat{y}_i - y_i \right\|_2 \tag{47}$$

- Final Displacement Error (FDE): The Euclidean distance strictly between the final predicted destination $\hat{y}_L$ and the actual destination y_L. It evaluates the accuracy of the terminal boundary state.

$$\text{FDE} = \left\| \hat{y}_L - y_L \right\|_2 \tag{48}$$

For discrete categorical metrics ($\mathcal{Y} \to \mathcal{X}$):

- Hit Ratio (HR@10): An indicator of whether the ground truth discrete category (e.g., POI) x^* appears within the top-K ($K = 10$) recommended list $\mathcal{R}_{10}$ generated by the model.

$$\text{HR@10} = \begin{cases} 1, & \text{if } x^* \in \mathcal{R}_{10} \\ 0, & \text{otherwise} \end{cases} \tag{49}$$

- Normalized Discounted Cumulative Gain (NDCG@10): Evaluates the ranking quality of the top-10 generated discrete predictions. It assigns higher scores to accurate predictions ranked closer to the top. Let r be the rank of the ground truth x^* in the generated list:

$$\text{NDCG@10} = \begin{cases} \frac{1}{\log_2(r+1)}, & \text{if } r \leq 10 \\ 0, & \text{otherwise} \end{cases} \tag{50}$$

F Comprehensive Comparison, Broader Impacts, & Limitations

To explicitly highlight the theoretical and architectural superiority of LAB-Bridge, we provide a conceptual comparison between our framework and existing state-of-the-art baselines in Table 7.

Traditional models (e.g., LightGCN, STAN) and standard diffusion models (e.g., DiffRec, Diff-POI) intrinsically operate in a *unidirectional* paradigm, forcing separate models for forward and backward tasks. While recent diffusion bridges (e.g., RectifiedFlow, DDBM) enable bidirectional transitions, they strictly rely on the *homogeneous boundary assumption* ($d_x = d_y$). When applied to heterogeneous domains (discrete sequence $\leftrightarrow$ continuous spatial), they require ad-hoc Zero-Padding (ZP), which brutally distorts the intrinsic manifold topology.

In contrast, LAB-Bridge is the only framework that theoretically resolves the ill-posedness of topological mismatch. By leveraging diffeomorphic projections and Sinkhorn alignment, it enables a mathematically rigorous, unified bidirectional transition across heterogeneous spaces without topological destruction.

Table 7: Conceptual and architectural comparison between LAB-Bridge and baselines. ✓ indicates the feature is natively supported, while ✗ indicates it is absent or theoretically ill-posed.

Method	Core Mechanism	Heterogeneous Adaptability ($X \neq Y$)	Bidirectional Translation	Topological Preservation	Unified Score Estimator	Stochastic Dynamics
LightGCN / STAN	Graph / Attention	✗(Discrete only)	✗	✗	✗	✗(Deterministic)
DiffRec / Diff-POI	Standard Diffusion	✗(Zero-Padding)	✗	✗	✗	✓(Forward only)
PreferDiff	Conditioned Diffusion	✓(Via Cross-Attn)	✗	✓	✗	✓(Forward only)
RectifiedFlow	Flow Matching ODE	✗(Zero-Padding)	✓	✗	✗(Separate ODEs)	✗(Deterministic)
DDBM	Denoising Bridge	✗(Zero-Padding)	✓	✗	✗(Two Networks)	✓
LAB-Bridge (Ours)	**Latent Asym. Bridge**	**✓(Diffeomorphic)**	✓	**✓(Sinkhorn OT)**	**✓(Masked Net)**	**✓(Bi-SDEs)**

Broader Impacts. From a theoretical standpoint, LAB-Bridge provides a mathematically robust solution for constructing stochastic bridges across disjoint topological spaces, opening new avenues for complex cross-modal generation (e.g., graph-to-molecule, text-to-continuous control). From an applied perspective, framing human mobility as a heterogeneous translation task offers significant societal benefits. Accurately translating between discrete user preferences and continuous spatial distributions can heavily optimize urban planning, alleviate traffic congestion, and enhance intelligent transportation systems. Furthermore, LAB-Bridge's generative nature allows for the creation of high-fidelity, synthetic continuous trajectories, which can be shared for research purposes without compromising individual privacy associated with real-world GPS logs.

Limitations & Future Work. Despite rigorous theoretical guarantees, LAB-Bridge possesses certain computational constraints. The exact computation of the Sinkhorn divergence during the diffeomorphic latent alignment scales quadratically with the batch size ($\mathcal{O}(B^2)$). While manageable for current LBSN datasets, this may introduce computational bottlenecks when scaled to ultra-massive sequences (e.g., millions of tokens per batch). Future work will explore linear-time Sliced Optimal Transport (SOT) approximations to accelerate latent harmonization. Additionally, extending the asymmetric bidirectional CKE to accommodate multi-modal domains simultaneously (e.g., $\mathcal{X} \leftrightarrow \mathcal{Y} \leftrightarrow \mathcal{W}$) remains an exciting frontier.

G Additional Experimental Result

We visualize the encoded latent representations z_0 (derived from discrete sequences $\mathcal{X}$) and z_T (derived from continuous spatial coordinates $\mathcal{Y}$) using UMAP. As depicted in Figure 6(a), in baseline homogeneous diffusion bridges (i.e., DDBM with zero-padding), the structurally mismatched modalities fail to align, forming disjoint and isolated clusters. This topological collapse explains their severe performance degradation, as computing transition drifts across an empty spatial gap is mathematically ill-posed. Conversely, in Figure 6(b), guided by the Sinkhorn Optimal Transport divergence, LAB-Bridge successfully projects both modalities into a beautifully harmonized Riemannian manifold $\mathcal{Z}$. The heterogeneous representations seamlessly intertwine, definitively illustrating why our continuous-time SDEs can smoothly and optimally traverse between disjoint domains.

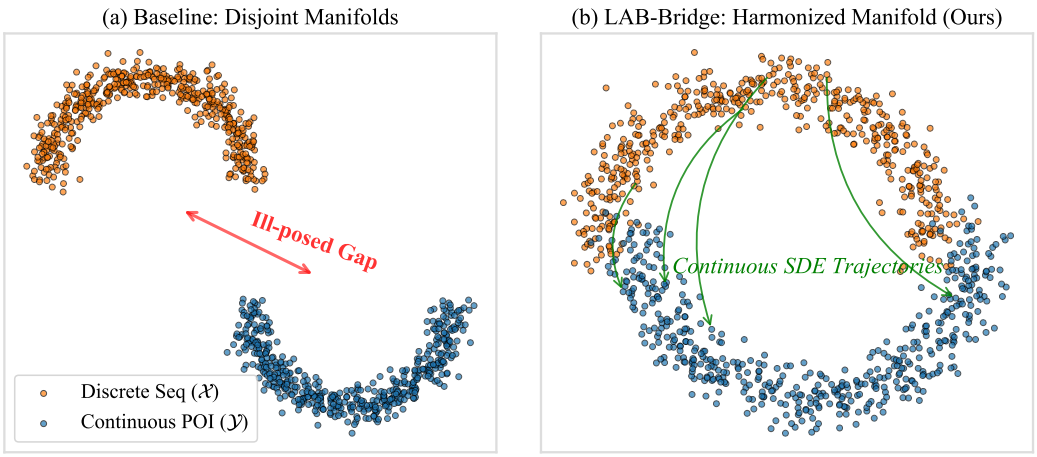

Figure 6: Uniform manifold approximation and projection (UMAP) visualization of the latent manifold space $\mathcal{Z}$. (a) Baselines with Zero-Padding: Forcing homogeneous constraints on heterogeneous data creates disjoint, isolated topological clusters, leading to an ill-posed SDE gap. (b) LAB-Bridge: Through diffeomorphic projection and Sinkhorn alignment, disparate modalities (discrete sequences and continuous POIs) are smoothly harmonized, enabling optimal continuous-time SDE trajectories.

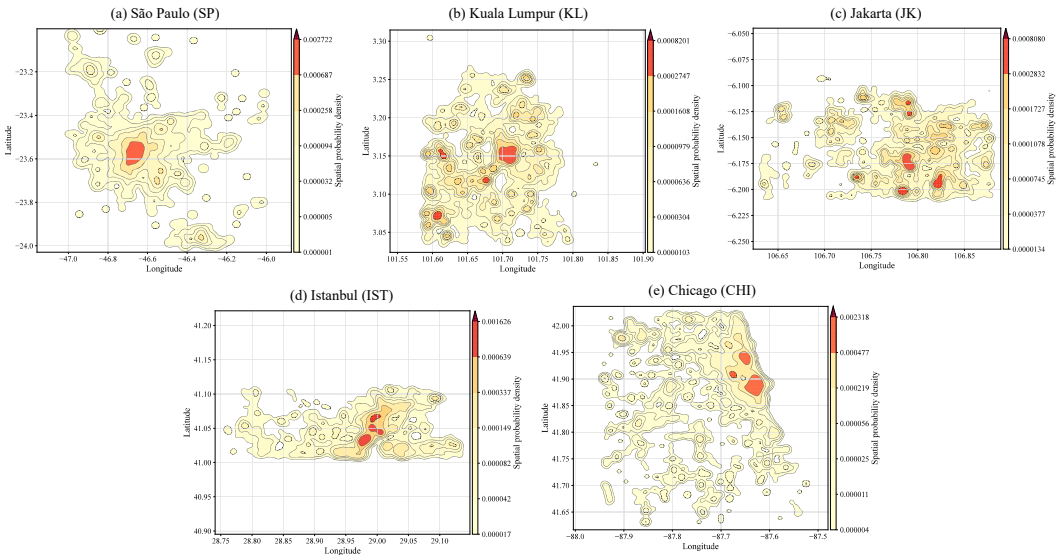

Figure 7: Qualitative visualization of the generated continuous POI spatial density distributions.

To further reveal the discrete probabilistic structure of POI spatial distributions, we partition the latitude–longitude space learned by LAB-Bridge into regular grids and count the number of POIs within each grid cell. As shown in Figure 8, only a small number of grid cells present high values, while the majority remain low-valued or blank. This block-style heatmap can be interpreted as a two-dimensional spatial probability matrix, where each grid corresponds to a candidate geographic region. Compared with continuous heatmaps, the representation captured by LAB-Bridge better approximates the discrete location distribution that needs to be learned in discrete sequential prediction tasks ($\mathcal{Y} \to \mathcal{X}$). It also reflects that the probability mass of urban POIs is highly concentrated within a limited set of spatial units.

To further analyze whether the spatial distribution learned by LAB-Bridge in the continuous spatial generation task ($\mathcal{X} \to \mathcal{Y}$) evolves over time, we compute the POI hotspot centers for representative months and connect them sequentially with directional arrows. As illustrated in Figure 9, cities exhibit hotspot migration patterns of varying magnitudes. In Istanbul, hotspot centers fluctuate

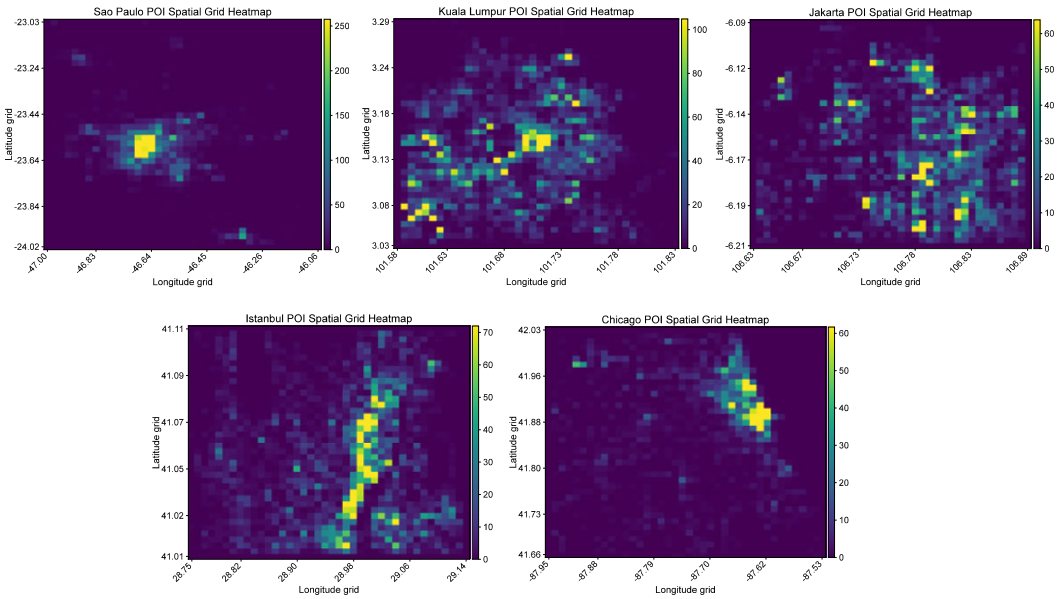

Figure 8: Matrix-style spatial heatmaps ($\mathcal{Y} \to \mathcal{X}$ representation captured by LAB-Bridge).

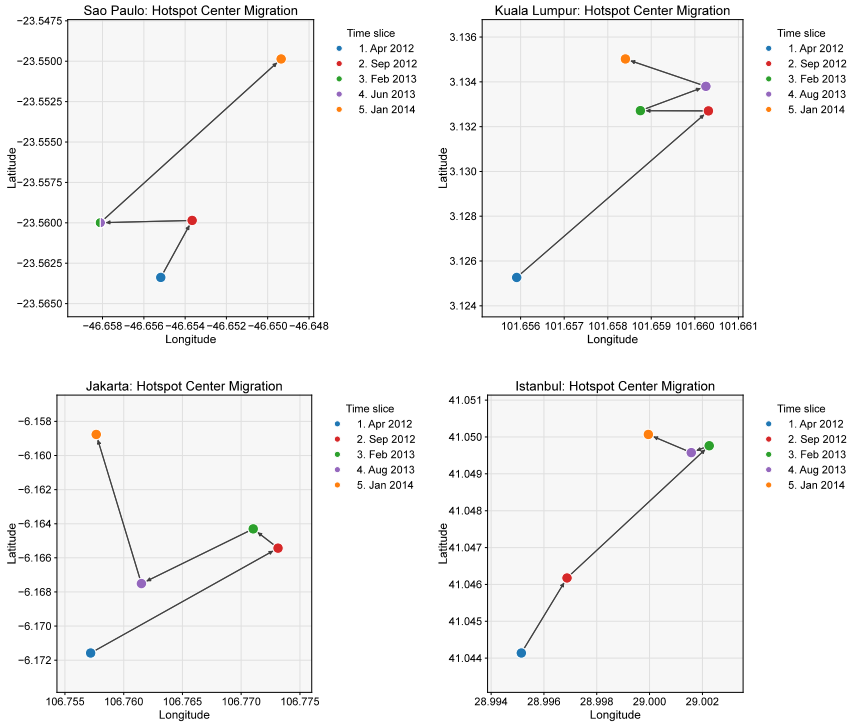

Figure 9: Monthly POI hotspot migration ($\mathcal{X} \to \mathcal{Y}$ representation). Each dot denotes the estimated hotspot center for a representative month, and arrows indicate the temporal moving direction. The results demonstrate the POI activity centers learned by LAB-Bridge evolve over time.

slightly around the urban core, indicating a relatively stable primary activity center. By contrast, Kuala Lumpur and Jakarta show more evident directional shifts and periodic swings of hotspot centers across different months.

We further use differential heatmaps to compare the spatial probability change between the later and earlier representative months. As shown in Figure 10, red regions indicate increased probability

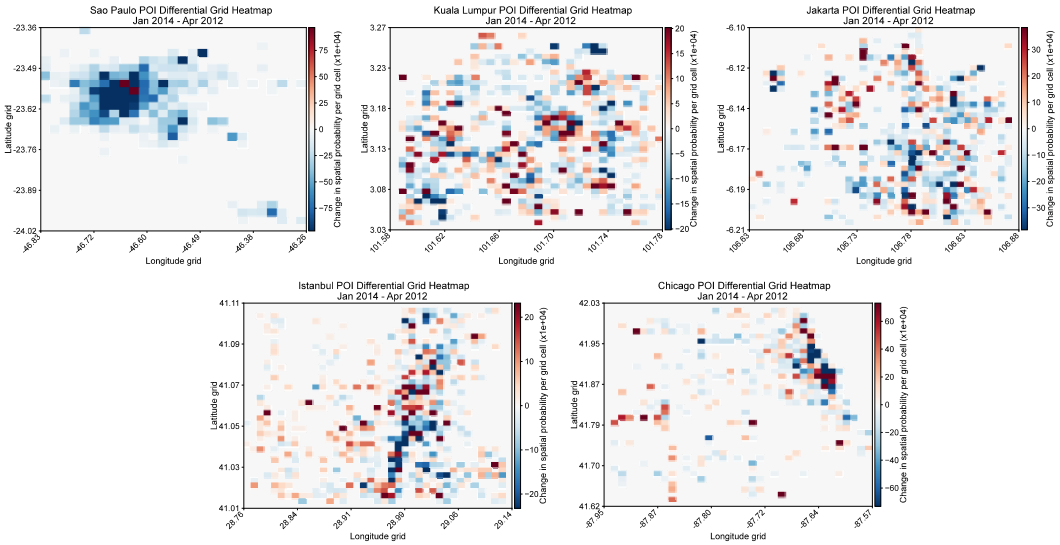

Figure 10: Monthly differential spatial probability (the representation captured by LAB-Bridge) on five city datasets. Red cells represent later-period probability gains, blue cells represent probability losses, and near-white regions indicate relatively weak changes.

mass in the later period, while blue regions indicate decreased probability mass. In several cities, local gains and losses appear alternately, suggesting that temporal change does not simply move a single center but redistributes spatial probability among multiple local regions. This redistribution is more evident in Jakarta and Chicago; in Istanbul, the changes mainly occur inside the core area; in Kuala Lumpur, the increased regions are consistent with the slight northward movement of the hotspot center. For Sao Paulo, the later representative month contains fewer records, so its differential map should be treated mainly as a visual reference for the direction of change.