

AutoSGNN: Automatic Propagation Mechanism Discovery for Spectral Graph Neural Networks

Shibing Mo^{1,2}, Kai Wu¹, Qixuan Gao², Xiangyi Teng^{2*}, Jing Liu^{1,2}

¹School of Artificial Intelligence, Xidian University

²Guangzhou Institute of Technology, Xidian University

msb@stu.xidian.edu.cn, kwu@xidian.edu.cn, gaoqx@stu.xidian.edu.cn, tengxiangyi@xidian.edu.cn, neouma@mail.xidian.edu.cn

Abstract

In real-world applications, spectral Graph Neural Networks (GNNs) are powerful tools for processing diverse types of graphs. However, a single GNN often struggles to handle different graph types—such as homogeneous and heterogeneous graphs—simultaneously. This challenge has led to the manual design of GNNs tailored to specific graph types, but these approaches are limited by the high cost of labor and the constraints of expert knowledge, which cannot keep up with the rapid growth of graph data. To overcome these challenges, we propose AutoSGNN, an automated framework for discovering propagation mechanisms in spectral GNNs. AutoSGNN unifies the search space for spectral GNNs by integrating large language models with evolutionary strategies to automatically generate architectures that adapt to various graph types. Extensive experiments on nine widely-used datasets, encompassing both homophilic and heterophilic graphs, demonstrate that AutoSGNN outperforms state-of-the-art spectral GNNs and graph neural architecture search methods in both performance and efficiency.

Code — <https://github.com/Explorermomo/AAAI2025-AutoSGNN>

Extended Version — <https://arxiv.org/abs/2412.12483>

Introduction

Graph Neural Networks (GNNs) have become central to machine learning, with recent advancements dividing them into spatial-based (Hamilton, Ying, and Leskovec 2017) and spectral-based (Defferrard, Bresson, and Vandergheynst 2016a) approaches. Spectral-based methods focus on creating advanced spectral graph filters for effective graph embedding, and have a explicit and explainable propagation mechanism for graphs, effectively characterizing the key features during the learning process.

Many traditional spectral GNNs assume homogeneity (McPherson, Smith-Lovin, and Cook 2001; Von Luxburg 2007), where interconnected nodes typically belong to the same category or share similar attributes. However, real-world graphs often exhibit significant heterogeneity, such as protein structure networks (Fout et al. 2017) that show

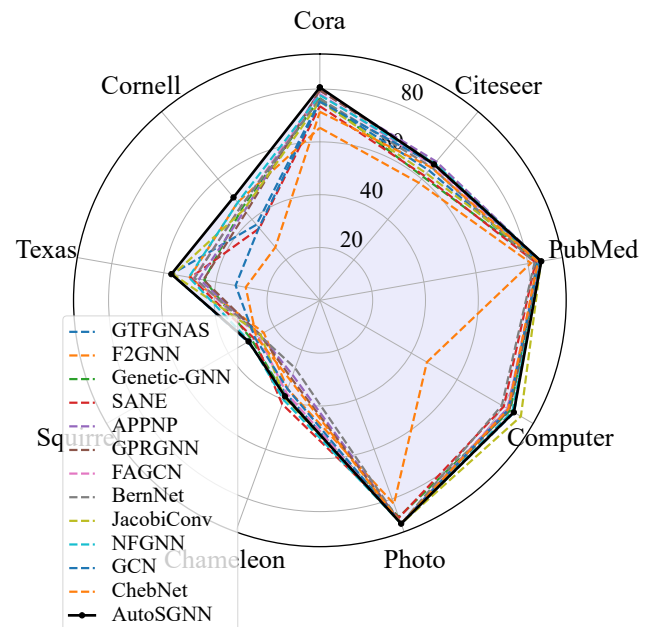


Figure 1: Radar chart of AutoSGNN and other spectral GNNs on 9 datasets.

stronger heterophilic connections between different types of amino acids. GNNs designed with the homogeneity assumption may struggle to perform well on such heterogeneous graphs (Mao et al. 2024a,b). Thus, the generalization ability of spectral GNNs is tested, and we find that a single spectral GNN is difficult to adapt to different graph data (see Figure 1, the advantageous application scenarios of different neural networks have differences). However, the current different spectral GNNs are manually designed to adapt to particular scenarios. Automatic and efficient GNN configuration schemes are urgently needed to reduce labor costs.

To solve this issue, in the past five years, many Neural Architecture Search (NAS) methods specifically designed for GNN neural architecture search (GNN-NAS) have been proposed (Liu and Liu 2023; Huan, Quanming, and Weiwei 2021). GNN-NAS can be summarized as a combinatorial optimization problem. Limited by the quantity and quality of SGNNs designed by current experts, it is difficult to

*Corresponding author.

adapt to a wider range of application scenarios. Additionally, while many efforts have been made to design adaptive spectral GNNs (Chien et al. 2021; He et al. 2021), their generalization capability remains constrained by issues such as homogeneity assumptions and oversmoothing problem.

Recently, Large Language Models (LLMs) have garnered significant attention for their powerful capabilities and versatility in solving natural language processing tasks such as text generation and question answering. Studies such as (Liu et al. 2024; Romera-Paredes et al. 2024; Shojaee et al. 2024) have demonstrated that LLMs can be used to discover scientific equations and mathematical algorithms, providing opportunities for numerous fields, including graph learning.

In this paper, we propose AutoSGNN, a framework for the automatic discovery of spectral GNNs. By leveraging LLMs, AutoSGNN can identify or generate suitable spectral GNNs tailored to various conditions. Specifically, we design a search space for spectral GNNs that encompasses the feature fitting terms, the graph Laplacian regularization terms, and the aggregation terms. This search space is described through a combination of text and code, enabling LLMs to understand and generate appropriate spectral GNNs. Then, based on a unified spectral GNN architecture, we establish an evolutionary framework to collaboratively evolve the design thoughts and code of spectral GNNs. With the assistance of LLMs, the linguistic description and code of heuristics evolve iteratively, guided by curated prompts, to generate the most suitable spectral GNNs for different graphs. In summary, AutoSGNN distinguishes itself from previous works through some key features:

- AutoSGNN introduces an automated approach for designing feature fitting terms, graph Laplacian regularization terms, and aggregation terms in spectral GNNs. This method is capable of generating suitable spectral GNNs for a wide range of graph datasets that encompass both homogeneous and heterogeneous graphs.
- We propose the fourth type of method for GNN-NAS, distinct from existing three types of approaches that rely on gradient descent, reinforcement learning, or evolutionary algorithms. AutoSGNN achieves a high degree of parallelism throughout the entire process, from generation to evaluation.
- In contrast to GNN-NAS methods, which are essentially combinatorial optimization techniques, AutoSGNN goes beyond by truly generating GNN frameworks through learning and understanding the fundamental design principles of filters.
- We conducted a comprehensive evaluation of AutoSGNN across nine widely studied graph datasets. Extensive experimental results demonstrate that AutoSGNN exhibits exceptional competitiveness in both algorithmic performance and time complexity, outperforming or matching state-of-the-art spectral GNNs and GNN-NAS methods.

Related Work

Spectral Graph Neural Network Spectral GNNs design spectral graph filters in the spectral domain. ChebNet

(Defferrard, Bresson, and Vandergheynst 2016b) employs higher-order Chebyshev polynomials to approximate the filters. GCN (Kipf and Welling 2017) uses a first-order approximation of the Chebyshev filters to simplify the process. GPRGNN (Chien et al. 2021) approximates polynomial filters through gradient descent on the polynomial coefficients. FAGCN (Bo et al. 2021) learns the weights of low-pass and high-pass filters by approximating them with attention coefficients. BernNet (He et al. 2021) approximates any non-negative coefficient filter in the normalized Laplacian spectrum of the graph using k-order Bernstein polynomials. NFGNN (Zheng et al. 2024) creates a local filter by combining the Kronecker delta function with polynomial approximations, and accelerates computational efficiency while balancing local and global spectral information by reparameterizing the feature parameters.

Despite the emergence of various spectral GNNs, their graph learning mechanisms can mostly be viewed as optimizing the feature fitting functions of various graph kernels with graph regularization terms. (Zhu et al. 2021) optimizes two proposed objective functions, allowing adjustable high-pass or low-pass filters to fit a wide range of GNNs.

Neural Architecture Search Recently, NAS has been applied to GNNs to automate their design process. GraphNAS (Gao et al. 2021) and AutoGraph (Li and King 2020) utilize reinforcement learning and genetic algorithms, respectively, to find the optimal GNN architecture. Recent methods like SANE (Huan, Quanming, and Weiwei 2021) employ differentiable NAS, reducing computational costs through gradient-based optimization. CTFGNAS (Liu and Liu 2023) introduces an alternative evolutionary Graph NAS (GNAS) algorithm, whose search space includes not only micro-level network layer components but also topological connections and feature fusion strategies. These approaches demonstrate that NAS has the potential to enhance GNN performance across various tasks.

LLM for Graph In recent years, LLMs such as ChatGPT (Ouyang et al. 2022) have been applied to fields like mathematical formula discovery due to their outstanding text processing and generation capabilities. (Romera-Paredes et al. 2024) applies LLMs to heuristic algorithm discovery. By adding heuristic prompts that approximate the chain of thought, (Liu et al. 2024) can discover heuristic algorithms more rapidly. Meanwhile, (Shojaee et al. 2024) utilized LLMs to discover scientific equations. The inherent rich domain knowledge and contextual understanding of LLMs enable LLM-based search architectures to be more intelligent, flexibly generating suitable architectures based on different datasets and task requirements. Additionally, Evolutionary Strategy (ES) can provide an optimization framework (Li et al. 2024a,b; Wu et al. 2023; Guo et al. 2024) for further enhancing LLMs in a black-box setting, endowing LLMs with flexible global search capabilities.

Preliminaries

Notations A graph can be represented as $G = (V, E, A)$, where $V = \{\nu_i\}_{i=1}^n$ denotes the set of nodes with $|V| = n$,

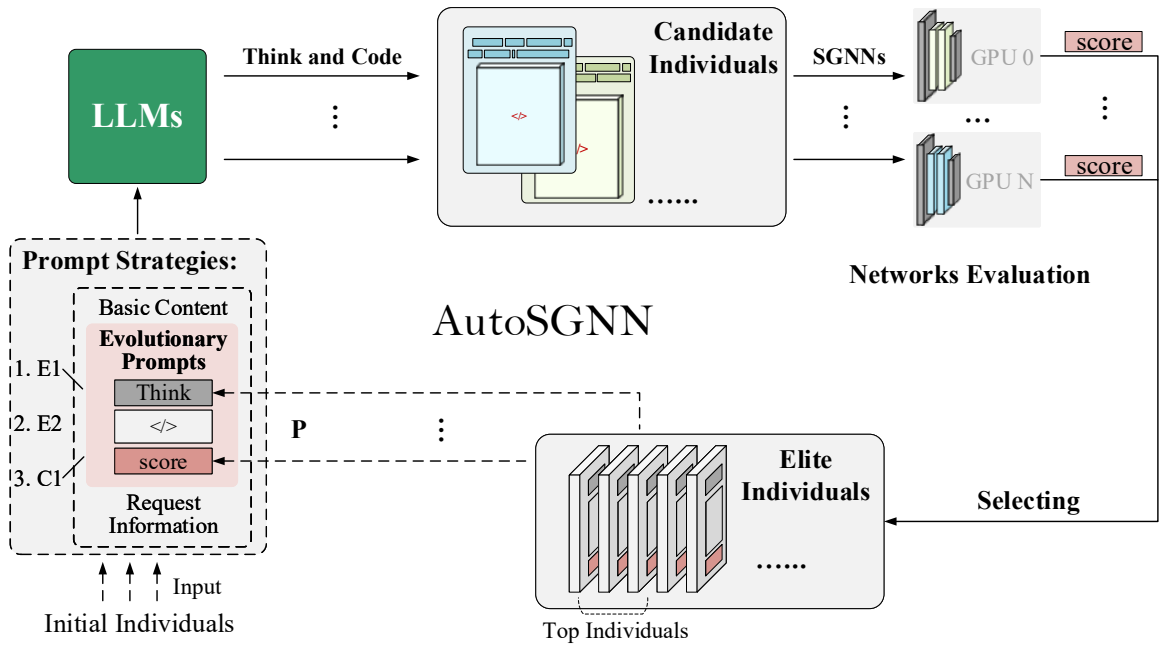


Figure 2: The overview of AutoSGNN. In this process, the LLM is used to generate population individuals, while other components are responsible for evaluation, selection, and describing requirements to optimize the LLM preferences, thus achieving the overall optimization process.

E is the set of edges between nodes, and the topology of the graph G can be represented by the adjacency matrix $A \in R^{n \times n}$. When $A_{ij} = 1$, it means that node v_i is connected to node v_j ; otherwise $A_{ij} = 0$. The features of each node can be represented as $X \in R^{n \times f}$, where f is the node feature dimension. The pairwise angle matrix is denoted as $D = \text{diag}(d_1, \dots, d_n)$, where $d_i = \sum_j A_{i,j}$.

Graph Fourier Transform The (combinatorial) graph Laplacian is defined as $L = D - A$, which is Symmetric Positive Semi-Definite (SPSD). Its eigendecomposition gives $L = U\Lambda U^T$, where the columns u_i of $U \in R^{n \times n}$ are orthonormal eigenvectors, namely the graph Fourier basis, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, where $\lambda_1 \leq \dots \leq \lambda_n$, and these eigenvalues are also called frequencies. Therefore, the graph Fourier transform of signal x can be defined as $\tilde{x} = U^{-1}x = U^T x$, while the inverse Fourier transform is $x = U\tilde{x}$, and the transform of the signal x through filter $\tilde{g} = f(\lambda) = \text{diag}(f(\lambda_1), \dots, f(\lambda_n))$ can be expressed as $z = U\tilde{g}U^T x$. In addition to L , there exist some common variants of this graph Laplacian matrix such as the symmetric normalized Laplacian matrix $L_{sym} = I - D^{-1/2}AD^{-1/2}$, and random walk normalized Laplacian Matrix $L_{rw} = I - D^{-1}A$. Applying the renormalization trick to affinity and Laplacian matrices respectively leads to $\tilde{L}_{sym} = I - \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ and $\tilde{L}_{rw} = I - \tilde{D}^{-1}\tilde{A}$, where $\tilde{A} = A + I$, $\tilde{D} = D + I$. In addition to this, (Zheng et al. 2024) uses $\tilde{L} = \frac{2L_{sym}}{\lambda_{max}} - I$ to limit the range of eigenvalues in order to satisfy numerical stability. For the processing of graph Laplacian matrices, it is often necessary to consider the structure of the graph, signal processing requirements and other conditions in order to

select a suitable form.

Spectral Graph Neural Networks Spectral-based GNNs aim to learn specific spectral filters given a graph structure and node labels. By optimizing the optimal solution of feature fitting functions of various graph kernels with a graph regularization term, they retain appropriate frequency components for downstream tasks. By summarizing spectral GNNs, (Zhu et al. 2021) proposed two unified frameworks for spectral GNNs. For GNNs with layer-wise feature transformation (e.g., GCN), the output of the K th layer directly uses the output of the $(K-1)$ th layer. The learning process can be represented as:

$$Z^K = \left\langle \text{Trans}(\text{Agg}\{G; Z^{(K-1)}\}) \right\rangle_K \quad (1)$$

where Z^K is the graph embedding of the K th layer, and $Z^{(0)} = X$. Whereas for some networks that separate $\text{Trans}()$ and $\text{Agg}\{\}$ (e.g., FAGCN (Bo et al. 2021)), the K th layer may perform operations such as pooling, concatenation, or attentional aggregation on the outputs of the $(K-1)$ th layer, and then the learning process can be represented as:

$$Z^K = \left\langle \text{Agg}\{G; Z^{(K-1)}\} \right\rangle_K \quad (2)$$

where $Z^{(0)} = \text{Trans}(X)$, and $\text{Trans}()$ is the feature transformation operation at the corresponding layer, including the non-linear activation function and the learnable weight matrix W of the specific layer. $\text{Agg}\{\}$ means aggregating the $(K-1)$ layer output $Z^{(K-1)}$ along graph G for the K th learning operation according to the set spectral filter. The operator

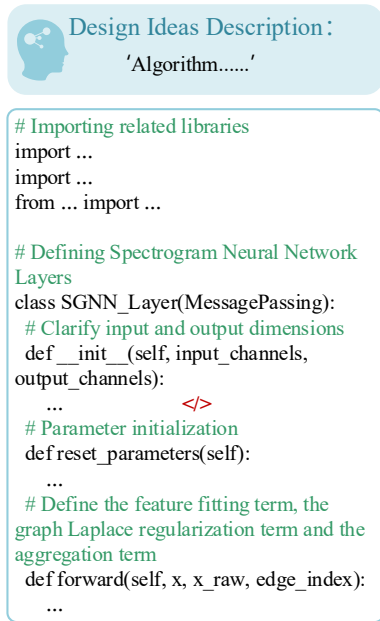


Figure 3: Individual Content Schematic.

$\langle \rangle_K$ depends on specific GNNs and represents the generalized combinatorial operation after K convolutions.

By simplifying and integrating Eqs.(1) and (2), a spectral GNNs unified architecture can be obtained:

$$Z^K = \left\langle TRANS(AGG\{G; X^{raw}; Z^{(K-1)}\}) \right\rangle_K \quad (3)$$

where $X^{raw} = Linear(X^{(0)})$. $AGG\{\}$ can also combine operations such as pooling, concatenation or attention aggregation, and $TRANS()$ can combine dimensional scaling factors such as $Linear()$. In summary, it is not difficult to see that when suitable graph regularization terms, spectral filter approximation terms, and aggregation terms are determined, a spectral GNN can be obtained.

AutoSGNN

Figure 2 shows an overview of AutoSGNN, consisting of five parts: Elite Individuals, Candidate Individuals, Prompt Strategies, LLMs and Networks Evaluation.

Elite and Candidate Individuals

Spectral graph learning mechanisms can mostly be seen as optimizing the feature fitting functions of various graph kernels with graph regularization terms. Therefore, in AutoSGNN, the LLM is used to search for suitable graph regularization, feature approximation terms, and aggregation terms for different graphs, thus generating the most appropriate spectral GNNs.

We use code text descriptions to represent spectral GNNs. Through the formatted output of the LLMs, candidate individuals in the entire evolutionary framework are formed, as shown in Figure 3. The N individuals in candidate individuals, each consist of two parts: the design ideas and the code of the spectral GNNs (containing the feature fitting

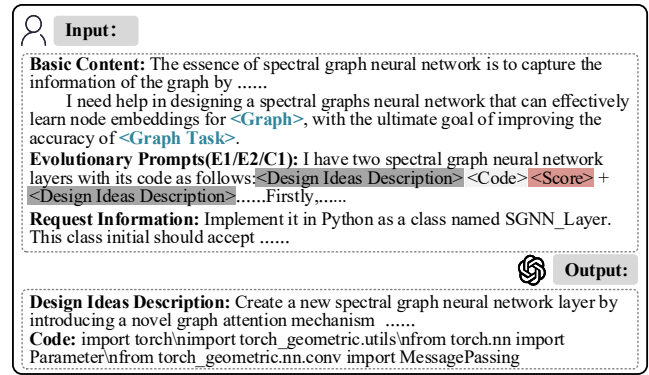


Figure 4: Prompt Strategies for LLMs.

terms, graph Laplacian regularization terms, and aggregation terms). The spectral GNNs is specifically represented as a Python class, corresponding to the unified architecture of spectral GNNs in Eq.(3). By initializing the network's input and output dimensions, it facilitates message passing between layers, and the Design Ideas Description provides a brief introduction to the neural network.

Prompt Strategies

Although LLMs have been widely applied, their black-box nature and limited search capabilities hinder their performance in complex optimization problems. Evolution Strategies(ES), a classic black-box optimization technique (Pošík, Huyer, and Pál 2012), can provide an optimization framework for LLMs in black-box environments, endowing LLMs with flexible global search capabilities. AutoSGNN combines the advantages of ES and LLMs, designing three types of evolutionary prompts(E1, E2, C1) to quickly understand the preferences of different datasets, enabling the selection or generation of optimal spectral GNNs. As shown in Figure 4, the three types of prompts consist of basic information, evolutionary strategies, and requirement information. Specific prompts for different datasets are detailed in **Appendix B**.

Basic Content A short introduction to the application of spectral GNNs, including the information of graph and downstream tasks. Some expert knowledge can be provided by setting up tips.

Evolutionary Prompts

- $E1$. Exploring new spectral GNN that are completely different from the given spectral GNN. First, select P_1 elite individuals from the Elite Individuals, then prompt the LLM to design a spectral GNN different from the chosen elite individuals, accompanied by a brief textual description. It's like the mutation operation in an evolutionary algorithm.
- $E2$. Based on the observation of the given elite individuals, new spectral GNN is explored. First, P_2 individuals are selected from elite individuals. Then, the LLM is instructed to identify the common ideas of the existing spectral GNNs. Finally, a new spectral GNN is designed,

	Cora	Citeseer	PubMed	Computer	Photo	Chameleon	Squirrel	Texas	Cornell	Rank
GTFGNAS	75.72±2.06	64.24±1.85	83.35±1.41	<u>84.98±1.41</u>	90.52±0.96	37.54±5.84	31.47±1.73	<u>57.03±3.47</u>	37.38±2.23	4
F2GNN	65.33±1.26	58.16±2.28	81.16±0.69	46.81±4.01	82.03±3.13	35.72±1.86	24.52±1.41	50.06±4.47	<u>48.44±7.58</u>	13
Genetic-GNN	76.58±1.06	60.98±1.67	84.01±0.46	83.30±2.10	87.33±1.37	38.54±2.00	29.81±1.09	44.45±6.34	45.74±7.87	8
SANE	73.52±3.58	61.57±2.62	82.91±0.26	80.41±2.15	87.56±1.70	41.79±2.33	28.79±1.90	49.71±2.92	35.61±13.32	11
APPNP	<u>79.41±0.38</u>	68.59±0.30	85.02±0.09	81.99±0.26	91.11±0.26	30.12±0.92	26.01±0.48	46.44±3.10	44.37±4.90	7
GPRGNN	79.51±0.36	67.63±0.38	<u>85.07±0.09</u>	82.90±0.37	91.93±0.26	31.46±0.89	26.14±0.51	44.76±4.33	43.45±4.65	5
FAGCN	78.10±0.21	66.77±0.18	84.09±0.02	82.11±1.55	90.39±1.34	37.24±3.54	26.91±3.08	48.44±1.78	46.38±1.82	6
BernNet	76.37±0.36	65.83±0.61	82.57±0.17	79.57±0.28	91.60±0.35	27.33±1.14	26.42±0.39	48.21±3.17	46.50±5.57	10
JacobiConv	75.30±0.02	62.75±0.02	85.05±0.01	87.89±0.01	<u>92.34±0.01</u>	39.13±0.02	25.91±0.01	56.34±0.05	44.80±0.04	3
NFGNN	77.69±0.91	<u>67.74±0.52</u>	<u>85.07±0.13</u>	84.18±0.40	92.16±0.82	<u>40.10±1.34</u>	30.91±0.62	50.37±6.54	48.37±7.34	2
GCN	75.21±0.38	67.30±0.35	84.27±0.01	82.52±0.32	90.54±0.21	35.73±0.87	28.37±0.77	32.40±2.31	35.90±3.64	9
ChebNet	71.39±0.51	65.67±0.38	83.83±0.12	82.41±0.28	90.09±0.28	32.36±1.61	27.76±0.93	28.51±4.03	26.21±3.12	12
AutoSGNN	80.61±1.52	67.23±1.31	85.13±0.26	84.80±0.82	92.51±0.36	38.67±3.52	<u>31.17±1.82</u>	57.10±4.86	50.87±6.50	1

Table 1: The accuracy of node classification under the sparse splitting ratios of 2.5%:2.5%:95%, with the best results highlighted in bold black and the second-best results underlined. The results are ranked by employing the Wilcoxon-Holm analysis (Ismail Fawaz et al. 2019) at a significance level of $p = 0.05$.

based on these common ideas, but differing from the existing spectral GNNs by introducing new elements, along with a brief textual description. It’s like the crossover operation in an evolutionary algorithm.

For E1 and E2, we need to select P_1 and P_2 individuals to form the prompts. We randomly select 50% individuals from the top 20% based on scores, and 50% from the remaining individuals. This approach helps AutoSGNN maintain the optimization trend.

- C1. By comparing individuals, LLM preferences are set. Inspired by algorithms such as DPO (Rafailov et al. 2024), two elite individuals are selected for comparative learning to directly set LLM preferences. First, two individuals are randomly chosen from the top 1/3 and bottom 1/3 by score within Elite Individuals, respectively. Then, the LLM is asked to compare their similarities and differences and hypothesize why the higher-score individual is superior. Finally, a more optimal spectral GNN is designed, accompanied by a brief textual description.

Request Information The request information supplements the initialization of the generated spectral GNNs, including input dimensions, output dimensions, input variables, output variables, and the end-of-code marker, "Do not give additional explanations." By specifying these requirements to the LLM, the basic framework of the returned code is constrained, improving the executability of the generated spectral GNNs.

LLMs

In AutoSGNN, a wide variety of LLMs, such as ChatGPT (Ouyang et al. 2022), can be flexibly integrated. These LLMs can be open-source models (e.g., GPT-4 or the LLaMA series), proprietary commercial products, or variants fine-tuned for specific tasks or domains. This flexibility allows AutoSGNN to fully leverage the powerful capabilities of LLMs in generating, reasoning, and optimizing GNN architectures.

Networks Evaluation

By evaluating the score of candidate individuals on different datasets, it helps to adjust the LLMs to generate preferences

for different situations. The scoring metric can be selected as downstream task metrics such as classification accuracy and loss value. After scoring the candidate individuals, an elite strategy is used to select individuals, retaining a certain number of individuals into Elite Individuals, which are initialized by selecting some classic methods of spectral GNNs (Kipf and Welling 2017; Bo et al. 2021; Chien et al. 2021; He et al. 2021; Zheng et al. 2024).

The entire execution process of AutoSGNN is highly parallel, from the parallel responses of the LLM to the parallel evaluation of each GNN model composed of the generated spectral GNNs on different GPUs. Furthermore, to enhance efficiency, we set a timeout duration (*Timeout Duration* = 600s). If an SGNN fails to complete its training and testing within this timeframe, the process is terminated, and the SGNN individual is discarded.

Experiments

Experiment Setting

Datasets We utilize five widely-used homophilic graphs, including the citation graphs Cora, CiteSeer, and PubMed, as well as the Amazon co-purchase graphs Computers and Photo. In addition, we employed four heterophilic benchmark datasets, including the Wikipedia graphs Chameleon and Squirrel, and the webpage graphs Texas and Cornell from WebKB. Detailed information about these datasets can be found in **Appendix A**. Following (Zheng et al. 2024; Chien et al. 2021), for the node classification task under the transductive setting, we randomly sparse split the node set into train/val/test datasets with ratio 2.5%:2.5%:95%. We have also included the experimental results for the randomly dense split (60%:20%:20%) in **Appendix F**.

Baselines We evaluated the performance of AutoSGNN by comparing it with several baselines: 1) **graph NAS methods**: GTFGNAS (Liu and Liu 2023), F2GNN (Wei, Zhao, and He 2022), Genetic-GNN (Shi et al. 2022), and SANE (Huan, Quanming, and Weiwei 2021). 2) **Spectral GNNs**: Since AutoSGNN is designed to generate spectral GNNs for specific tasks and datasets, we selected five baselines—APPNP (Gasteiger, Bojchevski, and Günnemann 2019), GPRGNN (Chien et al. 2021), FAGCN (Bo et al.

Graph	Propagation Mechanism	Relevant Parameter
Cora	$Z_0 = \alpha X$ $Z_k = Z_k + \sum_{k=1}^{K-1} \alpha(1-\alpha)^k \hat{A} Z_{k-1} W_k$ $Z_K = (1-\alpha)^K \hat{A} Z_{K-1} W_K + Z_{K-1}$	<i>Initial</i> $\alpha = 0.15, K = 4$ $\hat{A} = \tilde{D}^{-1/2} \tilde{A} (\tilde{D}^T)^{-1/2}, \tilde{A} = A + 4I, \tilde{D} = D + 4I$
Citeseer	$Z_0 = Att \cdot X$ $Z_K = Z_0 + \sum_{k=1}^{K-1} (Att \cdot \hat{A} Z_{k-1} W_k + 0.2 X_{raw})$	<i>Initial</i> $Att \in R^{1 \times n}, K = 4$ $\hat{A} = \tilde{D}^{-1/2} \tilde{A} (\tilde{D}^T)^{-1/2}, \tilde{A} = A + 2I, \tilde{D} = D + 2I$
PubMed	$Z_0 = \alpha X + \gamma X_{raw}$ $Z_K = Z_0 + \sum_{k=1}^K \beta \cdot ReLU(\bar{A} Z_{k-1} W_k)$	<i>Initial</i> $\alpha = 0.25, \beta = 0.4, \gamma = 0.25, K = 4$ $\tilde{A} = A + 2I$ $\bar{A}_{(i,j)} = \frac{\hat{A}_{(i,j)}}{\sum_{(i,j) \in E} \hat{A}_{(i,j)} + \epsilon}$ if $\hat{A}_{(i,j)} \geq (\mu - \sigma)$ $\bar{A}_{(i,j)} = 0$ if $\hat{A}_{(i,j)} < (\mu - \sigma)$ $\mu = \frac{\sum_{(i,j) \in E} \hat{A}_{(i,j)}}{ E }, \sigma = \sqrt{\frac{\sum_{(i,j) \in E} (\hat{A}_{(i,j)} - \mu)^2}{ E }}$
Computer	$Z = \tilde{\alpha}_0 X + \tilde{\alpha}_1 \hat{A} X W_1 + \tilde{\alpha}_2 \hat{A}^2 X W_2$	<i>Initial</i> $\alpha = 0.1, K = 2$ $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_K = \frac{\alpha^0, \alpha^1, \dots, \alpha^K}{\sum_{k=0}^K \alpha^k }$ $\hat{A} = \tilde{D}^{-1/2} \tilde{A} (\tilde{D}^T)^{-1/2}, \tilde{A} = A + 2I, \tilde{D} = D + 2I$
Photo	$Z_0 = (\beta + \gamma) X$ $Z_K = Z_0 + \sum_{k=1}^K \beta \gamma \hat{A} Z_{k-1} W_k$	<i>Initial</i> $\gamma = 0.3, \beta = 0.7, K = 3$ $\hat{A} = \tilde{D}^{-1/2} \tilde{A} (\tilde{D}^T)^{-1/2}, \tilde{A} = A + 2I, \tilde{D} = D + 2I$
Chameleon	$Z = \tanh((X W_1 + b) + \frac{1}{1+e^{-\alpha}} \tilde{A} X_{raw} W_2)$	<i>Initial</i> $\alpha = 1$ $\hat{A} = D^{-1/2} A (D^T)^{-1/2}$
Squirrel	$X_{Att} = \text{softmax}(X \cdot Att, \text{dim} = 1)$ $Z_0 = \hat{A}(X W_0 + b_0) W_1$ $Z_1 = \hat{A}(Z_0 W_2 + b_1) W_3$ $Z = Z_1 \cdot X_{Att} \cdot W_3 + b_1$	<i>Initial</i> $Att \in R^{n \times 1}$ $\hat{A} = \tilde{D}^{-1/2} \tilde{A} (\tilde{D}^T)^{-1/2}, \tilde{A} = A + 2I, \tilde{D} = D + 2I$
Texas	$Z = W_2 \cdot ELU((W_1 \cdot X_{raw} + \sum_{k=1}^K \tilde{A}^k \cdot X) \cdot Att)$	<i>Initial</i> $Att \in R^{n \times 1}, K = 4$ $\hat{A} = D^{-1/2} A (D^T)^{-1/2}$
Cornell	$Z = \hat{A} X W_1 + \sum_{j \in N(i)} \alpha_{(i,j)} W_2 X_j^{raw}$	$\hat{A} = D^{-1/2} A (D^T)^{-1/2}$

Table 2: The summary results of the spectral graph neural networks searched by AutoSGNN for different graphs for node classification task under the sparse splitting ratios of 2.5%:2.5%:95%.

2021), BernNet (He et al. 2021), JacobiConv(Wang and Zhang 2022) and NFGNN (Zheng et al. 2024)—along with two classic spectral GNNs, GCN (Kipf and Welling 2017) and ChebNet (Defferrard, Bresson, and Vandergheynst 2016b).

Reproducibility For AutoSGNN, the iterative search algebra is set to 30, and the evolutionary strategy prompts for $E1$ and $E2$ both contain $P1 = P2 = 4$ elite individuals. The GPT-3.5-turbo pre-trained LLM is used, with the parallel response number set to 4. Through parallel experiments with 3 prompts, the number of candidate individuals generated per cycle is 12. We use node classification accuracy on the validation dataset as the fitness metric, saving the top 30 individuals per cycle as elite individuals. We employ 2-layer generated spectral GNN layers and use 2-layer MLP for feature transformation. Similar to other spectral GNNs baselines, we follow the experimental setup in (Zheng et al. 2024) and use the best hyperparameter combinations provided in the original papers for each dataset. For GNN-NAS baselines, we follow the experimental setup in (Liu and Liu 2023), and like AutoSGNN, we set the population size and iterative search algebra to 30. To prevent overfitting during the evaluator’s training for each SGNN, we set an early stopping criterion of 200 epochs.

All experiments, in order to remove the effects of randomness, are run independently for 10 trials and the resulting

means and standard deviations are reported. All search experiments are conducted on a machine equipped with four NVIDIA GeForce RTX 3090 GPUs, two Intel(R) Xeon(R) Silver 4210 CPUs (2.20 GHz), and 252GB of RAM.

AutoSGNN v.s. Human-Designed GNN

To demonstrate the effectiveness of AutoSGNN, we validate the network architecture obtained through the search process using sparse node classification tasks. In Table 1, compared to the GNN-NAS methods, AutoSGNN outperforms in most datasets. For the nine datasets, seven of them are ranked first or second best result, except for the Citeseer and Chameleon datasets. For advanced manually designed spectral GNNs, it is observed that they do not always perform well on every dataset. For instance, GPRGNN achieved the second-best node classification accuracy on the PubMed dataset but performed poorly on the Squirrel dataset. This indicates that different network architectures may have their own limitations. Selecting different network architectures for different situation can better adapt to various graph datasets.

Case Study on Cora

As shown in Table 2, AutoSGNN can search for suitable spectral GNNs for different datasets and provide corresponding hyperparameters. More detailed information can be found in **Appendix C**. For the Cora dataset, the network generated by AutoSGNN can be regarded as a full-frequency

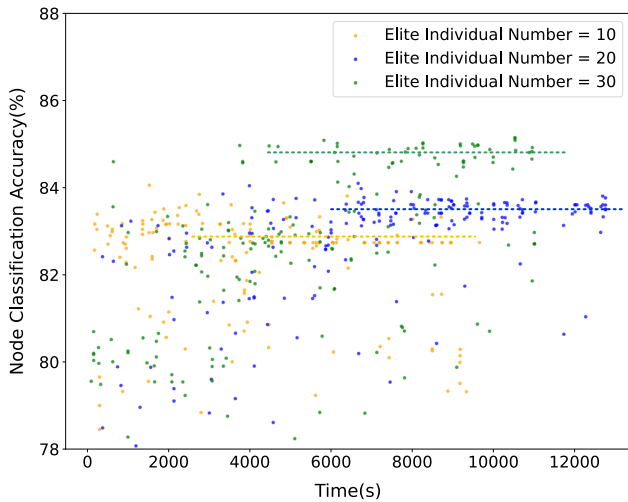


Figure 5: The iterative search process of AutoSGNN with different Elite Individual numbers for node classification experiment on the PubMed dataset. Each dot represents the node classification result of a searched candidate individual on the PubMed dataset. The dashed line represents the overall convergence trend of elite individuals.

learning filter composed of multiple filters. When $k = 0$, $Z_0 = \alpha X$, the initial graph features X are learned, which can be considered as a low-frequency filter. When $k = K$, in the equation $Z_K = (1 - \alpha)^K \hat{A}Z_{K-1}W_K + Z_{K-1}$, the term $(1 - \alpha)^K \hat{A}Z_{K-1}W_K$ can be seen as a high-frequency filter. For the intermediate layer k , the term $\alpha(1 - \alpha)^k \hat{A}Z_{k-1}W_k$ acts as a filter learning specific frequencies. Therefore, the entire propagation mechanism can be represented as a process of learning and processing graph signals across the entire spectrum. Additionally, AutoSGNN determines the hyperparameters $K = 4$ and initial $\alpha = 0.15$, and the graph kernel \hat{A} , which determine the relevant properties in the frequency domain.

Parameter Analysis

This section explores the impact of various components and hyperparameters on AutoSGNN.

No. of Elite Individuals As shown in Figure 5, we compared the search process for Elite Individual numbers of 10, 20, and 30. It is evident that a larger elite individual number allows for a broader search space, increasing the likelihood of finding the most suitable network architecture. Conversely, when the Elite Individual number is smaller, the population tends to converge to network architectures with generally mediocre performance.

Choices of LLMs From Figure 6, for the PubMed dataset, the spectral GNNs generated using AutoSGNNs from ChatGPT 4o achieved a classification score of 85.43%, about 0.3% higher than the score obtained using ChatGPT 3.5. We can also observe that the use of more powerful LLMs in AutoSGNN results in a more stable search process. By

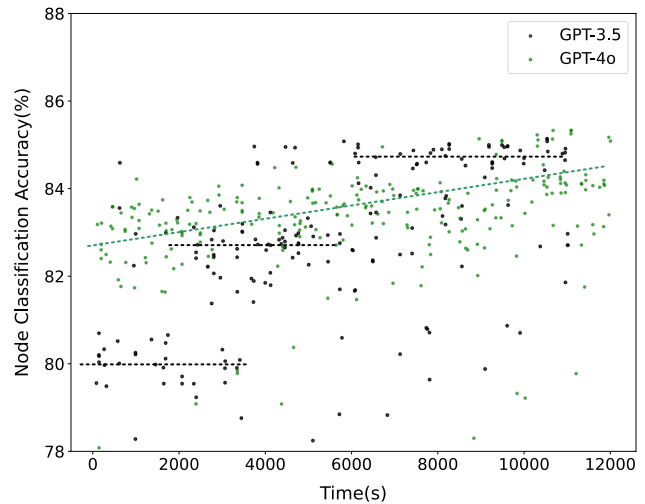


Figure 6: The iterative search process of AutoSGNN with different LLMs for the node classification experiment on the PubMed dataset.

AutoSGNN	w.o. E1	w.o. E2	w.o. C1
85.13±0.26	84.88±0.43	84.54±0.59	83.95±0.67

Table 3: The accuracy of AutoSGNN and its variants on the PubMed dataset, where w.o. means remove prompt of this type.

utilizing ChatGPT-4o, the prompts’ requirements are better understood, enabling a more accurate connection between the various preferences of each dataset, thus facilitating the search for an appropriate spectrum GNN. Additionally, by comparing the number of points in different colors in the figure, we can see that the use of more advanced LLMs significantly enhances the executability of the response code.

Number of Elite Individuals in Prompt Operators. We explore the impact of $P1$ and $P2$, extracted per iteration by the evolutionary prompts ($E1/E2$) on AutoSGNN. We set the number of individuals for the operations to $P1 = P2 = 2, 3$ and 4 , respectively. The experimental results are shown in Figure 7. From the figure we can see that when $P1 = P2 = 4$, the entire search process of AutoSGNN is more effective. When $P1 = P2 = 2$, the performance is the second best, and when $P1 = P2 = 3$, the performance is the worst. We believe this may be because when $P1 = P2 = 2$, all types of prompts can be considered as preference-setting prompts ($C1$), leading to better search effectiveness compared to when $P1 = P2 = 3$. However, relying solely on preference-setting prompts ($C1$) is not sufficient. As shown in Table 3, the absence of evolutionary prompts ($E1/E2$) still impacts the performance of AutoSGNN.

Ablation Study

We explore the impact of the three types of prompts on the AutoSGNN search process. Since removing one type of prompt causes AutoSGNN to generate only 8 candidate

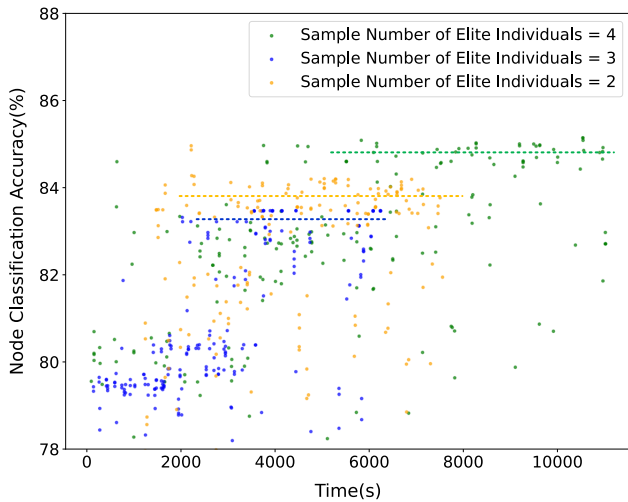


Figure 7: The iterative search process of AutoSGNN with different number of elite individuals in prompt operators for node classification experiment on the PubMed dataset. Each dot represents the node classification result of a searched candidate individual on the PubMed dataset. The dashed line represents the overall convergence trend of elite individuals.

individuals per iteration, we set the number of iterations to 45 to ensure a fair evaluation, thereby generating the same number of candidate individuals. The results in Table 3 show that removing any type of prompt strategy affects the search performance of AutoSGNN. Particularly, the removal of the preference-setting prompt ($C1$) has the most significant weakening effect. This may be because without the preference-setting prompt ($C1$), the LLMs cannot quickly and effectively align with task preferences, leading to many ineffective searches.

Time Complexity Analysis

To compare the search efficiency of GNN-NAS methods and AutoSGNN, as shown in Figure 8, we recorded the time required for each method to complete a full run under the experimental settings described earlier. In Figure 8, AutoSGNN is competitive. More detailed information can be found in **Appendix E**.

Generalization of AutoSGNN

In our exploration of spectral GNNs generated by AutoSGNN, we examined their generalization capability across different types of graph. The result in **Appendix D** illustrates that SGNNs optimized for a specific dataset generally perform the best on that dataset. However, when these SGNNs are applied to different types of graphs, they frequently fail to achieve optimal performance. This finding highlights a limitation in the generalizability of SGNNs: Their effectiveness is often contingent upon the alignment between their information processing mechanisms and the specific information requirements of downstream tasks. Thus, automatically tailoring SGNNs to match the characteristics of differ-

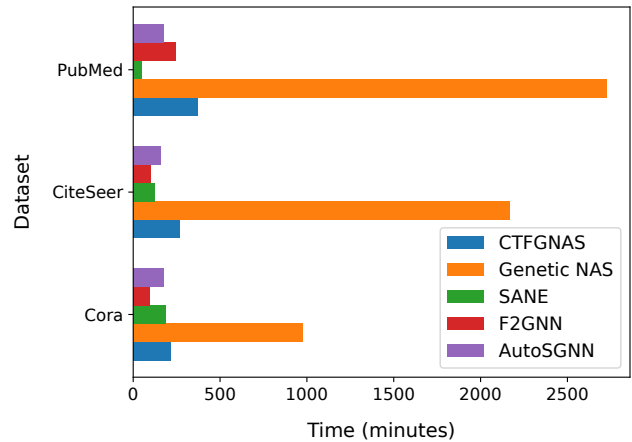


Figure 8: Computational time (clock time in minutes) of GNN-NAS methods and AutoSGNN tested on the Pubmed dataset.

ent graph types is advisable. For further details, please refer to **Appendix D**.

Conclusion and Future Work

This paper proposes a spectral GNNs generation method (AutoSGNN), which leverages the synergy between LLMs and ES. By utilizing three types of prompt, it sets the preferences of LLMs to adaptively search for spectral GNNs suitable for different graph. AutoSGNN is applied to nine widely used graph datasets, and the experimental results show that the spectral GNNs generated by AutoSGNN are more suitable than advanced manually designed spectral GNNs. Compared to GNN-NAS methods, AutoSGNN outperforms in both accuracy and time efficiency. Notably, unlike GNN-NAS which can be viewed as a combinatorial optimization problem, AutoSGNN is a true automatic propagation mechanism generation method assisted by LLMs.

In summary, we have developed a tailored search space for spectral GNNs specifically designed for LLMs. By incorporating ES, we enable LLMs to effectively learn and utilize SGNNs. This novel approach enhances the ability of LLMs to perceive, analyze, and describe graphs, offering a new perspective on the construction of graph foundation models. In future work, we plan to explore further applications of LLMs in various graph-based tasks and investigate the use of AutoSGNNs with real-world graph data.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62471371, 62306224 and 62206205, in part by the Young Talent Fund of Association for Science and Technology in Shaanxi, China under Grant 20230129, in part by the Science and Technology Program of Guangzhou, China under Grant 2023A04J1060, in part by the Guangdong High-level Innovation Research Institution Project under Grant 2021B0909050008, in part by the Guangzhou Key Research

and Development Program under Grant 202206030003, and in part by the Fundamental Research Funds for the Central Universities under Grant XJSJ23097.

References

- Bo, D.; Wang, X.; Shi, C.; and Shen, H. 2021. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Chien, E.; Peng, J.; Li, P.; and Milenkovic, O. 2021. Adaptive universal generalized PageRank graph neural network. In *International Conference on Learning Representations*.
- Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016a. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*.
- Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016b. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*.
- Fout, A.; Byrd, J.; Shariat, B.; and Ben-Hur, A. 2017. Protein interface prediction using graph convolutional networks. In *Advances in Neural Information Processing Systems*.
- Gao, Y.; Yang, H.; Zhang, P.; Zhou, C.; and Hu, Y. 2021. Graph neural architecture search. In *International Joint Conference on Artificial Intelligence*.
- Gasteiger, J.; Bojchevski, A.; and Günnemann, S. 2019. Predict then propagate: Graph neural networks meet personalized pagerank. In *International Conference on Learning Representations*.
- Guo, H.; Ma, Z.; Chen, J.; Ma, Y.; Cao, Z.; Zhang, X.; and Gong, Y.-J. 2024. ConfigX: Modular Configuration for Evolutionary Algorithms via Multitask Reinforcement Learning. *arXiv preprint arXiv:2412.07507*.
- Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*.
- He, M.; Wei, Z.; Xu, H.; et al. 2021. BernNet: Learning arbitrary graph spectral filters via bernstein approximation. In *Advances in Neural Information Processing Systems*.
- Huan, Z.; Quanming, Y.; and Weiwei, T. 2021. Search to aggregate neighborhood for graph neural network. In *2021 IEEE 37th International Conference on Data Engineering*.
- Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4): 917–963.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.
- Li, X.; Wu, K.; Li, Y. B.; Zhang, X.; Wang, H.; and Liu, J. 2024a. Pretrained Optimization Model for Zero-Shot Black Box Optimization. In *Advances in Neural Information Processing Systems*.
- Li, X.; Wu, K.; Zhang, X.; and Wang, H. 2024b. B2Opt: Learning to Optimize Black-box Optimization with Little Budget. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Li, Y.; and King, I. 2020. Autograph: Automated graph neural network. In *International Conference on Neural Information Processing*.
- Liu, F.; Xialiang, T.; Yuan, M.; Lin, X.; Luo, F.; Wang, Z.; Lu, Z.; and Zhang, Q. 2024. Evolution of heuristics: Towards efficient automatic algorithm design using large language model. In *International Conference on Machine Learning*.
- Liu, Y.; and Liu, J. 2023. A surrogate evolutionary neural architecture search algorithm for graph neural networks. *Applied Soft Computing*, 144: 110485.
- Mao, H.; Chen, Z.; Jin, W.; Han, H.; Ma, Y.; Zhao, T.; Shah, N.; and Tang, J. 2024a. Demystifying structural disparity in graph neural networks: Can one size fit all? In *Advances in Neural Information Processing Systems*.
- Mao, H.; Chen, Z.; Tang, W.; Zhao, J.; Ma, Y.; Zhao, T.; Shah, N.; Galkin, M.; and Tang, J. 2024b. Position: Graph foundation models are already here. In *International Conference on Machine Learning*.
- McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1): 415–444.
- Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P. F.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*.
- Pošik, P.; Huyer, W.; and Pál, L. 2012. A comparison of global search algorithms for continuous black box optimization. *Evolutionary Computation*, 20(4): 509–541.
- Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2024. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*.
- Romera-Paredes, B.; Barekatin, M.; Novikov, A.; Balog, M.; Kumar, M. P.; Dupont, E.; Ruiz, F. J.; Ellenberg, J. S.; Wang, P.; Fawzi, O.; et al. 2024. Mathematical discoveries from program search with large language models. *Nature*, 625(7995): 468–475.
- Shi, M.; Tang, Y.; Zhu, X.; Huang, Y.; Wilson, D.; Zhuang, Y.; and Liu, J. 2022. Genetic-GNN: Evolutionary architecture search for graph neural networks. *Knowledge-based Systems*, 247: 108752.
- Shojaee, P.; Meidani, K.; Gupta, S.; Farimani, A. B.; and Reddy, C. K. 2024. LLM-SR: Scientific equation discovery via programming with large language models. *arXiv preprint arXiv:2404.18400*.
- Von Luxburg, U. 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17: 395–416.

Wang, X.; and Zhang, M. 2022. How powerful are spectral graph neural networks. In *International conference on Machine Learning*.

Wei, L.; Zhao, H.; and He, Z. 2022. Designing the topology of graph neural networks: A novel feature fusion perspective. In *Proceedings of the ACM Web Conference*.

Wu, K.; Li, X.; Liu, P.; and Liu, J. 2023. DECN: Evolution inspired deep convolution network for black-box optimization. *arXiv preprint arXiv:2304.09599*.

Zheng, S.; Zhu, Z.; Liu, Z.; Li, Y.; and Zhao, Y. 2024. Node-oriented Spectral Filtering for Graph Neural Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 46(1): 388–402.

Zhu, M.; Wang, X.; Shi, C.; Ji, H.; and Cui, P. 2021. Interpreting and unifying graph neural networks with an optimization framework. In *Proceedings of the Web Conference*.