Integrating Generative Counterfactual Explanations into Graph Neural Networks for Enhanced Adversarial Robustness

Assignee Research

June 11, 2026

Abstract

Generating high-fidelity and biologically plausible synthetic single-cell RNA sequencing (scRNA-seq) data, especially with conditional control, is challenging due to its high dimensionality, sparsity, and complex biological variations. Existing generative models often struggle to capture these unique characteristics and ensure robustness to structural noise in cellular networks. We introduce LapDDPM, a novel conditional Graph Diffusion Probabilistic Model for robust and high-fidelity scRNA-seq generation. LapDDPM uniquely integrates graph-based representations with a score-based diffusion mode

1 Introduction

This paper examines: LapDDPM: A Conditional Graph Diffusion Model for scRNA-seq Generation with Spectral Adversarial Perturbations. Research question: Can integrating generative counterfactual explanations into graph neural networks enhance robustness metrics against adversarial perturbations on standard graph classification datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

16 papers retrieved. 14 claims extracted; 11 independently verified. Quality review score: 7.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Early approaches for generating synthetic cellular profiles often adapted models from general machine learning, such as	✓	0.17
VAE-based models learn a low-dimensional latent representation and reconstruct gene expression, often accounting for spa	✓	0.24
GANs aim to learn a mapping from a simple prior distribution to the complex data distribution through an adversarial tra	✓	0.23
Flow-based models have been explored for their exact likelihood estimation and invertible mappings.	×	0.15
These models often face challenges in capturing the intricate multi-modal distributions, preserving biological heterogen	✓	0.24
GNNs have been applied to various tasks in single-cell biology, including cell type annotation, trajectory inference, an	✓	0.23
GNNs are typically used as feature extractors or classifiers in single-cell biology applications.	×	0.14
Diffusion Probabilistic Models (DPMs) have emerged as a powerful class of generative models, demonstrating state-of-the-	✓	0.27
The overall training procedure combines diffusion, reconstruction, and KL divergence losses, with the encoder being trai	✓	0.26
Given a scRNA-seq dataset consisting of N cells and D genes, represented as a count matrix X $\in$RN$\times$D, a graph G = (V, E) i	✓	0.30
Prior to graph construction, genes expressed in fewer than a specified threshold of cells are filtered out to reduce spa	✓	0.21
The raw count data is normalized and log-transformed for stable numerical operations during feature extraction.	✓	0.28
A k-NN graph is constructed on the cells using PCA-reduced space.	✓	0.21
The adjacency matrix A $\in${0, 1}N$\times$N is formed based on the k nearest neighbors of each cell.	×	0.11

References

- `http://arxiv.org/abs/2312.05905v2`
- `http://arxiv.org/abs/2403.20287v5`
- `http://arxiv.org/abs/2506.13344v1`