# COMP511 Project Proposal: When do Graph Neural Networks Outperform Spectral Methods under Structural Noise?

Finn Creeggan*
finn.creeggan@mail.mcgill.ca
McGill University
Montreal

Sabrina Du*
sabrina.du@mail.mcgill.ca
McGill University
Montreal

Jamie Wise*
jamie.wise@mail.mcgill.ca
McGill University
Montreal

## 1 Introduction and Motivation

Graphs provide a natural representation for complex relational systems, including biological interaction networks, social networks, and neural connectomes. A central goal in graph learning is to recover meaningful latent structure from observed patterns of connectivity. In practice, however, this problem is rarely clean [1]. Real-world graphs are often sparse, noisy, and only partially observed, so the recorded edges may reflect an imperfect measurement process rather than the true underlying structure [2]. As a result, extracting reliable signals from graph data is fundamentally challenging and highly relevant across many scientific and applied domains [1].

This challenge matters because many of the settings in which graph learning would be most useful are precisely those in which data "quality" is weakest. In areas like biology and finance, collecting graph data can be expensive, technically difficult, and prone to measurement error [2, 3]. In these regimes, model performance is shaped not only by expressive power, but also by robustness to dataset noise and the ability to learn from limited information [4]. In many situations, simpler methods remain competitive because they are more interpretable [5], are less prone to overfitting noise when there is limited data [6], and achieve better calibration [7].

Motivated by this idea, this project studies whether the empirical gains often attributed to GNNs persist under controlled increases in structural noise, or whether simpler spectral methods become equally effective, or even preferable, when graph observations are corrupted. Answering this question is important both scientifically and practically: scientifically, it helps clarify a source of performance in modern graph learning; practically, it helps ensure that the complexity of a model is truly justified by data quality.

## 2 Related Work

**Defining structural noise in real-world graphs.** Structural noise in graphs primarily arises through missing edges and spurious connections [8, 9]. Both can emerge from a variety of sources depending on the application domain. For example, in biological networks, experimental instruments may fail to detect protein–protein interactions due to limited sensitivity; in social networks, bot-generated posts may be misidentified as originating from human accounts; and in connectomics, reconstruction pipelines may omit synaptic connections due to segmentation errors. Regardless of its source, structural noise degrades the observed topology of the graph and generally reduces performance on downstream tasks such as community detection [10, 11].

**Limitations of GNNs under noise.** Despite these challenges, Graph Neural Networks (GNNs) have emerged as a dominant paradigm for learning on graphs. Through iterative message passing, GNNs aggregate information from local neighborhoods, enabling the learning of complex representations that encode structural information [12]. Their strong empirical performance across a wide range of applications has led to the perception that GNNs provide a more expressive framework than classical spectral methods [13].

However, recent work has shown that GNNs are not inherently robust to structural noise [14], demonstrating empirically that progressively perturbing synthetic datasets by deleting edges reduces the performance of all GNN subtypes on community detection tasks. [15, 16] showed that GNNs' weakened performance under structural noise is largely due to message-passing operations amplifying noise by propagating incorrect information across the graph. Together, these works suggest that the performance gains of GNNs may not stem from robustness to noise, but rather from their ability to exploit clean or moderately clean structure, raising questions about their reliability in high-noise real-world regimes.

**Spectral methods and their convergence with GNNs.** At the same time, there has been significant progress in improving classical spectral methods to better handle structural noise in graphs. Techniques such as regularized spectral clustering [17] and eigenvector filtering [18] have been developed to mitigate issues such as eigenvector localization and instability. These methods aim to recover informative low-dimensional structure by carefully selecting or regularizing the spectral components used for embedding.

Interestingly, recent work has suggested that the performance gap between GNNs and simpler spectral methods may be smaller than often assumed. In particular, approaches based on label propagation over spectral positional encodings have shown that predictive accuracy in many community detection tasks can match or outperform GNNs [19]. This challenges the idea that strong performance requires deep nonlinear transformations. Supporting this view, [20] found that by removing the intermediate nonlinearities and collapsing the layer-wise transformations of a standard GCN to create a Simplified Graph Convolution (SGC), one can retain performance comparable to full GCNs. The authors further highlight that this approach essentially reduces GCNs to a fixed low-pass filter of the Laplacian eigenbasis followed by a linear classifier. Together, these findings suggest that a substantial portion of GNN performance may arise from effective filtering and prediction over informative eigenvectors, rather than from intrinsic nonlinear expressivity.

**The gap.** Despite this evidence, there remains a lack of systematic investigation into how GNNs and modern spectral methods compare under controlled increases in structural noise. In particular, it is unclear whether carefully designed spectral filtering methods

can match or outperform GNNs as structural and node feature noise increases. Addressing this question is essential for understanding when complex nonlinear architectures are necessary versus when simpler, more interpretable spectral methods suffice.

## 3 Problem Definition

Extensive research has characterized how GNN performance and robustness on community detection vary with graph properties and structural noise. However, the nature of the representations learned by GNNs as these properties shift remains poorly understood. Specifically, it is unclear whether GNN performance gains stem from capturing genuinely nonlinear structure or from learning relationships between graph components that could largely be recovered by improved spectral methods.

To shed light on the representation learning performed by GNNs on community detection tasks, we will conduct a systematic evaluation of spectral methods and GNN architectures on community detection tasks as structural noise increases in synthetic and real-world datasets. We will conduct these experiments in a supervised setting and control for graph properties that disadvantage or favour GNNs, such as heterophily [21] and sparsity [22], respectively. This research will ultimately help researchers determine the "breakpoint" where the computational overhead of GNNs is justified over traditional spectral techniques for real-world graphs with a specific amount of noise.

## 4 Dataset and Experimental Design

We evaluate our hypothesis primarily on controlled synthetic graph benchmarks generated using the Stochastic Block Model (SBM) and the Lancichinetti–Fortunato–Radicchi (LFR) model. These benchmarks are well suited to our question because they provide planted ground-truth communities and allow key graph properties, such as sparsity and heterophily, to be held approximately constant while structural noise is varied systematically. SBM graphs provide cleaner and more regular community structure, whereas LFR graphs are more challenging because they exhibit heavier-tailed degree distributions and more irregular community sizes.

Each synthetic graph contains approximately 1,000 nodes with average degree 25, making repeated experiments computationally feasible while still large enough to exhibit nontrivial community structure. Graphs are generated from fixed SBM or LFR parameter settings using controlled random seeds. For each graph family, we construct clean base graphs and store each instance as a sparse edge list or adjacency matrix together with planted node labels and metadata describing the generation settings, including perturbation level and seed.

To assess whether trends observed in the synthetic setting extend to more natural data, we also plan to evaluate selected methods on **real-world graphs**. These include protein–protein interaction networks from STRING [23] and neuro-connectomics graphs derived from ABIDE resting-state fMRI data [24], with financial correlation networks as a possible additional domain. Because real-world community labels are typically noisier and less definitive than planted synthetic labels, we will estimate Edge Signal-to-Noise Ratio (ENSR [25]) as a proxy for the clarity of graph structure and estimate standard Signal-to-Noise Ratio (SNR) as a proxy for graph feature

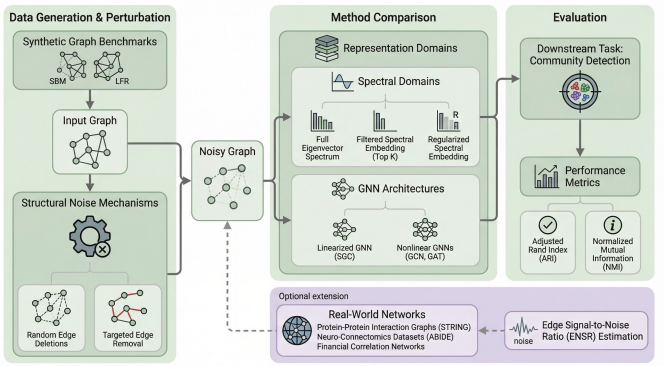


**Figure 1: Overview of our project.**

noise. Together, these datasets provide a progression from tightly controlled benchmarks to more realistic settings in which graph noise is intrinsic rather than synthetically imposed.

## 5 Methodology

We evaluate community detection methods on a controlled synthetic benchmark designed to test how performance changes as graph structure becomes less reliable, before moving on to real-world graphs. The benchmark has two parts. In the first, we progressively remove edges from synthetic graphs and measure how well different methods can still recover the planted communities. In the second, we keep the graph fixed but vary how informative the node features are. Both experiments use the same overall pipeline and are evaluated on community detection with ARI. For fair comparison, we will also compute FLOP budgets required for all methods to ensure that performance gains are presented relative to their computational cost. Thus, changes in performance can be tied to controlled changes in structure or features rather than to differences between datasets.

As mentioned previously, our synthetic graphs (LFR and SBM) serve different purposes. SBM graphs provide a cleaner and more regular community structure, whereas LFR graphs are more challenging due to their more irregular degree patterns and community sizes. For each family, we begin with clean base graphs and then create progressively noisier versions of those same graphs by deleting edges. This enables the study of how performance changes as the underlying graph is perturbed.

**Structural noise experiment** In the structural-noise experiment, we apply two edge-deletion strategies. The first is random edge deletion, where removed edges are chosen uniformly at random. The second is targeted edge deletion based on betweenness centrality, where edges that are more central to connectivity are removed first. These two perturbation types let us compare robustness under both untargeted and more adversarial structural degradation. Each graph instance is evaluated by every method in the benchmark. For each graph, we use a fixed 70/15/15 split of nodes into training, validation, and test sets, generated once with a fixed random seed and reused across methods. This ensures that all methods are evaluated on the same partition of each graph. For the

progress report, models are run with a fixed default hyperparameter configuration rather than a full hyperparameter search. We plan to use Bayesian hyperparameter tuning for the full proposal.

**Spectral methods.** To support the spectral methods, each graph is loaded together with its planted labels and precomputed spectral representations. The pipeline uses three Laplacian-based representations: the full eigenspectrum of the symmetric normalized Laplacian, a k-cut representation chosen using the eigengap heuristic, and a regularized eigenspectrum based on a Tikhonov-shifted Laplacian. These representations are computed on the full graph and then passed to downstream classifiers. Here, the spectral family consists of combinations of these embeddings with logistic regression or random forest classifiers, for a total of six spectral models.

**GNN-based methods.** We also compare against three graph neural network methods. In the current pipeline, GCN and GAT have been implemented, while SGC is planned but not yet implemented. These models are evaluated in a transductive setting: each method has access to the full graph structure during training, but only the labels of the training nodes are used as supervision. The task is thus not to generalize to an entirely new graph but to predict community labels of unseen nodes within the same graph.

**Feature noise experiment.** In the second experiment, we study feature quality directly. For a selected subset of structural-noise conditions, we generate 5-dimensional node features whose informativeness can be controlled continuously. Each community is assigned a fixed prototype feature vector, and Gaussian noise ($\epsilon$) is then added. When the informativeness is high, nodes in the same community have similar features; when it is low, the features become mostly noise. Features ($X_i$) are generated as

$$X_i = \alpha f_5(y_i) + (1 - \alpha)\epsilon_i \tag{1}$$

where $y_i$ is the planted community label, $f_5$ is the fixed 5-dimensional mapping, $\epsilon_i$ $N(0, I_5)$ and $\alpha \in [0, 1]$ is the amount of signal. We vary $\alpha$ while fixing the graph at three levels of structural noise, allowing us to study how methods respond to changes in feature quality under low, medium, and high-noise structural conditions.

The evaluation metric throughout is Adjusted Rand Index (ARI). ARI compares the predicted partition to the planted community assignment and is invariant to label permutation, which makes it appropriate for community recovery. Raw results are recorded at the level of individual graph instances, then summarized in two stages: first at the graph level and then at the condition level. Importantly, the independent unit of uncertainty in our benchmark is the base graph, not the node split. Condition-level means and standard deviations are therefore computed across the five base graphs within each setting. This captures variation across independently generated graph instances, rather than repeated partitions of the same graph, and is therefore more informative than treating different node partitions of one graph as independent samples.

## 6 Preliminary results

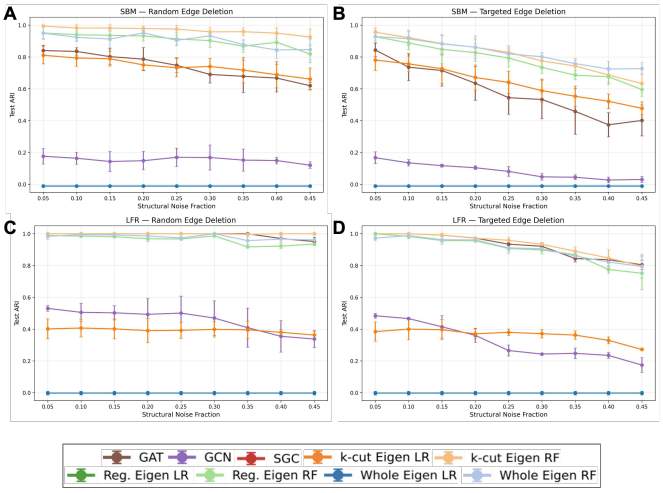


**Figure 2: Community detection performance with increasing graph structural noise.** We apply two edge-deletion strategies to increase graph structural noise: random edge deletion (A and C) and targeted edge deletion (B and D) based on betweenness centrality. ARI measures community detection performance across spectral and GNN methods noise increases on SBM and LFR graphs.

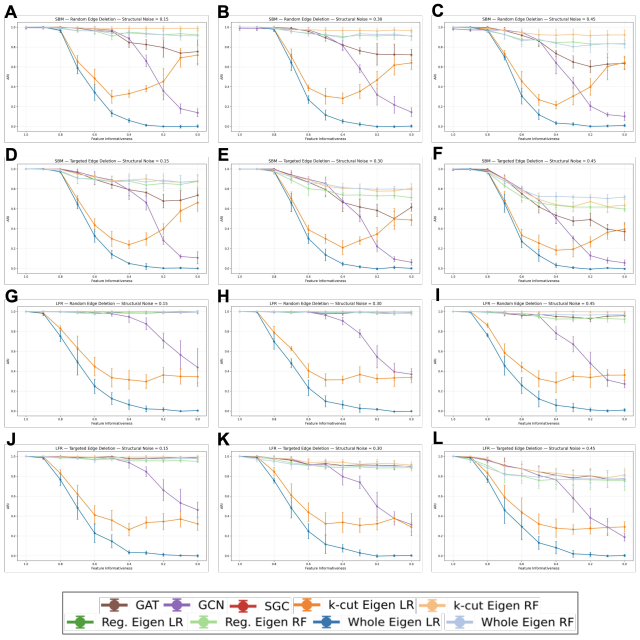


**Figure 3: Community detection performance with increasing node feature noise.** We decrease node feature informativeness by adding varying levels of Gaussian noise to node feature vectors, while fixing the type and level of structural noise. ARI measures community detection performance across spectral and GNN methods as feature informativeness decreases for SBM and LFR graphs.

## 7 Expected Contribution

This work provides a systematic and principled evaluation of GNNs relative to classical and modern spectral methods under varying levels of structural noise. By disentangling the roles of nonlinearity and spectral filtering, the project aims to clarify when GNNs offer genuine advantages, highlight the strength of advanced spectral approaches, and inform the design of simpler, more interpretable graph learning models. More broadly, this research contributes to a deeper understanding of the mechanisms driving performance in graph-based learning, with implications for both theoretical development and real-world applications.

## References


[1] Jiakang Li, Songning Lai, Zhihao Shuai, Yuan Tan, Yifan Jia, Mianyang Yu, Zichen Song, Xiaokang Peng, Ziyang Xu, Yongxin Ni, Haifeng Qiu, Jiayu Yang, Yutong Liu, and Yonggang Lu. A comprehensive review of community detection in graphs. *Neurocomputing*, 600:128169, 2024.

[2] Roger Guimerà and Marta Sales-Pardo. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52):22073–22078, 2009.

[3] Kartik Anand et al. The missing links: A global study on uncovering financial network structures from partial data. *Journal of Financial Stability*, 35:107–119, 2018.

[4] Siyi Qian et al. Robust training of graph neural networks via noise governance. *CoRR*, abs/2211.06614, 2023.

[5] Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, and Chudi Zhong. Interpretable machine learning: Fundamental principles and 10 grand challenges, 2021.

[6] Danilo Bzdok and John P. A. Ioannidis. Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in Neurosciences*, 42(4):251–262, 2019.

[7] Ben Van Calster, David J. McLernon, Maarten van Smeden, Laure Wynants, and Ewout W. Steyerberg. Calibration: the achilles heel of predictive analytics. *BMC Medicine*, 17(1):230, 2019.

[8] Wenrui Li, Changgee Chang, Suprateek Kundu, and Qi Long. Accounting for network noise in graph-guided bayesian modeling of structured high-dimensional data. *Biometrics*, 80(1):ujae012, 2024.

[9] Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning discrete structures for graph neural networks, 2019.

[10] Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. Revisiting the Bethe-Hessian: Improved community detection in sparse heterogeneous graphs, 2019.

[11] Leto Peel, Daniel B. Larremore, and Aaron Clauset. The ground truth about metadata and community detection in networks. *Science Advances*, 3(5):e1602548, 2017.

[12] Jan Böker, Ron Levie, Ningyuan Huang, Soledad Villar, and Christopher Morris. Fine-grained expressivity of graph neural networks, 2023.

[13] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2018.

[14] Jaidev Goel, Pablo Moriano, Ramakrishnan Kannan, and Yulia R. Gel. Community detection robustness of graph neural networks. *CoRR*, abs/2509.24662, 2025.

[15] 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.

[16] Linling Jiang, Wenchang Zhang, Hua Wang, and Fan Zhang. Robust memory-based graph neural networks for noisy and sparse graphs. *Engineering Applications of Artificial Intelligence*, 160:111874, 2025.

[17] Huan Qing. Community detection in multi-layer networks by regularized debiased spectral clustering. *Engineering Applications of Artificial Intelligence*, 152:110627, 2025.

[18] Mohammadreza Fattahi, Hamid Saeedi-Sourck, and Vahid Abootalebi. Large-scale graph signal denoising: A heuristic approach. *Digital Signal Processing*, 158:104914, 2025.

[19] Qian Huang, Horace He, Abhay Singh, Ser-Nam Lim, and Austin R. Benson. Combining label propagation and simple models out-performs graph neural networks. *CoRR*, abs/2010.13993, 2020.

[20] Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr., Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks, 2019.

[21] Jiong Zhu, Junchen Jin, Donald Loveland, Michael T. Schaub, and Danai Koutra. How does heterophily impact the robustness of graph neural networks? Theoretical connections and practical implications, 2021.

[22] Luana Ruiz, Ningyuan Huang, and Soledad Villar. A spectral analysis of graph neural networks on dense and sparse graphs, 2022.

[23] Damian Szklarczyk, Rebecca Kirsch, Michael Koutrouli, Hanan Nastou, Franziska Mehryary, Roman Hachilif, Annika L. Gable, Tao Fang, Alexander Doncheva, Aikaterini T. Pyysalo, et al. String database in 2025: protein networks with directionality and improved coverage. *Nucleic Acids Research*, 53(D1):D730–D737, 2025.

[24] Keqi Han, Yao Su, Lifang He, Liang Zhan, Sergey Plis, Vince Calhoun, and Carl Yang. Rethinking functional brain connectome analysis: do graph deep learning models help. *npj Artificial Intelligence*, 2:19, 2026.

[25] Mingze Dong and Yuval Kluger. Towards understanding and reducing graph structural noise for GNNs. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 5610–5633. PMLR, 2023.


Received 20 March 2026