Ph.D. Thesis Proposal

# Title:

Robust and Scalable Spectral Clustering via Adaptive Laplacian Learning and Perturbation-Aware Eigenanalysis

# 1. Problem Statement

Spectral clustering is a powerful but sensitive method for unsupervised learning. It depends heavily on the construction of the similarity graph, suffers from poor robustness to noise, and becomes computationally intractable at scale. This thesis aims to address the following problem:  
  
 Can we design a framework to learn or adapt the graph Laplacian in a data-driven, robust, and scalable manner such that spectral clustering remains effective in noisy, high-dimensional, and large-scale settings?

# 2. Significance and Motivation

Spectral clustering is foundational across data-driven sciences. Its practical effectiveness is hindered by its sensitivity to parameter choices, poor robustness, and scalability limits. Solving these problems would enable reliable unsupervised learning for applications in biology, vision, NLP, and network science.  
  
 The problem lies at the intersection of spectral graph theory, numerical linear algebra, and optimization—bridging theory with scalable implementations.

# 3. Related Work

- Ng et al. (2002): Normalized spectral clustering.  
- Von Luxburg (2007): Theoretical foundations.  
- Kalofolias (2016), Dong et al. (2016): Laplacian learning.  
- Drineas & Mahoney (2005): Nyström method for scalability.  
- Fanuel et al. (2017): Convex relaxation for robustness.  
- Wang & Singer (2016): Rotation recovery under noise.

# 4. Research Objectives

Objective 1: Adaptive Laplacian Learning  
- Learn similarity graphs from data with sparsity and smoothness constraints.  
  
Objective 2: Perturbation-Aware Spectral Embedding  
- Analyze eigenvector sensitivity and robustness guarantees.  
  
Objective 3: Scalable Approximation Techniques  
- Develop low-rank, randomized methods for large graphs.  
  
Objective 4: End-to-End Pipeline  
- Integrate learning, robustness, and approximation into a practical clustering tool.

# 5. Prerequisites and Tools

Theoretical:  
- Linear algebra (eigenvalue theory, SVD, matrix perturbations)  
- Spectral graph theory, convex optimization  
  
Practical:  
- Python, MATLAB  
- Libraries: NumPy, SciPy, PyTorch, NetworkX

# 6. Expected Contributions

- New robustness bounds for spectral embeddings  
- Algorithms for scalable spectral clustering  
- Open-source toolkit for robust spectral clustering  
- Applications to real-world data  
- Publications in top-tier ML/AI conferences and journals