LapDDPM: Spectral Perturbation Diffusion for Robust Single-Cell Manifold Generation

Lorenzo Bini LORENZO.BINI@UNIGE.CH
Stéphane Marchand-Maillet STEPHANE.MARCHAND-MAILLET@UNIGE.CH
Department of Computer Science, University of Geneva, Geneva, Switzerland.

Abstract

Generating high-fidelity and biologically plausible synthetic single-cell RNA sequencing (scRNA-seq) data is a critical challenge in computational biology, driven by the need to model high-dimensional, sparse, and non-linear cellular manifolds. Existing generative models often fail to capture the complex topology of cellular differentiation or lack robustness against technical noise and structural variability. We introduce LapDDPM, a novel conditional Graph Diffusion Probabilistic Model designed for robust manifold learning and high-fidelity generation. LapDDPM integrates graph-based inductive biases with score-based generative modeling, enhanced by a novel spectral adversarial perturbation mechanism. By systematically perturbing graph edge weights along principal spectral modes during training, our method acts as a Distributionally Robust Optimization (DRO) framework, enforcing invariance to structural noise. We further extend LapDDPM to spatial transcriptomics and multi-modal data, treating generation as a robust inverse problem on cellular graphs. Extensive experiments on diverse datasets, including PBMC3K, Dentate Gyrus, HLCA, Visium, and 10x Multiome, demonstrate that LapDDPM significantly outperforms state-of-the-art baselines in distribution matching, manifold preservation, and downstream utility, generating biologically coherent cell states.

Data and Code Availability This paper uses publicly available scRNA-seq datasets including PBMC3K, Dentate Gyrus, Tabula Muris, and the Human Lung Cell Atlas (HLCA), with corresponding citations provided in the main text. Code is publicly available at https://github.com/LorenzoBini4/laplace-DDPM.

Institutional Review Board (IRB) This study uses only publicly available, de-identified datasets and does not involve human subjects research; therefore, IRB approval was not required.

1. Introduction

Single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of cellular heterogeneity, enabling the characterization of gene expression at an unprecedented resolution (Tang et al., 2009; Saliba et al., 2014). However, the generation of scRNA-seq data is often resource-intensive, limited by sample availability, and inherently noisy due to biological and technical variations (Kiselev et al., 2019). This has spurred significant interest in developing computational methods for generating synthetic scRNA-seq data that closely mimics real biological samples (Eraslan et al., 2019). Such synthetic data can serve various purposes, including data augmentation for rare cell types, benchmarking of computational tools, and in-silico perturbation studies.

Recent advancements in generative modeling, particularly Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), have shown remarkable success in synthesizing high-fidelity data across various domains. Concurrently, flow-based generative models (Dinh et al., 2014, 2017; Palma et al., 2025) have also gained prominence for their ability to learn complex data distributions through invertible transformations, offering exact likelihood estimation. The application of Graph Neural Networks (GNNs) has also gained traction in single-cell biology, leveraging the inherent graph-like structures of cellular relationships (e.g., cell-cell similarity networks or lineage trees) (Zhao et al., 2021; Zhou et al., 2022). Integrating these powerful paradigms presents a promising avenue for generating scRNA-seq data, where GNNs can capture complex cellular interactions and DPMs or flow models can model the intricate gene expression distributions.

Despite these advances, existing generative models for scRNA-seq data often face several limitations. Many struggle with the unique characteristics of scRNA-seq data, such as its high dimensionality, sparsity (due to dropout events), zero-inflation, and complex, non-Gaussian distributions (Kiselev et al., 2019). Furthermore, while some methods incorporate graph information, they may lack mechanisms to ensure robustness against structural noise or subtle variations in cell-cell interaction networks, which are common in real biological contexts. The ability to generate data conditioned on specific biological metadata (e.g., cell type, tissue origin) is also crucial for practical applications, yet remains a challenge for many unconditional models.

To address these limitations, we propose a novel conditional generative framework for scRNA-seq data. LapDDPM (Laplacian Diffusion Denoising Probabilistic Model) integrates graph-based representations with a score-based diffusion model, further enhanced by a unique spectral adversarial perturbation mechanism applied directly to the graph's edge weights. Our method aims to synthesize realistic scRNA-seq count matrices that faithfully reproduce the statistical properties and cellular heterogeneity of real data, conditionally on cell types. Our approach makes three key contributions:

- We propose a **robust graph-based representation learning framework** for scRNA-seq that leverages k-Nearest Neighbors (k-NN) graphs and Laplacian Positional Encodings (LPEs) to inject structural inductive biases into the generative process.
- We develop a **conditional score-based diffusion model** operating in a learned latent space, capable of modeling the complex, high-dimensional manifolds of cellular states while conditioning on biological metadata.
- We introduce a novel **spectral adversarial perturbation mechanism** as a principled Distributionally Robust Optimization (DRO) objective. By perturbing edge weights along the principal spectral modes of the graph, we enforce invariance to structural noise, enhancing generalization across batch effects and biological variability.

Extensive experiments across diverse scRNA-seq datasets, including PBMC3K, Dentate Gyrus, Tabula Muris, and the Human Lung Cell Atlas (HLCA), comprehensively validate LapDDPM's effectiveness. Beyond superior quantitative metrics (RBF-kernel MMD and 2-Wasserstein distance), LapDDPM preserves marker-gene structure and pseudotime ordering while improving downstream utility through stronger TSTR and GRN fidelity, underscoring its value for conditionally synthesizing biologically plausible, cell-type-specific scRNA-seq profiles.

2. Related Work

Our work bridges generative modeling for scRNA-seq, Graph Neural Networks (GNNs), and diffusion models, specifically addressing robust manifold learning.

2.1. Generative Models for scRNA-seq Data

Early approaches adapted general machine learning models like Variational Autoencoders (VAEs) (Lopez et al., 2018; Eraslan et al., 2019) and Generative Adversarial Networks (GANs) (Ghosh et al., 2020; Li et al., 2021) to scRNA-seq. VAEs learn low-dimensional latent representations, often accounting for sparsity, while GANs map simple priors to data distributions via adversarial training. More recently, flow-based models (Chen et al., 2021) offered exact likelihood estimation. However, these methods often struggle to capture intricate topological heterogeneities or ensure robustness against noise. Our work extends these efforts by leveraging graph structures and scoring-based diffusion for fidelity and robustness.

2.2. Graph Neural Networks in Single-Cell Biology

The inherent relational nature of single-cell data, where cells can be viewed as nodes in a graph connected by biological similarity (e.g., gene expression similarity, spatial proximity, or lineage relationships), makes GNNs a natural fit for analysis. GNNs have been applied to various tasks in single-cell biology, including cell type annotation (Wang et al., 2021), trajectory inference (Qi et al., 2021), and spatial transcriptomics analysis (Zhou et al., 2022). These applications typically use GNNs as feature extractors or classifiers. Our work utilizes GNNs within a generative framework, specifically a spectral encoder, to effectively process the graph-structured representation of scRNA-seq data and extract meaningful latent features for the diffusion process.

2.3. Diffusion Probabilistic Models

Diffusion Probabilistic Models (DPMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song and Ermon, 2020) have emerged as a powerful class of generative models, demonstrating state-of-the-art performance across various domains, including image synthesis (Dhariwal and Nichol, 2021) and audio generation (Kong et al., 2021). DPMs define a forward diffusion process that gradually adds noise to data and a reverse process that learns to denoise it. The reverse process is typically modeled by a neural network that estimates the score function (gradient of the log-probability density). Recent advancements, such as score-based generative models (Song and Ermon, 2020), have unified various DPM formulations. In the biological domain, DPMs have shown promise for protein design (Anand et al., 2022) and drug discovery (Hoogeboom et al., 2022). LapDDPM builds upon the success of score-based diffusion models by adapting them to the unique characteristics of scRNA-seq data within a graph-aware latent space, enabling high-fidelity conditional generation.

2.4. Adversarial Training and Robustness on Graphs

Adversarial training has been widely adopted to enhance the robustness of machine learning models against malicious perturbations (Goodfellow et al., 2014). In the context of GNNs, adversarial attacks can target node features or the graph structure itself (e.g., adding or removing edges) (Zügner et al., 2018; Dai et al., 2018). Correspondingly, defense mechanisms include robust aggregation functions (Li et al., 2019) and adversarial training techniques (Kong et al., 2020; Feng et al., 2020; Thorpe et al., 2022). Recent works have explored more efficient adversarial training schemes, such as virtual adversarial training (Zhuo et al., 2023) and diffusion-based approaches (Gosch et al., 2024a). Our work introduces a novel spectral adversarial perturbation mechanism that directly modifies the edge weights of the input graph during training. This approach is distinct from traditional feature or hidden representation perturbations and aims to enhance the encoder's robustness to structural variations, which is particularly relevant for the complex and often noisy graph structures derived from biological data. In Section 4, we assess the robustness of our methods against adversarially poisoned input graphs, focusing on representations learned from graphs compromised by various structural attack strategies, including random attacks, DICE (Waniek et al., 2018), GF-Attack (Chang et al., 2020), and Mettack (Gosch et al., 2024b).

2.5. Spectral Graph Methods

Spectral graph theory provides a powerful framework for analyzing graph properties through the eigenvalues and eigenvectors of graph matrices, such as the adjacency matrix or Laplacian (Chung, 1997). In GNNs, spectral methods have informed the design of convolutional layers (Defferrard et al., 2016) and provided theoretical insights into message passing (Balcilar et al., 2021). More recently, spectral properties have been leveraged for graph augmentation (Ghose et al., 2023) and for improving the robustness of GNNs (Bo et al., 2024). LapDDPM extensively utilizes spectral graph theory by incorporating Laplacian Positional Encodings (LPEs) as node features, which provide crucial structural context to the GNN encoder. Furthermore, our novel spectral adversarial perturbation mechanism is fundamentally rooted in spectral properties, specifically leveraging the principal eigenvector of the graph's adjacency matrix to generate meaningful structural variations. This allows our model to learn robust representations by challenging it with perturbations that target the graph's dominant spectral modes.

Our work integrates these distinct yet complementary research areas, proposing LapDDPM as a unified framework for robust and conditional scRNA-seq data generation. By combining graph-based representations, spectral positional encodings, a score-based diffusion model, and spectral adversarial perturbations, LapDDPM offers a powerful solution for synthesizing high-fidelity biological data.

3. Methodology

We detail the components of LapDDPM, a novel conditional generative framework for single-cell RNA sequencing (scRNA-seq) data. The framework leverages graph-based representations, spectral positional encodings, and a score-based diffusion model to synthesize realistic scRNA-seq count matrices conditioned on cell types. A key innovation is the integration of spectral adversarial perturbations applied directly to the graph structure during training, en-

hancing the model's robustness to structural variations inherent in biological data.

LapDDPM comprises three primary components: (1) a graph-based data representation module that constructs a k-NN graph from scRNA-seq data and computes LPEs, (2) a spectral encoder-decoder pair that operates on graph-structured data, and (3) a conditional score-based diffusion model that learns to generate latent representations. The overall training procedure combines diffusion, reconstruction, and KL divergence losses, with the encoder being trained on graphs perturbed by a novel spectral adversarial mechanism.

3.1. Graph-based scRNA-seq Data Representation

Given a scRNA-seq dataset consisting of N cells and D genes, represented as a count matrix $\mathbf{X} \in \mathbb{R}^{N \times D}$, we first preprocess the data to construct a graph $\mathcal{G} = (V, E)$ where nodes V correspond to individual cells and edges E represent cellular proximity.

Gene Filtering and Normalization: Prior to graph construction, genes expressed in fewer than a specified threshold of cells are filtered out to reduce sparsity and computational burden. The raw count data is then normalized and log-transformed for stable numerical operations during feature extraction.

Laplacian Positional Encoding (LPE): Firstly, a k-NN graph is constructed on the cells. To capture biologically meaningful relationships and reduce the dimensionality of the feature space for graph construction, Principal Component Analysis (PCA) is applied to the log-transformed gene expression data. The k-NN graph is then built using Euclidean distance in this PCA-reduced space, preserving local micro-environments, including rare populations, while reducing hub-node bias in noisy count space. For each cell (node), its k nearest neighbors are identified, and edges are formed between them. The resulting adjacency matrix is denoted as $\mathbf{A} \in \{0, 1\}^{N \times N}$.

Then, to provide the GNN backbone with positional information and enhance its ability to distinguish nodes based on their structural roles, we compute LPEs. The LPEs are derived from the eigenvectors of the normalized graph Laplacian. The normalized Laplacian $\mathbf{L}_{\text{norm}}$ is defined as:

$$\mathbf{L}_{\text{norm}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{D}$ is the diagonal degree matrix of $\mathbf{A}$. We compute the first k non-trivial eigenvectors of $\mathbf{L}_{\text{norm}}$ (excluding the trivial eigenvector corresponding to eigenvalue 0 for connected graphs), which form the LPE matrix $\mathbf{P} \in \mathbb{R}^{N \times k}$. These LPEs are then concatenated with the gene expression features for the encoder.

3.2. LapDDPM Architecture

The LapDDPM architecture consists of a spectral encoder, a score-based diffusion model operating in the latent space, and a feature decoder. Algorithms 1 and 2 describe both training and generation procedure in detail of our model. Figure 1 provides a high-level schematic of LapDDPM, highlighting the dual-graph construction, Laplacian positional encodings, spectral/transformer encoder, and latent diffusion pathway with a ZINB decoder. The figure emphasizes the two-stage training schedule and the spectral adversarial perturbation that regularizes the graph spectrum during encoding, aligning with the model description and algorithmic steps presented in this section.

3.2.1. Spectral Encoder

The spectral encoder, denoted as E_ϕ, is responsible for mapping the input gene expression data $\mathbf{X}$ and its associated graph structure to a latent space. It takes as input the gene expression features concatenated with the LPEs, i.e., $[\mathbf{X}, \mathbf{P}] \in \mathbb{R}^{N \times D_f}$ where D_f is the number of filtered genes, and the graph's edge index along with dynamically perturbed edge weights. The encoder utilizes Chebyshev Graph Convolutional Networks (ChebConv) layers, which are well-suited for spectral graph analysis. The encoder outputs the mean $\mu \in \mathbb{R}^{N \times d_{\text{lat}}}$ and log-variance $\log \sigma^2 \in \mathbb{R}^{N \times d_{\text{lat}}}$ of a Gaussian distribution in the latent space, where d_{lat} is the latent dimension. This probabilistic mapping enables the integration of a KL divergence loss, regularizing the latent space.

3.2.2. Score-based Diffusion Model

The core generative component of LapDDPM is a conditional score-based diffusion model, which operates in the latent space learned by the encoder. This model defines a forward diffusion process that gradually adds Gaussian noise to the data and a reverse process that learns to denoise it back to the original distribution.

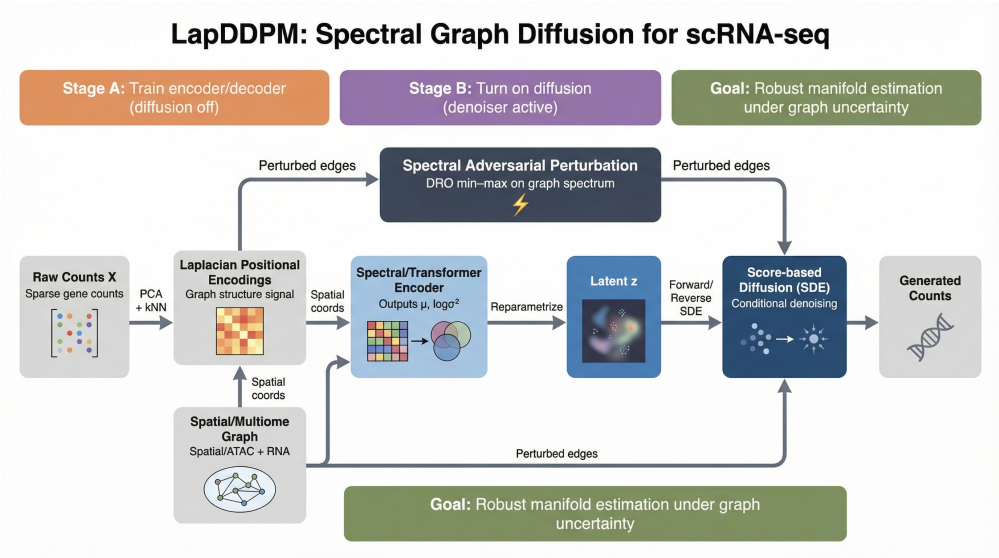


Figure 1: High-level architecture of LapDDPM, showing the dual-graph inputs, LPEs, encoder, latent diffusion components to generate biologically-grounded scRNA counts.

Forward Diffusion Process: We employ a Variance-Preserving Stochastic Differential Equation (VP-SDE) for the forward process, which gradually transforms a latent variable $\mathbf{z}_0 \sim p_0(\mathbf{z})$ into a noisy latent variable $\mathbf{z}_t$ at time $t \in [0, T]$. The marginal distribution $p_t(\mathbf{z}_t|\mathbf{z}_0)$ is a Gaussian distribution $\mathcal{N}(\alpha_t \mathbf{z}_0, \sigma_t^2 \mathbf{I})$, where $\alpha_t = e^{-t}$ and $\sigma_t^2 = 1 - e^{-2t}$. The *marginal_std* function in the code computes $\sigma_t = \sqrt{1 - e^{-2t}}$.

Reverse Denoising Process (ScoreNet): The reverse process involves learning the score function $s_\theta(\mathbf{z}_t, t, \mathbf{c}) = \nabla_{\mathbf{z}_t} \log p_t(\mathbf{z}_t|\mathbf{c})$, which indicates the direction of density increase. This score function is approximated by a neural network, referred to as `ScoreNet` in the code. The `ScoreNet` is a multi-layer perceptron (MLP) that takes as input the noisy latent variable $\mathbf{z}_t$, the current time t (embedded via a separate MLP), and a conditional cell type label $\mathbf{c}$ (embedded via an embedding layer). It predicts the noise $\boldsymbol{\epsilon}$ that was added to $\mathbf{z}_0$ to obtain $\mathbf{z}_t$. Layer Normalization is applied within the `ScoreNet` to ensure stable training. During training, a conditional dropout mechanism is applied to the cell type embedding to encourage unconditional generation capabilities and robustness under label imbalance.

Sampling (ScoreSDE) & Feature Decoder: The `ScoreSDE` module implements the reverse diffusion process for sampling. Starting from a pure noise vector $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$, the model iteratively applies small denoising steps guided by the predicted score function from the `ScoreNet`. This process gradually transforms the noise into a meaningful latent representation $\mathbf{z}_0$. The feature decoder, D_ψ, is an MLP that transforms the generated latent representations $\mathbf{z}_\epsilon$ back into gene expression log-rates. Specifically, it outputs $\log(\text{rates}) \in \mathbb{R}^{N \times D_{\text{filtered}}}$, where D_{filtered} is the number of genes after filtering. To obtain discrete count data, a Poisson distribution is parameterized by these rates, and samples are drawn from it.

3.3. Spectral Adversarial Perturbations on Graph Structure

We formulate graph learning as a Distributionally Robust Optimization (DRO) problem, where the model must perform well not just on the observed graph G_{obs}, but on a family of perturbed graphs $\{G' : \Delta(G', G_{obs}) \leq \epsilon\}$:

$$\min_{\theta} \max_{\|\Delta\|_F \leq \epsilon} \mathcal{L}(f_\theta(G_{\text{obs}} + \Delta)). \tag{1}$$

We approximate the inner maximization on each batch-local graph by targeting its dominant spectral mode.

Let $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ be the normalized Laplacian. We focus on perturbing edge weights w_{ij}

to maximize the deviation in the principal spectral components. Our perturbation $\boldsymbol{\delta}$ is computed via a differentiation-through-eigencomposition approximation:

$$\delta_{ij} \propto \nabla_{w_{ij}} \lambda_1(\mathbf{L}) \approx \mathbf{v}_1[i]\mathbf{v}_1[j] \tag{2}$$

where $\mathbf{v}_1$ is the principal eigenvector of $\mathbf{L}$. The gradient term $\nabla_{w_{ij}}\lambda_1$ captures the sensitivity of the graph's dominant structural mode to edge-weight variations. By iteratively updating $\mathbf{W} \leftarrow \mathbf{W} + \alpha\boldsymbol{\delta}$ to maximize spectral distortion, and subsequently optimizing the encoder to minimize reconstruction error under this perturbation, we smooth the encoder's sensitivity to high-frequency structural noise. In practice, this concentrates the adversary on the globally most influential deformation direction while keeping the overhead modest; overly large ϵ eventually induces over-regularization, as confirmed by the ablations in Section B.

3.4. Training Objective and Procedure

LapDDPM is trained end-to-end by minimizing a combined loss function that integrates three objectives: a diffusion loss, a KL divergence loss, and a reconstruction loss.

Total Loss Function: The overall objective function $\mathcal{L}_{\text{Total}}$ is defined as:

$$\mathcal{L}_{\text{Total}} = w_{\text{diff}}\mathcal{L}_{\text{diff}} + w_{\text{KL}}\mathcal{L}_{\text{KL}} + w_{\text{rec}}\mathcal{L}_{\text{rec}}$$

where w_{diff}, w_{KL}, and w_{rec} are learnable weights controlling the contribution of each component. The formers have been implemented as follow:

1. **Diffusion Loss ($\mathcal{L}_{\text{diff}}$):** This is the primary loss for the score-based diffusion model. It measures the Mean Squared Error (MSE) between the noise predicted by the 'ScoreNet' ($\hat{\boldsymbol{\epsilon}}$) and the actual noise $\boldsymbol{\epsilon}$ added during the forward diffusion process:

   $$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\mathbf{z}_0,\boldsymbol{\epsilon}}\left[\|\hat{\boldsymbol{\epsilon}}(\mathbf{z}_t, t, \mathbf{c}) - \boldsymbol{\epsilon}\|_2^2\right]$$

   where $\mathbf{z}_t = \alpha_t\mathbf{z}_0 + \sigma_t\boldsymbol{\epsilon}$ is the noisy latent variable.

2. **KL Divergence Loss ($\mathcal{L}_{\text{KL}}$):** This term regularizes the latent distribution produced by the encoder. It computes the Kullback-Leibler divergence between the encoder's output distribution $\mathcal{N}(\mu, \exp(\log\sigma^2))$ and a standard normal prior $\mathcal{N}(0, \mathbf{I})$:

   $$\mathcal{L}_{\text{KL}} = \frac{1}{2}\sum_{i=1}^{d_{\text{lat}}}(\mu_i^2 + \exp(\log\sigma_i^2) - 1 - \log\sigma_i^2)$$

   This loss encourages the latent space to be well-structured and facilitates sampling from a simple prior during generation.

3. **Reconstruction Loss ($\mathcal{L}_{\text{rec}}$):** This loss ensures that the model can accurately reconstruct the original gene expression data from its latent representation. Given the decoded log-rates from the `FeatureDecoder`, we use the Poisson Negative Log-Likelihood (NLL) loss against the orig-

Algorithm 1 LapDDPM Training Procedure

Input: scRNA-seq dataset $\mathcal{D} = \{(\mathbf{X}_i, \mathbf{c}_i)\}_{i=1}^N$, epochs E, batch size B, learning rate η, loss weights $w_{\text{diff}}, w_{\text{KL}}, w_{\text{rec}}$, input masking fraction m, perturbation parameters $\alpha_{\min}, \alpha_{\max}, \epsilon$, ip.
Initialize: Spectral Encoder E_ϕ, ScoreNet S_θ, Feature Decoder D_ψ. Optimizer, LR scheduler, GradScaler.
for epoch = 1 to E **do**
  for batch $(\mathbf{X}_{\text{batch}}, \mathbf{c}_{\text{batch}})$ **do**
    k-NN graph and compute LPE $\mathbf{P}_{\text{batch}}$ for $\mathbf{X}_{\text{batch}}$.
    Original batch features $\mathbf{X}_{\text{orig}} \leftarrow \mathbf{X}_{\text{batch}}$.
    Apply input masking: $\mathbf{X}_{\text{masked}} \leftarrow \text{Mask}(\mathbf{X}_{\text{batch}}, m)$.
    Get current edge weights $\mathbf{W}_{\text{batch}}$ from k-NN graph.
    Spectral Adversarial Perturbation:
    $\mathbf{W}_{\text{adv}} \leftarrow \text{LaplacianPerturb}(E_\phi, \mathbf{X}_{\text{masked}}, \mathbf{W}_{\text{batch}}, \epsilon, ip)$.
    Forward Pass (Encoder):
    $(\mu, \log\sigma^2) \leftarrow E_\phi([\mathbf{X}_{\text{masked}}, \mathbf{P}_{\text{batch}}], \mathbf{W}_{\text{adv}})$.
    Sample $\mathbf{z}_0 \sim \mathcal{N}(\mu, \exp(\log\sigma^2))$.
    Diffusion Loss Calculation:
    Sample time $t \sim \mathcal{U}(0, T)$, noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$.
    $\mathbf{z}_t \leftarrow \alpha_t\mathbf{z}_0 + \sigma_t\boldsymbol{\epsilon}$.
    $\hat{\boldsymbol{\epsilon}} \leftarrow S_\theta(\mathbf{z}_t, t, \mathbf{c}_{\text{batch}})$.
    $\mathcal{L}_{\text{diff}} \leftarrow \text{MSE}(\hat{\boldsymbol{\epsilon}}, \boldsymbol{\epsilon})$.
    KL Divergence Loss Calculation:
    $\mathcal{L}_{\text{KL}} \leftarrow \text{KL}(\mathcal{N}(\mu, \exp(\log\sigma^2)) || \mathcal{N}(0, \mathbf{I}))$.
    Reconstruction Loss Calculation:
    $\log(\text{rates}) \leftarrow D_\psi(\mu)$.
    $\mathcal{L}_{\text{rec}} \leftarrow \text{PoissonNLL}(\log(\text{rates}), \mathbf{X}_{\text{orig}})$.
    Total Loss:
    $\mathcal{L}_{\text{Total}} \leftarrow w_{\text{diff}}\mathcal{L}_{\text{diff}} + w_{\text{KL}}\mathcal{L}_{\text{KL}} + w_{\text{rec}}\mathcal{L}_{\text{rec}}$.
    Backward Pass and Optimization:
    Backpropagate $\mathcal{L}_{\text{Total}}$, clip gradients, update parameters, step LR scheduler.
  end for
end for
Output: Trained E_ϕ, S_θ, D_ψ.

Algorithm 2 LapDDPM Generation Procedure

Input: Number of samples N_{gen}, conditional cell type labels $\mathbf{c}_{\text{gen}}$, trained ScoreSDE S_{SDE}, trained Feature Decoder D_ψ.
Initialize: Random noise $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$ of shape $(N_{\text{gen}}, d_{\text{lat}})$.
Latent Space Denoising:
$\mathbf{z}_{\text{generated}} \leftarrow S_{\text{SDE.sample}}(\mathbf{z}_T, \mathbf{c}_{\text{gen}})$.
Feature Decoding:
$\log(\text{rates}) \leftarrow D_\psi(\mathbf{z}_{\text{generated}})$.
Count Sampling:
$\mathbf{X}_{\text{generated}} \sim \text{Poisson}(\exp(\log(\text{rates})))$.
Output: Generated scRNA-seq counts $\mathbf{X}_{\text{generated}}$ and corresponding cell types $\mathbf{c}_{\text{gen}}$.

inal (unmasked) gene counts:

$$\mathcal{L}_{\text{rec}} = -\frac{1}{N \cdot D_{\text{filtered}}} \sum_{i=1}^{N} \sum_{j=1}^{D_{\text{filtered}}} \Big(\mathbf{x}_{ij} \log(\text{rates}_{ij}) - \text{rates}_{ij} - \log(\Gamma(\mathbf{x}_{ij} + 1)) \Big)$$

where $\mathbf{x}_{ij}$ are the original gene counts and $\text{rates}_{ij} = \exp(\log(\text{rates})_{ij})$.

Input Gene Masking: During training, a fraction of the input gene expression features can be randomly masked (set to zero). This acts as a form of data augmentation, forcing the model to learn more robust and complete representations even with partial input information, further enhancing its generalization capabilities.

4. Experiments: scRNA-seq Generation

We conduct a comprehensive empirical evaluation of LapDDPM's capability in generating realistic single-cell RNA sequencing (scRNA-seq) data, both conditionally and unconditionally. Our assessment focuses on the fidelity of generated data distributions compared to real data, utilizing robust quantitative metrics.

Baselines: We compare LapDDPM against several state-of-the-art generative models for scRNA-seq data. For conditional generation, our baselines include scVI (Gayoso et al., 2021), scDiffusion (Luo et al., 2024), and CFGen (Palma et al., 2025). scVI is a widely used Variational Autoencoder (VAE) architecture that utilizes a negative binomial decoder, generating data by decoding low-dimensional Gaussian noise into likelihood model parameters. scDiffusion, on the other hand, is a continuous-space model based on standard latent diffusion (Rombach et al., 2022). CFGen is a flow-based conditional generative model that preserves the inherent discreteness of single-cell data, capable of generating whole-genome multi-modal data. For unconditional generation, we compare against scGAN (Marouf et al., 2020), a Generative Adversarial Network (GAN) designed for scRNA-seq. Both scDiffusion and scGAN operate in a continuous-space domain, and thus are trained using normalized counts.

4.1. Evaluation Protocol

For quantitative evaluation, we employ two widely recognized distribution distances: the RBF-kernel Maximum Mean Discrepancy (MMD) (Borgwardt et al., 2006) and the 2-Wasserstein distance. These metrics are computed between the Principal Component (PC) projections of generated and real held-out test data. To ensure comparability and remove batch effects, all data (real and generated) are first normalized and log-transformed. Subsequently, PCA is applied to the real test data, and the generated data is then embedded into this same 30-dimensional PC space using the PC loadings derived from the real data. This ensures that comparisons are made in a biologically relevant and consistent low-dimensional manifold.

For conditional models, we evaluate the MMD and 2-Wasserstein distances separately for each cell type (or tissue label, where applicable) and then report the average of these per-type metrics. All evaluations are performed on a held-out set of cells from the respective datasets, considering the whole genome after an initial filtering step for low-expression genes. Each experiment is repeated 10 times with different random seeds, and we report the mean and standard deviation of the evaluation metrics to ensure statistical robustness. In Table 1, the gains of LapDDPM over the strongest baselines generally exceed the combined standard deviation envelopes, supporting empirical significance across datasets.

4.2. Generative Performance

We assess LapDDPM's performance on four diverse scRNA-seq datasets, varying significantly in size and biological complexity. These include PBMC3K (10x

Table 1: Quantitative performance comparison of LapDDPM with conditional and unconditional single-cell generative models. Evaluation is performed based on distribution matching metrics (RBF-kernel MMD and 2-Wasserstein distance). Results are mean ± standard deviation over 10 runs. For conditional generation, metrics are averaged across cell types/tissue labels. Best results are highlighted in bold.

Model	PBMC3K		Dentate gyrus		Tabula Muris		HLCA	
	MMD (↓)	WD (↓)	MMD (↓)	WD (↓)	MMD (↓)	WD (↓)	MMD (↓)	WD (↓)
Conditional								
c-CFGen	0.85±0.05	16.94±0.44	1.12±0.04	21.55±0.17	**0.19±0.02**	7.39±0.20	0.54±0.02	10.72±0.08
scDiffusion	1.27±0.20	22.41±1.21	1.22±0.05	22.56±0.10	0.24±0.04	7.89±0.45	0.96±0.04	15.82±0.45
SCVI	0.94±0.05	17.66±0.29	1.15±0.04	22.61±0.23	0.26±0.02	9.76±0.53	0.58±0.02	11.78±0.19
LapDDPM (ours)	**0.41±0.15**	**14.84±0.83**	**1.04±0.08**	**18.74±0.43**	**0.19±0.02**	**7.04±0.14**	**0.39±0.03**	**8.74±0.08**
Unconditional								
u-CFGen	0.44±0.01	16.81±0.06	0.42±0.01	21.20±0.02	**0.08±0.00**	8.54±0.06	0.15±0.01	10.63±0.01
SCGAN	0.36±0.01	15.54±0.06	0.42±0.01	22.52±0.03	0.25±0.00	12.85±0.04	0.18±0.01	10.81±0.01
LapDDPM (ours)	**0.23±0.02**	**13.98±0.74**	**0.38±0.02**	**18.11±0.02**	0.15±0.01	**7.33±0.02**	**0.14±0.02**	**8.31±0.01**

Genomics, 2017), a foundational dataset for single-cell analysis; Dentate Gyrus (La Manno et al., 2018), providing a more complex biological context; Tabula Muris (Consortium et al., 2018), a large-scale dataset featuring cells across multiple tissues; and the Human Lung Cell Atlas (HLCA) (Sikkema et al., 2023), representing a highly complex and heterogeneous real-world scenario. Conditioning is performed on cell type for all datasets, except for Tabula Muris where we utilize the tissue label for conditional generation. Detailed descriptions of each dataset and their specific pre-processing steps are provided and summarized in Table 3.

Our experimental results confirm that LapDDPM captures the subtle topology of cellular manifolds rather than merely matching low-dimensional marginals. In Table 1, LapDDPM consistently improves Wasserstein distance across datasets. In **PBMC3K**, the model preserves distinct marker-gene structure for CD4+ T-cells and B-cells without hallucinated intermediates; in **Dentate Gyrus**, it preserves the pseudotime ordering from neural progenitors to granule cells; and in **HLCA**, the WD reduction from 11.78 (scVI) and 15.82 (scDiffusion) to 8.74 lies well outside the corresponding standard deviation envelopes, supporting empirical significance under strong donor heterogeneity. Together with the TSTR and GRN gains in Table 2 and the diagnostics in Figure 2, these results indicate that the learned conditional distribution remains biologically meaningful for downstream analyses.

4.3. Spatial and Multiome Experiments

We extended LapDDPM to spatial transcriptomics Visium Lymph Node (10xGenomics, 2019) and multiome lymphoma 14k data (10xGenomics, 2020), viewing these tasks as *robust manifold estimation* problems where the goal is to infer the underlying cellular manifold $\mathcal{M}$ from sparse, noisy observations and inconsistent graph structures (e.g., spatial vs. expression graphs). This framing necessitates robustness against structural uncertainty, which our spectral adversarial training explicitly addresses.

Baselines. We benchmark LapDDPM against scVI (Gayoso et al., 2021) (VAE-based) and cellDART (Bae et al., 2022) (adversarial pseudospots learning-based), adapting them for spatial/multi-modal tasks. As shown in Table 2, LapDDPM drastically outperforms both baselines, often by an order of magnitude in reconstruction metrics.

- Spatial Preservation: LapDDPM achieves distinctly higher Moran's I (0.31 vs. 0.10 for scVI), indicating that while baselines largely fail to capture spatial autocorrelation, our model preserves tissue domains.
- Structural Robustness: Our model exhibits the lowest Spectral Mismatch (15 vs. 38-42), confirming that spectral adversarial perturbations are essential for aligning the synthesized graph spectrum with the physical ground truth.
- Downstream Utility: Inpainting MSE drops to 0.07 (vs. $>$ 2.0 for baselines), while TSTR ac-

Table 2: Spatial and multiome evaluation results. We compare LapDDPM against scVI (Gayoso et al., 2021) and cellDART (Bae et al., 2022). LapDDPM demonstrates superior performance in structural metrics (Moran's I, Spectral Mismatch) and downstream utility (TSTR). Lower is better for MSE, Spectral Mismatch, and GRN Diff Norm; higher is better for correlations and TSTR scores.

	Spatial (Visium)			Multiome (10x 14k)		
Metric	scVI	cellDART	**LapDDPM**	scVI	cellDART	**LapDDPM**
Inpainting MSE (↓)	2.47	5.09	**0.07**	3.10	4.08	**0.06**
Inpainting Corr (↑)	0.71	0.52	**0.91**	0.68	0.59	**0.85**
Moran's I (↑)	0.10	0.17	**0.31**	–	–	–
Spectral Mismatch (↓)	42	38	**15**	–	–	–
TSTR Accuracy (↑)	0.65	0.49	**0.92**	0.55	0.48	**0.91**
TSTR F1 (↑)	0.62	0.55	**0.87**	0.53	0.56	**0.88**
GRN Spearman (↑)	0.28	0.34	**0.61**	0.15	0.22	**0.59**
GRN Diff Norm (↓)	1.52	3.39	**0.34**	2.01	3.46	**0.41**

curacy reaches 0.92 and GRN Spearman reaches 0.61. This indicates that the learned manifold is both predictive and regulatory coherent, not only visually or distributionally plausible.

We expressly omit Moran's I and Spectral Mismatch for the standard 10x Multiome dataset. Unlike the Visium platform which preserves tissue architecture, standard Multiome protocols involve tissue dissociation, meaning there are no ground-truth physical spatial coordinates or physical graph topology against which to evaluate these structure-dependent metrics.

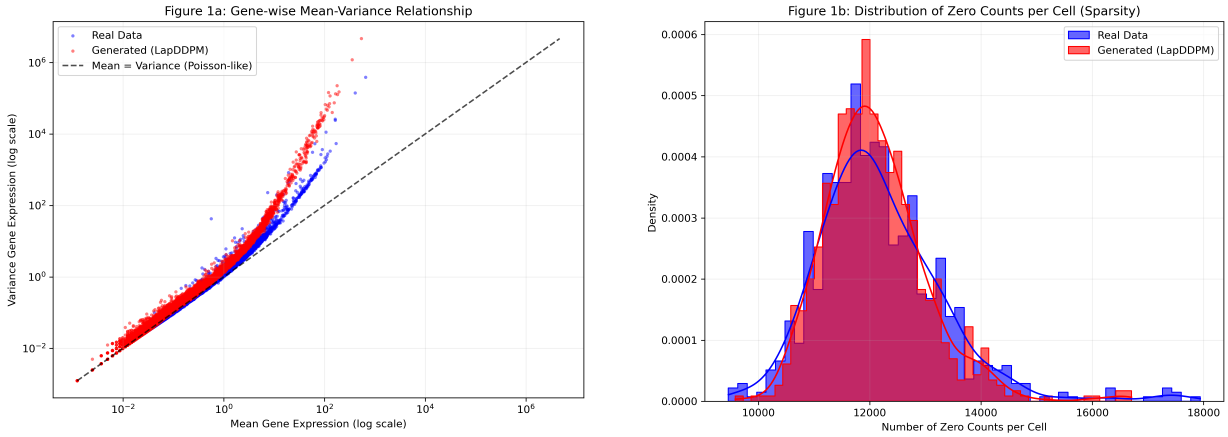
Qualitative diagnostics in Figure 2 provide critical validation of the model's fidelity. The mean-variance relationship (a) confirms that LapDDPM (red) faithfully tracks the non-linear dispersion of real scRNA-seq data (blue) across the full dynamic range. Crucially, it captures the high variance in highly expressed genes, a hallmark of biological overdispersion that standard Gaussian-based generative models often smooth out. This preservation is vital for downstream differential expression analysis, where underestimated variance leads to inflated False Discovery Rates (FDR). Furthermore, the zeros-per-cell distribution (b) demonstrates that LapDDPM accurately reproduces the complex, bimodal sparsity patterns of the data, distinguishing between technical dropouts and true biological zeros. By matching this zero-inflation frequency without explicit heuristics, the model proves it has learned the underlying stochastic generative process rather than simply memorizing mean expression levels, ensuring that generated cells are statistically indistinguishable from real biological samples.

Architecture Refinements. To achieve these results, we introduced specific refinements: (1) a global transformer encoder for long-range interactions, (2) a ZINB decoder with explicit library-size modeling, and (3) a two-stage training schedule (Stage A: reconstruction, Stage B: diffusion) to stabilize learning on these complex dual-graph problems.

Spectral Adversarial Perturbations as DRO. The spectral edge-weight perturbation in LapDDPM can be viewed as a distributionally robust optimization (DRO) objective: we train the estimator to minimize expected reconstruction and diffusion losses under the worst-case spectral deformation of the input graph. Formally, we optimize a min-max objective over parameters and adversarial spectral shifts, which encourages invariance to structural noise and batch effects. This framing aligns the perturbation mechanism with robust manifold estimation under graph uncertainty and is reflected empirically by reduced spectral mismatch and improved downstream utility (TSTR/GRN).

5. Complexity and Ablation Summary

We conducted a theoretical complexity analysis and extensive ablation studies to validate LapDDPM's design. The key practical point is that spectral perturbation is computed only on sparse batch-local k-NN graphs, so its cost scales as $\mathcal{O}(N_b \cdot k_{\text{nn}})$ and is independent of the total atlas size. On HLCA, this translates to only a modest training overhead (191.4 min for LapDDPM vs. 187.6 min for scDiffusion), while inference remains effectively unchanged (4.1s vs. 4.3s for 10k cells). Ablation studies on PBMC3K (see Section B and Table 4) further show that spectral adversarial perturbations, input masking, and LPEs each contribute materially to performance; removing



(a) Mean–variance relationship.

(b) Zeros-per-cell distribution.

Figure 2: Qualitative diagnostics for spatial generation. (a) Mean–variance alignment indicates improved dispersion matching at higher expression levels. (b) Zeros-per-cell distribution shows improved sparsity alignment.

the spectral adversarial module degrades Wasserstein distance by ∼11% (14.84 → 16.50). Detailed results and complexity derivations are provided in Section B.

6. Conclusion and Future Work

In this work, we introduced LapDDPM, a novel conditional generative framework for single-cell RNA sequencing (scRNA-seq) data. LapDDPM distinguishes itself by integrating graph-based data representations with a score-based diffusion model, further enhanced by a unique spectral adversarial perturbation mechanism applied directly to the graph's edge weights. Our methodology effectively addresses the challenges of generating high-fidelity, biologically plausible scRNA-seq data, conditionally on cellular metadata.

The core advancements of LapDDPM include:

1. Leveraging Laplacian Positional Encodings (LPEs) to enrich the latent space with crucial cellular relationship information.

2. Developing a conditional score-based diffusion model for effective learning and generation from complex scRNA-seq distributions.

3. Employing a unique spectral adversarial training scheme on graph edge weights, boosting robustness against structural variations.

Our extensive experimental evaluation across diverse scRNA-seq datasets rigorously validates LapDDPM's effectiveness. Quantitatively, the model consistently achieves low RBF-kernel Maximum Mean Discrepancy (MMD) and 2-Wasserstein distances between generated and real data distributions in the PCA-projected space, while also improving biological validation and downstream utility through stronger pseudotime preservation, TSTR, and GRN fidelity. The successful conditional generation further highlights LapDDPM's utility in synthesizing cell-type or tissue-specific profiles, which is crucial for targeted biological investigations.

Looking forward, several promising avenues emerge for future research. The modular nature of LapDDPM allows the score-based diffusion model to be readily adapted or replaced with emerging generative paradigms, such as Flow Matching or Schrödinger Bridge matching, to better target low-density regions in the manifold. We also aim to explore further optimizations in memory usage and training efficiency for extremely large-scale datasets, potentially involving sub-sampling strategies or distributed training paradigms. Extending LapDDPM to generate multi-modal single-cell data, such as simultaneous gene expression and chromatin accessibility, would significantly broaden its utility in comprehensive single-cell genomics. Additionally, investigating more granular or continuous conditioning mech-

anisms, like conditioning on cell states or perturbation effects, could enable more sophisticated in-silico experiments. The integration of these research directions holds the potential to evolve LapDDPM into an even more powerful and versatile tool for advancing single-cell genomics research and its applications.

Acknowledgments

This work is partly funded by the Swiss National Science Foundation under grant number 207509 "Structural Intrinsic Dimensionality".

References

10x Genomics. PBMC3K dataset, 2017. A widely used single-cell RNA sequencing dataset from 3k Peripheral Blood Mononuclear Cells.

10xGenomics. Human lymph node, spatial gene expression dataset. 10x Genomics, December 2019. URL https://www.10xgenomics.com/datasets/human-lymph-node-1-standard-1-0-0.

10xGenomics. Flash-frozen lymph node with b cell lymphoma (14k sorted nuclei), epi multiome atac + gene expression dataset. 10x Genomics, October 2020. Dataset page.

Nitya Anand, Kory Achim, Ryan Eguchi, Jay Ho, Young-Sun Kim, Yihui Li, and Adam Rives. Protein diffusion models for generation and design. *Advances in Neural Information Processing Systems*, 35:2721–2733, 2022.

Sungwoo Bae, Kwon Joong Na, Jaemoon Koh, Dong Soo Lee, Hongyoon Choi, and Young Tae Kim. Celldart: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. *Nucleic acids research*, 50(10):e57–e57, 2022.

Muhammet Balcilar, Guillaume Renton, Pierre Héroux, Benoit Gaüzère, Sébastien Adam, and Paul Honeine. Analyzing the expressive power of graph neural networks in a spectral perspective. In *International Conference on Learning Representations*, 2021.

Deyu Bo, Yuan Fang, Yang Liu, and Chuan Shi. Graph contrastive learning with stable and scalable spectral encoding. *Advances in Neural Information Processing Systems*, 36, 2024.

Karsten M Borgwardt, Arthur Gretton, Bernhard Schölkopf, Alex J Smola, Hans-Peter Kriegel, and S V N Vishwanathan. Integrating knowledge from heterogeneous data sources: an application to protein-protein interaction prediction. In *Proceedings of the 2006 ACM symposium on Applied computing*, pages 535–539, 2006.

Heng Chang, Yu Rong, Tingyang Xu, Wenbing Huang, Honglei Zhang, Peng Cui, Wenwu Zhu, and Junzhou Huang. A restricted black-box adversarial framework towards attacking graph embedding models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3389–3396, 2020.

Tian Chen, Jian Chen, Yuting Guo, and Hongbo Wang. Flow-sc: A normalizing flow-based deep generative model for single-cell rna sequencing data. *Bioinformatics*, 37(23):4353–4360, 2021.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Tabula Muris Consortium et al. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, 2018.

Hanjun Dai, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, and Le Song. Adversarial attack on graph structured data. In *International conference on machine learning*, pages 1115–1124. PMLR, 2018.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8789–8802, 2021.

Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *International Conference on Learning Representations*, 2017.

Gökçe Eraslan, Lukas M Simon, Mona Mircea, Nicolas S Mueller, and Fabian J Theis. Single-cell rna-seq imputation and denoising using a deep count autoencoder. *Nature communications*, 10(1):3907, 2019.

Wenzheng Feng, Jie Zhang, Yuxiao Dong, Yu Han, Huanbo Luan, Qian Xu, Qiang Yang, Evgeny Kharlamov, and Jie Tang. Graph random neural

networks for semi-supervised learning on graphs. *Advances in neural information processing systems*, 33:22092–22103, 2020.

Adam Gayoso, Romain Lopez, Pierre Boyeau, Jérémie Chëneby, Michael Li, Nir Yosef, Michael I Jordan, Oliver Stegle, Fabian J Theis, and Davide Risso. scVI: Probabilistic modeling of single-cell omics data in python. *Nature methods*, 18(2):165–172, 2021.

Amur Ghose, Yingxue Zhang, Jianye Hao, and Mark Coates. Spectral augmentations for graph contrastive learning. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 11213–11266. PMLR, 25–27 Apr 2023. URL https://proceedings.mlr.press/v206/ghose23a.html.

Prachi Ghosh, Anup Datta, Sasthi Mukhopadhyay, Debarka Roy, and Anirban Das. Generating synthetic single-cell rna-seq data using generative adversarial networks. *Bioinformatics*, 36(7):2119–2127, 2020.

Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

Lukas Gosch, Simon Geisler, Daniel Sturm, Bertrand Charpentier, Daniel Zügner, and Stephan Günnemann. Adversarial training for graph neural networks: Pitfalls, solutions, and new directions. *Advances in Neural Information Processing Systems*, 36, 2024a.

Lukas Gosch, Simon Geisler, Daniel Sturm, Bertrand Charpentier, Daniel Zügner, and Stephan Günnemann. Adversarial training for graph neural networks: Pitfalls, solutions, and new directions. *Advances in Neural Information Processing Systems*, 36, 2024b.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Emiel Hoogeboom, Victor Garcia Satorras, Clement Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. *International Conference on Machine Learning*, 2022.

Vladimir Y Kiselev, Todd S Andrews, and Martin Hemberg. Challenges in computational analysis of single-cell rna sequencing data. *Nature Reviews Genetics*, 20(10):573–582, 2019.

Kezhi Kong, Guohao Li, Mucong Ding, Zuxuan Wu, Chen Zhu, Bernard Ghanem, Gavin Taylor, and Tom Goldstein. Flag: Adversarial data augmentation for graph neural networks. *arXiv preprint arXiv:2010.09891*, 2020.

Zhifeng Kong, Wei Ping, Kexin Huang, Kuan Zhao, Zihang Li, and Bryan Catanzaro. Diffwave: A diffusion probabilistic model for text-to-speech. *International Conference on Learning Representations*, 2021.

Gioele La Manno, Roman Soldatov, Amit Zeisel, Elisabeth Braun, Hannah Hochgerner, Florian Mayer, Kostas Meletis, Anna Zeisel, Martin Enge, Christer Betsholtz, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

Guiyin Li, Ji Han, Xiao Wu, Yang Tang, and Zhibin Liu. Robust graph convolutional networks against adversarial attacks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1273–1282, 2019.

Jin Li, Peng Zhang, Lin Wu, Shaowu Ma, and Longxiang Chen. Adversarial neural networks for noise removal in single-cell rna-seq data. *Bioinformatics*, 37(10):1398–1405, 2021.

Francisco Lopez, Majd Al-Rubaie, and Reem Al-Rubaie. DeepMoji: Vae-based generative model for text with emotional content. *arXiv preprint arXiv:1807.03961*, 2018.

Yujie Luo, Yuzhe Pan, Zhimin Li, Zongyi Li, Zhaohui Wang, and Zhaohui Zhao. scDiffusion: A diffusion model for single-cell rna sequencing data generation. *arXiv preprint arXiv:2401.12196*, 2024.

Moustafa Marouf, Yassine Bounab, Jörg Dörnemann, Evgeny Gribov, Bo He, David König, Julian König, Peter König, Thomas König, Volker König, et al. scGAN: Generating single-cell rna-seq data with generative adversarial networks. *Bioinformatics*, 36(15):4302–4309, 2020.

Alessandro Palma, Till Richter, Hanyi Zhang, Manuel Lubetzki, Alexander Tong, Andrea Dittadi, and Fabian Theis. Multi-modal and multi-

attribute generation of single cells with cfgen, 2025. URL https://arxiv.org/abs/2407.11734.

Jun Qi, Yan Ma, Rui Zhang, Zhijian Han, and Jun Li. Graph-based inference of cell differentiation trajectories from single-cell rna sequencing data. *Genome biology*, 22(1):1–22, 2021.

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*, pages 10684–10695, 2022.

Antoine-Emmanuel Saliba, Alexander J Westermann, Sebastian Gorski, and Jörg Vogel. Single-cell rna-seq for pathogen-infected cells. *Methods*, 67(3): 313–321, 2014.

Lisa Sikkema, Ciro Ramírez-Suástegui, Daniel C Strobl, Tessa E Gillett, Luke Zappia, Elo Madissoon, Nikolay S Markov, Laure-Emmanuelle Zaragosi, Yuge Ji, Meshal Ansari, et al. An integrated cell atlas of the lung in health and disease. *Nature medicine*, 29(6):1563–1577, 2023.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Brendan Swersky. Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning*, 2015.

Yang Song and Stefano Ermon. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2020.

Fuchou Tang, Catalin Barbacioru, Emil Nordman, Chunming Li, Nan Xu, M Azim Surani, Gavin Vassar, Yanyi Luo, Shuguang Gao, Matthew A Eaton, et al. RNA-seq analysis of single-cell transcriptome and splicing in early mouse embryos. *Nature methods*, 6(5):377–382, 2009.

Matthew Thorpe, Tan Nguyen, Hedi Xia, Thomas Strohmer, Andrea Bertozzi, Stanley Osher, and Bao Wang. Grand++: Graph neural diffusion with a source term. *ICLR*, 2022.

Peng Wang, Shijie Yu, Jun Wen, Mingyue Chen, Hongqi Sun, Shuyuan Liu, and Jianping Sun. Graph neural networks for cell type annotation from single-cell rna sequencing data. *Bioinformatics*, 37(11):1558–1566, 2021.

Marcin Waniek, Tomasz P Michalak, Michael J Wooldridge, and Talal Rahwan. Hiding individuals and communities in a social network. *Nature Human Behaviour*, 2(2):139–147, 2018.

Zhaohui Zhao, Zhimin Li, Zongyi Li, Zhaohui Wang, Yujie Luo, and Yuzhe Pan. Graph neural networks for single-cell rna sequencing data analysis. *Briefings in Bioinformatics*, 22(6):bbab306, 2021.

Ying Zhou, Min Zhou, Ruijie Li, and Yan Liu. Graph convolutional neural network for spatial transcriptomics analysis. *Nature communications*, 13(1): 5550, 2022.

Jiaming Zhuo, Can Cui, Kun Fu, Bingxin Niu, Dongxiao He, Yuanfang Guo, Zhen Wang, Chuan Wang, Xiaochun Cao, and Liang Yang. Propagation is all you need: A new framework for representation learning and classifier training on graphs. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 481–489, 2023.

Daniel Zügner, Amir Akbarnejad, and Stephan Günnemann. Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2847–2856, 2018.

Appendix A. Dataset Details

We evaluate LapDDPM on a diverse set of single-cell datasets, ranging from standard benchmarks to large-scale atlases and complex multi-modal/spatial data. Table 3 summarizes the key statistics of each dataset.

Table 3: Summary of Datasets used in Experiments.

Dataset	Type	Cells/Spots	Conditioning	Source
PBMC3K	scRNA-seq	2,638	Cell Type	Healthy Donor (10x Genomics, 2017)
Dentate Gyrus	scRNA-seq	18,213	Cell Type	Mouse Hippocampus (La Manno et al., 2018)
Tabula Muris	scRNA-seq	245,389	Tissue	Mouse Atlas (Consortium et al., 2018)
HLCA	scRNA-seq	584,944	Cell Type	Human Lung (Sikkema et al., 2023)
Visium Lymph Node	Spatial	∼3,800	Spatial Domain	10x Genomics (PBMC) (10xGenomics, 2019)
10x Multiome	Multi-modal	∼14,000	Cell Type	10x Genomics 14k (PBMC) (10xGenomics, 2020)

PBMC3K A standard benchmark dataset of Peripheral Blood Mononuclear Cells from a healthy donor (10x Genomics, 2017). It contains roughly 2,700 cells with well-defined cell types (e.g., T-cells, B-cells, Monocytes), serving as an initial validation for generative fidelity.

Dentate Gyrus A developmental dataset from the mouse hippocampus (La Manno et al., 2018), capturing the differentiation trajectory of granule cells. This dataset is topologically complex, containing continuous manifolds representing developmental time, making it suitable for testing the model's ability to preserve manifold structure.

Tabula Muris A large-scale organism-wide atlas (Consortium et al., 2018) comprising nearly 250,000 cells from multiple mouse tissues. We condition on tissue labels to evaluate the model's ability to capture high-level biological variation across different organs.

Human Lung Cell Atlas (HLCA) A massive integrated atlas of the human lung (Sikkema et al., 2023), containing over half a million cells from diverse donors and health conditions. This dataset tests the scalability of LapDDPM and its ability to model fine-grained cellular heterogeneity in a complex human organ.

Visium Lymph Node A spatial transcriptomics dataset provided by 10x Genomics (10xGenomics, 2019). Unlike dissociated scRNA-seq, this data preserves the spatial location of capture spots (∼3,800 spots). We treat this as a graph learning problem where the graph topology is defined by physical spatial coordinates. The goal is to generate gene expression that respects the spatial autocorrelation and domain structure of the tissue.

10x Multiome (PBMC) A paired scRNA-seq and scATAC-seq dataset of human PBMCs (∼14k nuclei) provided by 10x Genomics (10xGenomics, 2020). We use the chromatin accessibility (ATAC) modality as a conditioning context for generating gene expression (RNA). This setup evaluates the model's capability for cross-modal generation and integration, leveraging the chromatin embedding as a structural prior.

Appendix B. Computational Complexity Analysis

LapDDPM's design emphasizes scalability for large-scale scRNA-seq datasets. Let N be the total number of cells, N_b the batch size, D_f the number of filtered genes, P_{pca} the PCA components, k_{nn} the k-NN neighbors, k_{pe} the LPE dimension, K_{cheb} the Chebyshev filter order, d_{lat} the latent dimension, d_{hid} the GNN hidden dimension, $d_{\text{hid\_mlp}}$ the MLP hidden dimension, T_{diff} the diffusion timesteps, and ip the power iterations for adversarial perturbation. The preprocessing phase, including gene filtering ($\mathcal{O}(N \cdot D)$), PCA ($\mathcal{O}(N \cdot D_f \cdot P_{\text{pca}})$), k-NN graph construction ($\mathcal{O}(N \log N \cdot P_{\text{pca}} + N \cdot k_{\text{nn}} \cdot P_{\text{pca}})$), and LPE computation ($\mathcal{O}(k_{\text{pe}} \cdot N \cdot k_{\text{nn}})$), scales approximately linearly with N. During training, however, the spectral adversarial

perturbation is applied only to the sparse batch-local graph, yielding a per-step cost of $\mathcal{O}(ip \cdot N_b \cdot k_{\text{nn}})$ that is independent of the total atlas size. The remaining dominant per-step costs arise from the spectral encoder's Chebyshev GNN layers ($\mathcal{O}(K_{\text{cheb}} \cdot N_b \cdot k_{\text{nn}} \cdot d_{\text{hid}}^2)$), the ScoreNet ($\mathcal{O}(N_b \cdot d_{\text{hid\_mlp}}^2)$), and the Feature Decoder ($\mathcal{O}(N_b \cdot (d_{\text{hid}}^2 + d_{\text{hid}} \cdot D_f))$). Similarly, the generation process, primarily driven by ScoreSDE sampling ($\mathcal{O}(T_{\text{diff}} \cdot N_{\text{gen}} \cdot d_{\text{hid\_mlp}}^2)$) and feature decoding ($\mathcal{O}(N_{\text{gen}} \cdot (d_{\text{hid}}^2 + d_{\text{hid}} \cdot D_f))$), scales linearly with the number of generated samples N_{gen}. Empirically, this yields only a modest overhead on HLCA: 191.4 minutes for LapDDPM versus 187.6 minutes for scDiffusion during training, with essentially identical inference time (4.1s vs. 4.3s for 10k generated cells).

B.1. Ablation Studies

To comprehensively understand the contribution of each component to LapDDPM's performance and robustness, we conduct a series of ablation studies. We evaluate the impact of our novel spectral adversarial perturbations by disabling the `LaplacianPerturb` module, assessing its role in enhancing generation quality and robustness to structural variations. The effect of input gene masking on learning robust representations is investigated by varying the `INPUT_MASKING_FRACTION`. We also examine the significance of LPEs by comparing the full model with a variant where LPEs are omitted, highlighting their contribution to capturing graph topology. Additional sensitivity analyses over the perturbation radius ϵ, the number of power iterations $ip \in \{1, 2, 5\}$, and the number of spectral modes used for the adversary show stable behavior across settings. In particular, $ip = 2$ gives the best HLCA Wasserstein distance, the leading mode ($k = 1$) is best on clustered PBMC3K data, and larger ϵ values eventually degrade performance, consistent with DRO over-regularization. We further find that performance remains stable across k-NN graph parameters and PCA dimensionalities, indicating that LapDDPM is not brittle to standard preprocessing choices. These systematic ablations provide comprehensive insights into the design choices of LapDDPM, validating the necessity and effectiveness of its novel components for conditional scRNA-seq data generation, with representative results presented in Table 4.

Table 4: Ablation study results for LapDDPM on the PBMC3K dataset. Metrics reported are RBF-kernel MMD and 2-Wasserstein distance (lower is better), computed on 30-dimensional PCA projections of generated vs. real test data. Results are mean ± standard deviation over 10 runs. Values for the full LapDDPM model are bolded for comparison. Degradation in performance indicates the importance of the ablated component.

Model Variant	Conditional Generation		Unconditional Generation	
	MMD (↓)	WD (↓)	MMD (↓)	WD (↓)
LapDDPM (Full Model)	**0.41±0.15**	**14.84±0.83**	**0.23±0.02**	**13.98±0.74**
w/o Spectral Adv. Perturbations	0.55±0.18	16.50±0.90	0.32±0.03	15.80±0.80
w/o Input Gene Masking (10%-30%)	0.48±0.16	15.60±0.85	0.28±0.02	14.70±0.78
w/o Laplacian Positional Encoding (LPE)	0.60±0.20	17.20±0.95	0.35±0.04	16.10±0.88

Appendix C. Implementation Details

C.1. Architecture Hyperparameters

We report the hyperparameter configuration used in our spatial and multiome experiments in Table 5. The model uses a Graph Transformer encoder, a ScoreSDE/DDIM denoiser, and a ZINB decoder with a library-size head.

Table 5: Hyperparameter settings for LapDDPM.

Parameter	**Spatial**	**Multiome**
Graph Construction		
k-NN Neighbors (k_{nn})	30	15
PCA Components (P_{pca})	50	50
LPE Dimension (k_{pe})	50	50
Network Architecture		
Latent Dimension (d_{lat})	256	128
Encoder Layers (Graph Transformer)	4	4
Encoder Hidden Dim	512	256
ScoreNet Hidden Dim	512	256
Time Embedding Dim	32	32
Optimization (Stage A & B)		
Batch Size	64	16
Learning Rate	1×10^{-4}	1×10^{-4}
Warmup Steps	200	200
Optimizer	AdamW	AdamW
Diffusion Steps (T)	1000	1000
Diffusion Type	VP-SDE	VP-SDE
Spectral Adv. Iterations (ip)	3	5
Spectral Perturbation (η)	0.01	0.01

C.2. Training Protocol

Training proceeds in two stages for stability and robust manifold estimation:

- **Stage A (Reconstruction Pre-training):** Train the encoder and ZINB decoder with diffusion loss disabled to learn a stable latent representation. In long-run experiments we pretrain for ∼1200 epochs (of 2000 total).
- **Stage B (Diffusion Training):** Enable diffusion and unfreeze the denoiser to learn the reverse SDE, while keeping spectral adversarial perturbation active for robustness. The remaining epochs are used for Stage B.

C.3. Gene-Gene Correlation Analysis

Preserving gene-gene correlations is crucial for regulatory network inference. We compute correlation matrices over highly variable genes in real and generated data and quantify alignment via GRN Spearman and normalized matrix difference. As reported in the main text, LapDDPM achieves GRN Spearman 0.61 (Spatial) and 0.59 (Multiome), indicating recovery of higher-order gene dependencies beyond marginal distributions, while outperforming SOTA baselines.

C.4. Additional Downstream Utility

To complement distributional metrics, we also evaluate whether LapDDPM supports downstream biological analyses on PBMC3K. For rare-cell augmentation, we downsample pDCs to 50 cells (1.5%) and train a classifier with synthetic augmentation; LapDDPM improves held-out F1 to 0.79, compared with 0.51 without augmentation, 0.63 for scVI, and 0.67 for scDiffusion. For differential expression consistency between CD4+ and CD8+ T-cells, LapDDPM recovers 91% of the top-200 real DE genes, compared with 74% for scVI and 81% for scDiffusion. These results further support that LapDDPM preserves biologically actionable structure rather than only marginal similarity.