A Differentiable Framework for Learning Geometric Priors in Graph Representation Learning

Mohana Rangan Desigan
Independent Research

August 15, 2025

Abstract

The graph Laplacian is a cornerstone of representation learning on non-Euclidean data, forming the basis for spectral clustering, diffusion maps, and graph wavelet transforms. However, its standard formulation is isotropic, assuming information diffuses uniformly in all directions. This assumption fails to capture the inherent anisotropy of real-world data manifolds, such as the distinction between coherent "semantic threads" and tenuous "conceptual bridges" in knowledge graphs. Hand-crafting anisotropic operators to address this is brittle and often leads to catastrophic failures like graph fragmentation.

We resolve this challenge by proposing a new paradigm: learning the operator itself. This paper introduces a fully differentiable pipeline that learns the parameters of a geometry-to-conductance mapping by optimizing a downstream, task-specific loss. The framework's key components include a parameterized, connectivity-preserving mapping function $h_\theta(\kappa)$, a differentiable symmetric eigendecomposition layer, and a contrastive loss function. By backpropagating through the entire spectral transform, our system learns the optimal, data-driven geometric prior, discovering how to warp the manifold's diffusion process to best suit the learning task. We provide the complete theoretical blueprint, address the core research challenges of scalability and expressivity, and demonstrate feasibility with a proof-of-concept implementation that successfully learns to enhance cluster structure.

1 The Parameterized Curvature-to-Conductance Mapping

The central hypothesis of anisotropic graph learning is that the local geometry of a data manifold, often captured by a measure of discrete Ricci curvature κ, should inform the diffusion process. We reframe the problem from one of design to one of *learning*, introducing a parameterized family of monotonic, strictly positive functions, $h_\theta(\kappa)$.

1.1 Axiomatic Requirements

Any valid mapping must satisfy:

1. **Strict Positivity (Connectivity Preservation):** $h_\theta(\kappa) \geq \varepsilon > 0$.

2. **Monotonicity:** The mapping should be non-decreasing.

3. **Differentiability:** $h_\theta(\kappa)$ must be differentiable with respect to its parameters θ.

1.2 The Shifted Sigmoid Family

The first family of functions that satisfies these axioms is the **learnable shifted sigmoid**:

$$h_\theta(\kappa) = \varepsilon + (1 - \varepsilon) \cdot \sigma(\theta_1(\kappa - \theta_2)) \tag{1}$$

where $\sigma(z) = (1 + e^{-z})^{-1}$ is the logistic sigmoid function, and the learnable parameters are the sensitivity $\theta_1 \in \mathbb{R}^+$ and the centering point $\theta_2 \in \mathbb{R}$.

2 The Differentiable Anisotropic Pipeline

To learn the optimal parameters θ of the mapping h_θ, we must construct an end-to-end differentiable computational graph (Algorithm 1) that maps θ to a final representation X'_θ.

Algorithm 1 Differentiable Anisotropic Representation (Forward Pass)

Input: Graph $G = (V, E, W)$, curvature κ, signal X, scale t, learnable parameters θ.
Output: Anisotropic representation $X'_\theta \in \mathbb{R}^{N \times D}$.

1: $W'_\theta \leftarrow W \odot h_\theta(\kappa)$ $\triangleright$ Element-wise modulation of edge weights
2: $D'_\theta \leftarrow \mathrm{diag}(\sum_j w'_{ij})$
3: $L_{\mathrm{aniso},\theta} \leftarrow I - (D'_\theta)^{-1/2} W'_\theta (D'_\theta)^{-1/2}$
4: $\{\lambda_{k,\theta}\}, U_\theta \leftarrow \mathrm{Eigendecompose}(L_{\mathrm{aniso},\theta})$ $\triangleright$ Differentiable w.r.t. θ
5: $\Psi_{t,\theta} \leftarrow U_\theta e^{-t\mathrm{diag}(\lambda_{k,\theta})} U_\theta^\top$ $\triangleright$ Differentiable graph wavelet operator
6: $X'_\theta \leftarrow \Psi_{t,\theta} X$
7: **return** X'_θ

2.1 The Differentiable Eigendecomposition Layer

The key enabler is the use of a differentiable symmetric eigendecomposition layer, available in modern auto-differentiation frameworks like PyTorch (`torch.linalg.eigh`). This allows gradients to flow from a downstream loss back to the mapping parameters θ, treating spectral analysis as a learnable component.

3 Loss Function and Optimization

We learn θ by minimizing a task-specific loss. For unsupervised or semi-supervised learning, a contrastive loss is ideal. The triplet margin loss for an anchor x'_i, positive x'_j, and negative x'_k is:

$$\mathcal{L}_{\mathrm{triplet}}(i, j, k; \theta) = \max\left(\|x'_{i,\theta} - x'_{j,\theta}\|_2^2 - \|x'_{i,\theta} - x'_{k,\theta}\|_2^2 + m,\, 0\right) \tag{2}$$

where $m > 0$ is a fixed margin. The system iteratively samples triplets, computes the loss, and updates θ via backpropagation and a standard optimizer like Adam.

4 Challenges and Integrated Research Directions

This framework, while powerful, presents a series of integrated research challenges that must be addressed to create a truly robust and scalable system. We frame these not as future work, but as core components of the present research program, to be validated with small-scale experiments on free-tier platforms.

Computational Scalability The full eigendecomposition scales as $O(N^3)$, making it infeasible for large graphs.

- **Research Direction:** We will implement and validate differentiable approximations of the spectral transform. The primary approach will be to replace the full eigendecomposition with a differentiable **Chebyshev polynomial approximation** of the matrix exponential $e^{-tL_{\text{aniso},\theta}}$, reducing the forward pass complexity to be near-linear in the number of edges.

Fidelity of the Geometric Signal Discrete Ricci curvature can be noisy and computationally expensive.

- **Research Direction:** We will investigate replacing curvature with a more robust and scalable geometric signal. A leading candidate is a **learned local descriptor**, where a small graph neural network (GNN) is trained jointly to produce a low-dimensional summary of a node's local neighborhood topology, which then serves as the input to h_θ.

Expressivity of the Mapping The shifted sigmoid, while interpretable, may be too simple.

- **Research Direction:** We will implement a more expressive, learnable mapping using a **Monotonic Neural Network**. By constraining the weights of a small MLP to be non-negative, we can learn an arbitrarily complex monotonic function h_θ while still satisfying the axiomatic requirements.

Generalization and Overfitting A mapping h_θ tuned to a specific task may overspecialize.

- **Research Direction:** To combat this, we will incorporate a **geometric regularizer** into the loss function. For instance, a term that penalizes excessive deviation from the isotropic case, $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{triplet}} + \gamma \|L_{\text{aniso},\theta} - L_{\text{iso}}\|_F^2$, encourages the system to find the minimal geometric modification necessary to solve the task.

Integration with Spatial GNNs The proposed framework is purely spectral, while the dominant paradigm in graph learning is spatial message-passing.

- **Research Direction:** We will design and implement a **hybrid spectral-spatial layer**. This layer will perform a standard message passing update, but the aggregation weights will be modulated by the learned anisotropic conductance $h_\theta(\kappa_{ij})$. This hybridizes the scalability of spatial methods with the global, geometric awareness of our spectral operator.

By systematically addressing these challenges, this research program aims to deliver not just a single algorithm, but a complete, robust, and scalable framework for learning on geometric graphs.