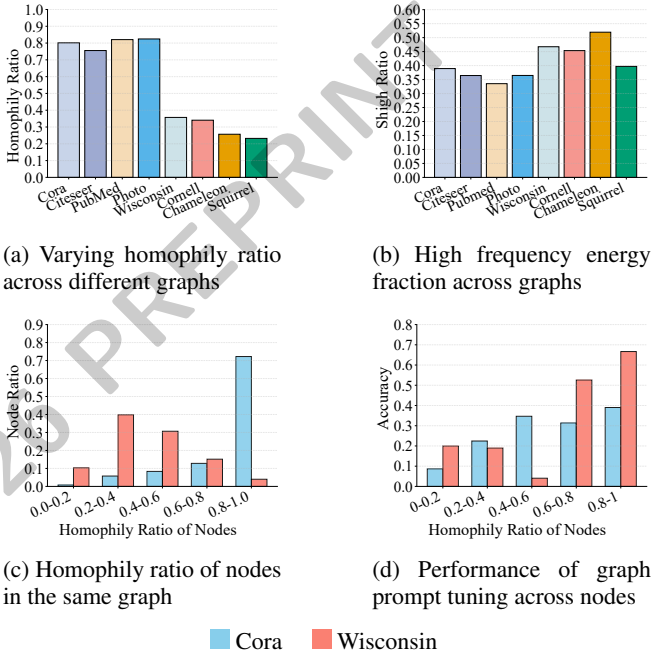
# Beyond Homophily: Spectrum-Based Graph Pre-Training and Cluster-Augmented Prompt Tuning

**Tingting Li**, **Yonghao Li**, **Xiangkun Wang**, **Sixiang Chen**, **Boyang Fan**, **Lingfei Ren***, **Hao Yu** and **Xin Yang***
Southwestern University of Finance and Economics
{224081200017, 225081200015}@smail.swufe.edu.cn, {liyonghao, renlf ,yangxin}@swufe.edu.cn, xiangkunwang18@gmail.com, fanby2014@126.com, yuhao2033@163.com

## Abstract

Graph pre-training and prompt tuning provide an effective route to label-efficient node classification by learning transferable backbones and adapting them with lightweight prompts. However, existing pre-train-and-prompt pipelines often generalize poorly across graphs with diverse homophily due to two fundamental limitations: (i) *spectral bias*, where learned representations overemphasize a single frequency band, and (ii) *prompt misalignment*, where spatial prompts cannot explicitly modulate critical spectral components and may even disrupt frequency-aware backbones. In this paper, we propose **SCA-GPPT**, a spectrum-aware framework that unifies **S**pectrum-based **G**raph **P**re-training and **C**luster-**A**ugmented **P**rompt **T**uning. First, we present a systematic spectral analysis of graph prompting, revealing the inherent limitations of spatial-domain prompts in capturing attenuated high-frequency signals. Second, we explicitly learn complementary low- and high-frequency spectral filters via parameterized Chebyshev polynomials and employ band-aware contrastive learning to model diverse spectral patterns during pre-training. For downstream adaptation, we freeze the pre-trained backbone and introduce cluster-augmented prompts that directly adapt Chebyshev coefficients while enforcing global structural consistency, enabling stable few-shot transfer across varying homophily regimes. Extensive experiments on real-world benchmarks demonstrate that SCA-GPPT consistently outperforms strong baselines under both transductive and inductive settings. Our code is available at https://github.com/swufeLTT/SCA-GPPT.




(a) Varying homophily ratio across different graphs

(b) High frequency energy fraction across graphs

(c) Homophily ratio of nodes in the same graph

(d) Performance of graph prompt tuning across nodes

Figure 1: Variation of homophily across different graphs.

## 1 Introduction

Graph learning has become a powerful paradigm for various machine learning tasks, especially node classification [Zang *et al.*, 2025]. Graph neural networks (GNNs) [Jiang *et al.*, 2019; Velickovic *et al.*, 2017; Ju *et al.*, 2025] are central to this paradigm due to their ability to aggregate information from local neighborhoods. However, with limited labeled data, GNNs are prone to overfitting and often generalize poorly, motivating graph pre-training methods that learn transferable representations from large-scale unlabeled graphs via self-supervised objectives, such as DGI [Veličković *et al.*, 2019] and GraphCL [You *et al.*, 2020]. Building on pre-trained models, graph prompt tuning has emerged as an efficient adaptation strategy that introduces lightweight, task-specific prompts while keeping the backbone fixed [Sun *et al.*, 2022; Liu *et al.*, 2023]. Recent unified frameworks, including HGPrompt [Yu *et al.*, 2024a], SAMGPT [Yu *et al.*, 2025a], and HS-GPPT [Luo *et al.*, 2025], further combine pre-training with prompt tuning to enhance adaptability across diverse graph structures.

Despite these advances, existing graph pre-training and prompt tuning methods face two critical challenges. **(1) Graph spectral bias**. Graph frequency distributions are

*Corresponding authors: Lingfei Ren, Xin Yang.

closely correlated with homophily: homophilic graphs are dominated by low-frequency components, whereas heterophilic graphs contain richer high-frequency information [Ren *et al.*, 2024; Zheng *et al.*, 2022]. As shown in Fig. 1a and Fig. 1b, datasets with lower homophily exhibit higher proportions of high-frequency energy. However, most pretraining methods focus on a single spectral band, limiting generalization across graphs with varying homophily levels [Chen *et al.*, 2025]. Moreover, without supervision, spectral GNNs struggle to learn optimal filter shapes, leading to suboptimal downstream adaptation. **(2) Graph prompt tuning misalignment**. Existing graph prompt methods are predominantly designed in the spatial domain and do not explicitly model or control the spectral properties of node representations [Yu *et al.*, 2025b]. Since heterophily varies across graphs and even across nodes within the same graph (Fig. 1c), standard prompt methods exhibit node- and dataset-sensitive performance (Fig. 1d). When combined with frequency-aware pre-trained models, spatial prompts often fail to modulate critical spectral bands and may even disrupt learned frequency information, particularly on heterophilic graphs. Spectrally isolated bias or prompt tuning is inadequate, as current methods lack joint optimization of spectral learning and adaptation, limiting homophily-agnostic generalization.

To address these challenges, we propose *Spectrum-based Graph Pre-training* and *Cluster Augmented Prompt Tuning (SCA-GPPT)*. During pre-training, we first employ a *smoothness-aware feature alignment module* to map heterogeneous node features from different domains into a unified semantic space. Complementary low and high-frequency filters are then learned via *parameterized Chebyshev polynomials*, together with band-aware contrastive learning, to capture diverse spectral patterns in the pretraining stage. For downstream adaptation, we freeze the pre-trained backbone and introduce a *spectral prompting mechanism* that directly adjusts Chebyshev filter coefficients, enabling fine-grained frequency adaptation without modifying backbone parameters. In addition, *cluster-augmented prompts* further exploit global structural consistency, yielding representations that are both discriminative and globally coherent. By jointly modeling feature alignment, spectral diversity learning, and prompt alignment, SCA-GPPT achieves robust generalization across graphs with varying homophily levels.

In summary, our contributions are threefold: (i) We pioneer a systematic spectral analysis of graph prompt tuning, revealing spectral bias and spatial prompt misalignment that necessitate spectral pre-training and prompt tuning across varying homophily. (ii) We propose SCA-GPPT, a spectrum-aware framework integrating spectral pre-training and cluster-augmented prompt tuning. It achieves fine-grained spectral alignment through learnable parameterized Chebyshev polynomials and adaptive clustering prompts, improving cross-graph transfer in both transductive and inductive scenarios. (iii) Extensive experiments on synthetic and real-world graphs with varying homophily demonstrate that SCA-GPPT consistently outperforms existing baselines in both transductive and inductive settings.



# 2 Preliminaries and Theoretical Analysis

## 2.1 Preliminaries

**Homophily.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ be a graph equipped with a node label vector $y_i \in \mathcal{Y}$. The homophily level quantifies the extent to which adjacent nodes share the same label. It is defined as the proportion of edges with the same label:

$$h(\mathcal{G}, y) = \frac{1}{|\mathcal{E}|} \sum_{(u,v)\in\mathcal{E}} \mathbb{I}\{y_u = y_v\}, \tag{1}$$

where $|\mathcal{E}|$ denotes the total number of edges and $\mathbb{I}\{\cdot\}$ is the indicator function that returns 1 when its argument is true and 0 otherwise.

**Spectral Graph Filters.** A spectral graph filter is a real-valued function $g(\cdot)$ applied to the spectrum of $\mathbf{L}$ [Bruna *et al.*, 2014]. Given a node feature matrix $\mathbf{L} \in \mathbb{R}^{N\times d}$, the filtered representation $\mathbf{Z} \in \mathbb{R}^{N\times d}$ is obtained by

$$\mathbf{Z} = g(\mathbf{L})\mathbf{X} = \mathbf{U}\, g(\Lambda)\, \mathbf{U}^\top \mathbf{X}, \tag{2}$$

where $g(\Lambda) = \mathrm{diag}(g(\lambda_1), \ldots, g(\lambda_N))$. Different choices of $g$ emphasise different spectral bands and thus capture different patterns of neighbour aggregation: low-pass filters promote similarity among connected nodes [Kipf and Welling, 2017], while high-pass filters highlight discrepancies between nodes and their neighbors [Bo *et al.*, 2021].

**Spectral Energy.** Let $\mathcal{G} = (\mathcal{V}, \mathbf{X}, \mathcal{E})$ be an undirected graph with $|\mathcal{V}| = n$ and Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$. For a node-wise scalar signal $\mathbf{x} \in \mathbb{R}^n$, we define its *spectral energy* as:

$$\mathcal{F}_G(\mathbf{x}) \;:=\; \frac{1}{2}\mathbf{x}^\top \mathbf{L}\mathbf{x}, \tag{3}$$

when using the normalized Laplacian $\mathbf{L}_{\mathrm{sym}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, one may analogously define $\mathcal{F}_G^{(\mathrm{norm})}(\mathbf{x}) := \frac{1}{2}\mathbf{x}^\top \mathbf{L}_{\mathrm{sym}}\mathbf{x}$.

**High-Frequency Energy.** Recall that the eigenvalues of the normalized Laplacian satisfy $\lambda_i \in [0, 2]$, Then we define the high-frequency energy fraction as

$$S_{\mathrm{high}}(\mathcal{G},\mathbf{x}) := \frac{\sum_{i\in\mathcal{B}_h} \lambda_i \hat{x}_i^2}{\sum_{i=1}^N \lambda_i \hat{x}_i^2}, \qquad S_{\mathrm{low}}(\mathcal{G},\mathbf{x}) := 1 - S_{\mathrm{high}}(\mathcal{G},\mathbf{x}), \tag{4}$$

where $\mathcal{B}_h := \{i | \lambda_i \geq \lambda_0\}$ is high spectral bands. For node classification, we instantiate $\mathbf{x}$ as the label indicator signals $\{\mathbf{y}_c\}_{c=1}^C$:

$$S_{\mathrm{high}}(\mathcal{G}, \mathbf{Y}) := \frac{\sum_{c=1}^C \sum_{i\in\mathcal{B}_h} \lambda_i \widehat{y}_{c,i}^2}{\sum_{c=1}^C \sum_{i=1}^N \lambda_i \widehat{y}_{c,i}^2}. \tag{5}$$

## 2.2 Theoretical Analysis

**Lemma 1** (Spectral Limitation of Spatial Prompts)**.** *Considering a general prompt as a perturbation $\delta$ to the feature matrix: $\mathbf{X}_p = \mathbf{X} + \delta$, where $\delta$ is learnable and lightweight. In the spectral domain, the prompted output becomes: $\mathbf{Z}_p = \boldsymbol{U} g(\Lambda) \boldsymbol{U}(\mathbf{X} + \delta) = g(\mathbf{L})\mathbf{X} + g(\mathbf{L})\delta$, the prompted high-frequency energy satisfies:*

$$\|\mathbf{Z}_{p,high}\|^2 \leq \varepsilon \|\delta\|^2, \tag{6}$$

*where $\varepsilon \to 0$ as the high-pass attenuation of $g$ increases. This implies that spatial prompts cannot effectively recover high-frequency information if it is absent from the pre-trained representations.*

Lemma 1 highlights that traditional space-domain prompt fine-tuning loses high-frequency energy in graphs and is only suitable for homophilic graphs. Figure 1d also confirms that nodes with lower homophily exhibit poorer fine-tuning performance.

**Theorem 1** (Transfer Ability to Spectral Discrepancy)**.** *Assume the optimal downstream decision function depends on filtered representations $g^\star(\mathbf{L})\mathbf{X}$ and the learned filter is $g$. For Lipschitz-continuous filters $g$ on $[0, 2]$, the representation mismatch between source and target satisfies*

$$\|g(\mathbf{L}_s)\mathbf{X}_s - g(\mathbf{L}_t)\mathbf{X}_t\|_F \;\le\; C_1\|\mathbf{X}_s - \mathbf{X}_t\|_F + 2C_2\left|S^{(s)}_{\text{high}} - S^{(t)}_{\text{high}}\right|, \quad (7)$$

*for some constants $C_1, C_2$ depending on $\|g\|_\infty$ and the Lipschitz constant of $g$.*

Theorem 1 highlights that cross-graph transfer ability is fundamentally limited by spectral discrepancy. Therefore, a pre-training objective that only emphasizes pure low-pass is brittle when $\left|S^{(s)}_{\text{high}} - S^{(t)}_{\text{high}}\right|$ is large.

# 3 Methodology

Preliminary analysis confirms that spatial prompt tuning causes loss of high-frequency energy and is only applicable to homophilic graphs. To address this issue from both spectral representation learning and prompt-based adaptation perspectives, we propose SCA-GPPT, as shown in Figure 2.

## 3.1 Smoothness-based Feature Alignment (SFA)

Node features often vary across domains due to diverse textual descriptions, metadata, or domain-specific attributes (e.g., personal profiles in pre-training versus transaction attributes in fine-tuning). To integrate such heterogeneous features, we propose an alignment module that maps them into a unified representation space through two stages: feature projection and semantic alignment.

**Feature Projection.** Motivated by GCOPE [Zhao *et al.*, 2024], we align the dimensions of features from different domains via a shared linear transformation. Specifically, for the feature matrix $\mathbf{X}^{(i)} \in \mathbb{R}^{n_i \times d_i}$ from domain $i$, we apply the projection operation:

$$\tilde{\mathbf{X}}^{(i)} \in \mathbb{R}^{N_i \times d_u} = \mathrm{Proj}\left(\mathbf{X}^{(i)}\right), \quad (8)$$

where $d_u$ denotes the predefined projection dimension, while $\mathrm{Proj}(\cdot)$ represents the feature projection operation, typically implemented as singular value decomposition (SVD) [Golub and Reinsch, 1971] or principal component analysis (PCA) [Pearson, 1901].

**Semantic Alignment.** Despite achieving dimensional alignment, semantic inconsistencies may still arise between the pre-training and fine-tuning stages [Liu *et al.*, 2024]. To further optimize this issue, we introduce a smoothness-based semantic alignment that draws from the perspective of graph signal processing to evaluate the contribution of each feature dimension. Formally, given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, we define the smoothness for the $k$-th feature dimension as:

$$s_k(\mathbf{X}) = \frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}}(\mathbf{X}_{u,k} - \mathbf{X}_{v,k})^2, \quad (9)$$

where a lower $s_k$ value indicates substantial variation between connected nodes, reflecting strong heterophily similar to high-frequency components.

## 3.2 Spectrum-based Pre-training

Theorem 1 demonstrates that a graph's transfer ability correlates with spectral energy differences across graphs. To achieve generalization capabilities across diverse spectral distributions, spectral information must be extracted. To this end, we designed a spectrum-based graph pre-training mechanism that ensures balanced integration of both low-frequency and high-frequency components in graphs, thereby enhancing the model's ability to discern subtle differences in spectral patterns.

**Adaptive Chebyshev Encoding (ACE).** Traditional heterophilic graph learning incurs parameters that scale linearly with the polynomial degree $K$ and may overemphasize specific frequency bands [He *et al.*, 2022]. Recent studies [Chen *et al.*, 2024; Wan *et al.*, 2024] build on cosine-parameterized Chebyshev polynomials to propose PolyGCL, a graph contrastive learning paradigm that effectively handles both homophilic and heterophilic graphs without relying on complex data augmentation used in conventional GCL [You *et al.*, 2020]. Consequently, we adopt PolyGCL as our backbone model. Specifically, given Chebyshev polynomials with interpolation $\sum_{k=0}^{K} w_k \mathrm{T}_k\left(\hat{\mathbf{L}}\right)\mathbf{X}$, where $\hat{\mathbf{L}}$ is normalized Laplace matrix, and the enhanced Chebyshev polynomials with Chebyshev interpolation $w_k$ is:

$$w_k = \frac{2}{K+1}\sum_{j=0}^{K}\gamma_j \mathrm{T}_k(x_j), \quad (10)$$

where $x_j = \cos\left(\frac{j+1/2}{K+1}\pi\right), j = 0, \cdots, K$ denoted the Chebyshev node for $\mathrm{T}_{K+1}$, and $\mathrm{T}_k(x)$ is Chebyshev polynomials recursively defined as $\mathrm{T}_k(x) = 2x\mathrm{T}_{k-1}(x) - \mathrm{T}_{k-2}(x)$, with $\mathrm{T}_0(x) = 1$ and $\mathrm{T}_1(x) = x$. As for the learnable parameter $\gamma_j$, we propose to use a learnable sliding cosine-parametric-based formulation to build our low-pass and high-pass filters based on parameters $\gamma^l_j$ and $\gamma^h_j$ as follows:

$$\begin{aligned}\gamma^l_j =& \sigma(\beta^l_a) - \frac{1}{2}\sigma(\beta^l_b)\\ &\times\left(1+\cos\left(1+\frac{\frac{1}{2}\tanh(\delta_l)+j}{K}\pi\right)\right),\\ \gamma^h_j =& \sigma(\beta^h_a) - \frac{1}{2}\sigma(\beta^h_b)\\ &\times\left(1+\cos\left(1+\frac{\frac{1}{2}\tanh(\delta_h)+j}{K}\pi\right)\right).\end{aligned} \quad (11)$$

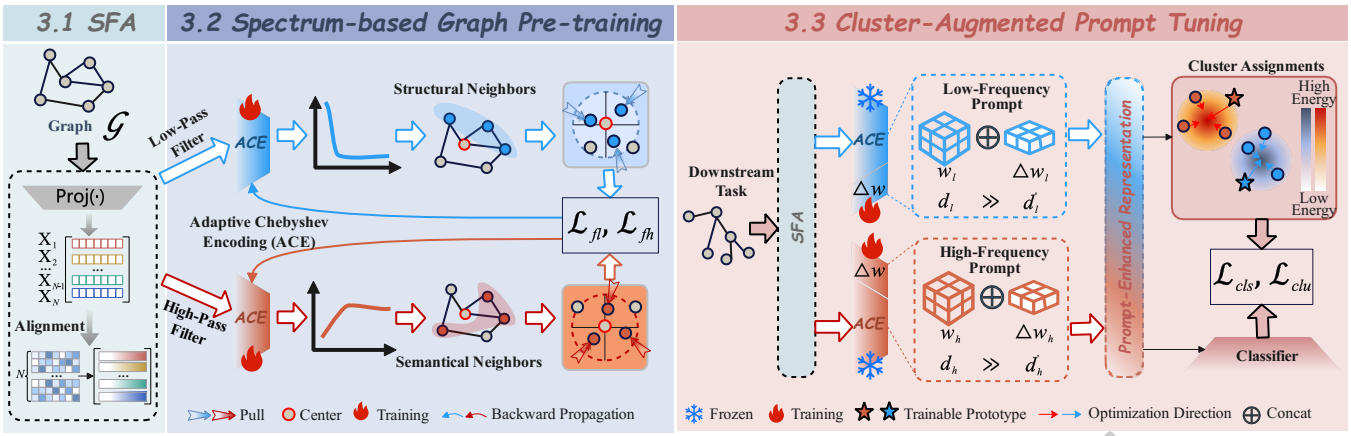



Figure 2: SCA-GPPT Framework: (1) Smoothness-based Feature Alignment (SFA) aligns heterogeneous node features across domains; (2) Adaptive Chebyshev Encoding (ACE) pre-training learns robust representations via joint low-/high-frequency modeling; (3) Cluster-augmented prompt tuning enables efficient cross-graph adaptation with a frozen backbone.

For initialisation, we set $\beta_j^h$ and $\beta_j^l$ to 0 and 2, respectively, whereas $\beta_b^l$ and $\beta_b^h$ are set to 2. These four parameters are all trainable during the learning process. To sum up, substituting $\gamma_i$ with $\gamma^l$ and $\gamma^h$ in Eq.10, respectively, the graph filter can be expressed as:

$$\begin{aligned}\mathbf{Z}_l &= \sum_{k=0}^{K} w_k^l \mathrm{T}_k\left(\hat{\mathbf{L}}\right) f_\theta^l(\mathbf{X}),\\ \mathbf{Z}_h &= \sum_{k=0}^{K} w_k^h \mathrm{T}_k\left(\hat{\mathbf{L}}\right) f_\theta^h(\mathbf{X}),\end{aligned} \tag{12}$$

where $f_\theta(\mathbf{X})$ denotes the use of MLP on the raw feature matrix $\mathbf{X}$.

**Local Homophily.** Specifically, we first utilise the smoothing properties of low-pass filters to pre-train them using traditional link prediction methods. For node $v_i$ in the graph, its positive set $\mathcal{N}_i'$ consists of nodes $v_j$ that are first-order neighbours of $v_i$, while its negative set $\mathcal{V} \setminus v_i$ comprises randomly sampled nodes that are not first-order neighbours. The cross-low objective is formulated to maximise the similarity between representations of such pairs:

$$\mathcal{L}_{fl} = -\frac{1}{2|\mathcal{V}|}\sum_{v_i\in\mathcal{V}}\frac{1}{|\mathcal{N}_i'|}\sum_{v_p\in\mathcal{N}_i'} \log\frac{s(z_i^l, z_p^l)}{\sum_{v_q\in\mathcal{V}\setminus v_i} s(z_i^l, z_q^l)}. \tag{13}$$

where $z_i^l$ and $z_p^l$ emphasise low-frequency components, capturing the local similarity and homophily within the spatial domain, $s(\cdot,\cdot)$ denotes a similarity function (e.g., cosine similarity). This objective encourages the model to align the embeddings of structurally proximate nodes, thereby fostering robustness in both homophilic and heterophilic graphs.

**Global Heterophily.** The high-pass filter reveals the diversity within local neighbourhoods and the global long-range similarity among semantic neighbours. We introduce semantically neighbours as positive sample pairs, determined based on feature-level similarity. Specifically, for each node $v_i$, we identify its $k$ most similar neighbours based on the aligned features, denoted as $\mathcal{N}_i'' = kNN(v_i, k)$. This enables the derivation of the all-pass component of the objective function, as shown in Eq.14:

$$\mathcal{L}_{fh} = -\frac{1}{2|\mathcal{V}|}\sum_{v_i\in\mathcal{V}}\frac{1}{|\mathcal{N}_i''|}\sum_{v_p\in\mathcal{N}_i''} \log\frac{s(z_i^h, z_p^h)}{\sum_{v_q\in\mathcal{V}\setminus v_i} s(z_i^h, z_q^h)}, \tag{14}$$

where $z_i^h$ and $z_p^h$ emphasizes high-frequency components, capturing long-range dependencies and dissimilarities within the spatial domain.

**Optimization and Balancing.** The overall loss combines these objectives with a balancing coefficient:

$$\mathcal{L} = \alpha\mathcal{L}_{fl} + (1-\alpha)\mathcal{L}_{fh}, \tag{15}$$

where $\alpha$ is a tunable hyperparameter that modulates the emphasis between local and global contributions. This formulation promotes inference efficiency, making it particularly suitable for large-scale graphs and real-time applications. Extensive experiments demonstrate that this approach yields superior adaptability and performance in diverse graph learning scenarios.

### 3.3 Cluster-Augmented Prompt Tuning

In the downstream stage, we use a parameter-efficient prompt tuning with adapters on the Chebyshev order $K$. The backbone, including $K$, remains frozen to preserve pre-trained knowledge. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, the frozen backbone produces spectral representations $\mathbf{Z}_l$ and $\mathbf{Z}_h$ capturing homophilic and heterophilic patterns.

To adapt the model without updating frozen parameters, we design lightweight prompts as adapters to $K$. These adapters modulate the Chebyshev filter weights $\{w_k\}_{k=0}^K$ associated with the frozen $K$. The prompt function $\mathcal{P}_K$ produces adaptive adjustments:

$$\{\Delta w_k\} = \mathcal{P}_K\left(\{w_k\}; \Theta_{\text{prompt}}\right), \tag{16}$$

where $\Theta_{\text{prompt}}$ are trainable parameters. The prompt-enhanced representation is:

$$\hat{\mathbf{Z}} = \sum_{k=0}^{K} (w_k + \Delta w_k)\, \mathrm{T}_k(\hat{\mathbf{L}})\, f_\theta(\mathbf{X}). \tag{17}$$

The adapter $\mathcal{P}_K$ is implemented as a bottleneck structure:

$$\mathcal{P}_K\left(\{w_k\}\right) = \mathbf{W}_{\text{up}}\, \sigma(\mathbf{W}_{\text{down}}\, \phi(\{w_k\}))\,, \tag{18}$$

where $\phi(\cdot)$ encodes weight information, $\mathbf{W}_{\text{down}} \in \mathbb{R}^{d \times d'}$, $\mathbf{W}_{\text{up}} \in \mathbb{R}^{d' \times d}$ with $d' \ll d$, and $\sigma(\cdot)$ is an activation function. For training stability, $\mathbf{W}_{\text{up}}$ is initialized to zero, ensuring minimal initial perturbation to the pre-trained filters.

After prompt modulation, the low-frequency and high-frequency representations are fused to form the final node embedding:

$$\mathbf{Z} = \beta \hat{\mathbf{Z}}_l + (1 - \beta)\hat{\mathbf{Z}}_h, \tag{19}$$

where $\beta \in [0, 1]$ controls the relative contribution of homophilic and heterophilic information. This design enables flexible adaptation to different graph structural regimes without modifying the backbone architecture.

While supervised classification loss provides signals for labeled nodes, it may not sufficiently regularize the global embedding space, especially in low-label or semi-supervised settings, so we introduce a cluster-augmented objective that enforces structural consistency among prompt-enhanced representations. Let $\mathbf{Z} = \{z_i\}_{i=1}^{|\mathcal{V}|}$ denote the final node embeddings, and $\{\mu_k\}_{k=1}^{K}$ be a set of learnable cluster centroids. We compute the soft assignment of node $v_i$ to cluster $k$ using a Student-*t* kernel:

$$q_{ik} = \frac{\left(1 + \frac{\|z_i - \mu_k\|^2}{\tau}\right)^{-1}}{\sum_{k'} \left(1 + \frac{\|z_i - \mu_{k'}\|^2}{\tau}\right)^{-1}}, \tag{20}$$

where $\tau$ is a temperature parameter controlling the smoothness of the assignment distribution.

Following the Deep Embedded Clustering (DEC) paradigm [Xie *et al.*, 2016], we construct a sharpened target distribution to enhance cluster discrimination:

$$p_{ik} = \frac{q_{ik}^2 / \sum_i q_{ik}}{\sum_{k'} q_{ik'}^2 / \sum_i q_{ik'}}, \tag{21}$$

and optimize the clustering objective by minimizing the Kullback–Leibler divergence between the target distribution and the soft assignments:

$$\mathcal{L}_{\text{clu}} = \sum_{i=1}^{|\mathcal{V}|} \sum_{k=1}^{K} p_{ik} \log \frac{p_{ik}}{q_{ik}}. \tag{22}$$

It is worth noting that the clustering objective does not introduce additional supervision, but instead serves as a structural regularizer that complements the classification loss [Caron *et al.*, 2018; Bo *et al.*, 2020]. While the classification objective enforces label-level separability, the clustering constraint encourages compactness and consistency in the embedding space, thereby improving training stability and generalization.

The overall downstream optimization objective is defined as:

$$\mathcal{L}_{\text{down}} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{clu}}, \tag{23}$$

where $\mathcal{L}_{\text{cls}}$ denotes the supervised classification loss and $\lambda$ is a balancing coefficient. During downstream training, only the prompt modules, clustering module, and classification head are updated, while the pre-trained spectral backbone remains frozen. This design enables efficient and stable adaptation with minimal trainable parameters.

**Theorem 2** (Stability Gain from Clustering Regularization)**.** *Assume the backbone is frozen and only prompt parameters are learned. If the clustering loss enforces a margin $\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k'}\| \geq m$ for $k \neq k'$, then the prompt solution has improved uniform stability, leading to a tighter generalization bound:*

$$\begin{aligned} \mathcal{R}(\hat{f}) - \hat{\mathcal{R}}(\hat{f}) \;\leq\; & O\left(\frac{\|\Theta_{\text{prompt}}\|}{\sqrt{|\mathcal{S}|}}\right) \\ & + O\left(\frac{1}{m\sqrt{|\mathcal{S}|}}\right), \end{aligned} \tag{24}$$

*where $\mathcal{S}$ is the labeled set and $\Theta_{\text{prompt}}$ are prompt parameters.*

Theorem 2 formalizes that clustering regularization reduces the effective hypothesis space explored by prompt tuning under few-shot supervision, thus improving generalization and robustness.

# 4 Experiments

## 4.1 Experimental Setup

**Datasets.** We conduct experiments on eight real-world datasets with various homophily levels. Among them, Cora, Citeseer, PubMed [Sen *et al.*, 2008], Amazon product network Photo [Shchur *et al.*, 2018] are homophilic graphs, Wisconsin, Cornell [Pei *et al.*, 2020], Chameleon, Squirrel [Rozemberczki *et al.*, 2021]are considered as heterophilic graphs.

**Baselines.** The baseline methods fall into three categories: (i) Traditional GNN models: GCN [Kipf and Welling, 2017], GAT [Velickovic *et al.*, 2017], H2GCN [Zhu *et al.*, 2020], and FAGCN [Bo *et al.*, 2021]; (ii) Graph "Pre-training + Fine-tuning" models: DGI [Veličković *et al.*, 2019], GraphCL [You *et al.*, 2020], and PolyGCL [Chen *et al.*, 2024]; (iii) Graph "Pre-training + Prompt-tuning" models: GPPT [Sun *et al.*, 2022], GPrompt [Liu *et al.*, 2023], GPF-plus [Yu *et al.*, 2024b], All-in-One [Zhao *et al.*, 2024], ProNoG [Yu *et al.*, 2025b], and HS-GPPT [Luo *et al.*, 2025]. Among these methods, H2GCN, FAGCN, and PolyGCL are designed to handle heterophily. For GPPT, GPrompt, GPF-plus, and All-in-One, we use GraphSAGE [Hamilton *et al.*, 2017] as the backbone. For HS-GPPT, we follow the official implementation released by the authors and keep its original backbone and training settings.

## 4.2 Performance Comparison

We focus on two scenarios: transductive and inductive settings. We define our task as 1-shot node classification. Results are averaged over 100 runs with different random seeds

| Method | Cora | Citeseer | PubMed | Photo | Wisconsin | Cornell | Chameleon | Squirrel |
|---|---|---|---|---|---|---|---|---|
| GCN | 37.77 ± 8.13 | 27.24 ± 6.21 | 45.68 ± 8.34 | 48.84 ± 11.48 | 28.61 ± 10.42 | 25.78 ± 7.68 | 24.69 ± 4.29 | 20.70 ± 1.33 |
| GAT | 47.33 ± 6.80 | 40.76 ± 6.42 | 53.01 ± 8.83 | 67.51 ± 8.73 | 29.50 ± 6.42 | 27.03 ± 5.77 | 24.84 ± 4.17 | 20.74 ± 1.70 |
| H2GCN | 35.49 ± 7.22 | 33.09 ± 6.92 | 46.09 ± 8.02 | 46.76 ± 11.04 | 31.17 ± 10.65 | 27.43 ± 9.02 | 22.25 ± 3.67 | 20.37 ± 1.14 |
| FAGCN | 33.26 ± 10.06 | 26.78 ± 7.16 | 42.43 ± 9.27 | 19.22 ± 11.36 | 31.30 ± 12.64 | 25.04 ± 9.56 | 23.31 ± 3.69 | 21.09 ± 2.07 |
| DGI | 32.90 ± 5.91 | 25.43 ± 4.79 | 39.95 ± 6.31 | 43.47 ± 7.61 | 27.07 ± 9.51 | 24.01 ± 6.69 | 24.23 ± 4.01 | 20.90 ± 1.71 |
| GraphCL | 29.86 ± 5.83 | 28.82 ± 5.63 | 40.35 ± 6.78 | 42.23 ± 8.46 | 23.72 ± 8.12 | 22.57 ± 7.13 | 20.97 ± 2.81 | 20.27 ± 1.36 |
| PolyGCL | 41.18 ± 8.37 | 34.00 ± 6.91 | 52.55 ± 14.07 | 48.25 ± 8.09 | 34.96 ± 11.36 | 24.76 ± 11.41 | 25.63 ± 5.01 | 22.83 ± 3.15 |
| GPPT | 31.44 ± 7.48 | 32.88 ± 8.19 | 40.35 ± 8.38 | 55.33 ± 8.80 | 25.33 ± 6.32 | 24.23 ± 6.52 | 23.89 ± 4.41 | 20.76 ± 1.64 |
| GPrompt | 39.91 ± 8.46 | 34.08 ± 7.55 | 43.20 ± 9.97 | 39.06 ± 8.17 | 24.84 ± 11.07 | 23.46 ± 8.99 | 21.78 ± 3.56 | 20.22 ± 1.25 |
| GPF-plus | 40.57 ± 8.30 | 35.14 ± 7.49 | 43.34 ± 9.32 | 39.26 ± 8.18 | 25.24 ± 10.69 | 24.11 ± 8.40 | 21.93 ± 3.72 | 20.36 ± 1.27 |
| All-in-One | 25.35 ± 4.67 | 23.67 ± 5.94 | 49.85 ± 9.18 | 11.68 ± 4.18 | 41.16 ± 12.62 | 24.37 ± 7.12 | 24.48 ± 3.78 | 21.22 ± 1.91 |
| ProNoG | 47.80 ± 8.01 | 35.89 ± 6.59 | 50.07 ± 8.44 | 59.88 ± 7.37 | 30.24 ± 9.26 | 26.71 ± 7.39 | 26.14 ± 4.42 | 21.39 ± 1.93 |
| HS-GPPT | 45.69 ± 9.10 | 45.97 ± 10.45 | **55.47** ± 10.50 | 63.78 ± 9.82 | 32.67 ± 9.07 | 33.02 ± 7.79 | 23.71 ± 4.03 | 20.54 ± 2.11 |
| Ours | **61.63**±9.37 | **53.61**±11.38 | 55.20± 11.81 | **68.22**±8.33 | **46.87**±12.55 | **41.61**±10.66 | **31.30**±5.50 | **23.95**±4.40 |

Table 1: Performance comparison on 1-shot node classification under the transductive setting. Here, bold values signify the best result across all methods, while underlined highlights the second-best results.

| Settings | In-Domain | | Cross-Domain | |
|---|---|---|---|---|
| Source<br>Target | Chameleon<br>Squirrel | Wisconsin<br>Cornell | citeseer<br>Squirrel | Chameleon<br>Citeseer |
| DGI | 21.15 ± 1.81 | 23.56 ± 6.25 | 20.53 ± 1.46 | 23.60 ± 4.52 |
| GraphCL | 20.47 ± 1.30 | 22.29 ± 7.45 | 20.48 ± 1.33 | 28.15 ± 4.48 |
| PolyGCL | 22.62 ± 2.67 | 27.88 ± 11.35 | 21.31 ± 2.29 | 26.99 ± 7.23 |
| GPPT | 20.67 ± 1.71 | 22.99 ± 5.63 | 20.69 ± 1.84 | 19.87 ± 3.19 |
| GPrompt | 21.24 ± 2.17 | 26.29 ± 9.57 | 20.73 ± 1.81 | 20.35 ± 4.72 |
| GPF-plus | 21.33 ± 2.18 | 25.99 ± 8.88 | 20.59 ± 1.63 | 22.13 ± 4.89 |
| All-in-One | 21.60 ± 2.44 | 22.27 ± 6.43 | 22.02 ± 3.16 | 18.08 ± 2.80 |
| ProNoG | 21.38 ± 1.98 | 26.07 ± 7.46 | 20.96 ± 1.81 | 20.95 ± 1.85 |
| HS-GPPT | 20.38 ± 1.68 | 26.36 ± 6.91 | 20.48 ±1.33 | 23.32 ± 3.88 |
| Ours | **24.16** ± 4.66 | **30.04** ± 11.06 | **24.55** ± 4.44 | **50.40** ± 10.88 |

Table 2: Performance comparison on 1-shot node classification under the inductive setting, "Source" denotes the pre-training dataset, and "Target" denotes the downstream dataset.

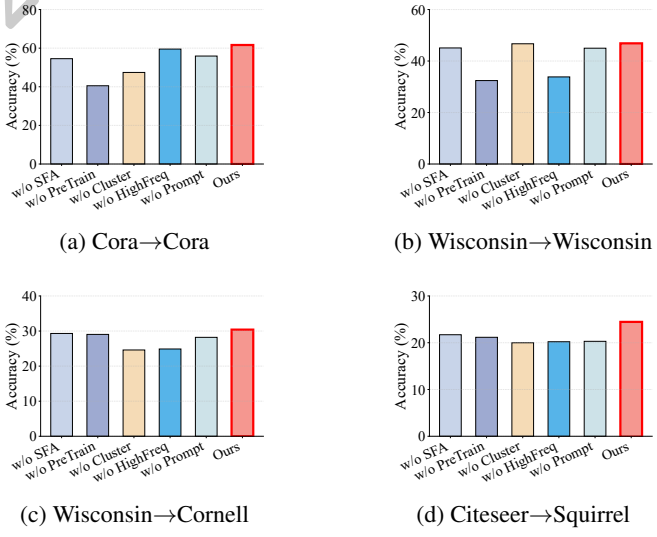
, using ACC as the metric Results of transductive learning are in Table 1, inductive learning in Table 2.

**Transductive Performance.** From the results, we observe that SCA-GPPT performs best on most datasets and remains highly competitive on PubMed across both homophilic and heterophilic graphs. On homophilic datasets such as Cora and Citeseer, SCA-GPPT achieves accuracies of 61.63% and 53.61%, respectively, outperforming the second-best methods by a substantial margin. On heterophilic datasets including Cornell, Chameleon, and Squirrel, our method remains robust (41.61%, 31.30%, and 23.95%), demonstrating its ability to leverage both structural and feature cues across diverse graph types. Overall, our approach achieves strong and competitive performance across datasets, further demonstrating its effectiveness and robustness in few-shot scenarios.

**Inductive Performance.** We evaluate inductive learning performance across graphs with varying levels of homophily, considering both in-domain settings where pre-training and downstream tasks share the same domain and cross-domain settings. Experimental results demonstrate that our model achieves strong inductive learning performance, even when discrepancies between pre-trained and downstream graphs pose challenges for knowledge transfer. This effectiveness stems from its ability to capture rich spectral representations during pre-training. It then adapts the spectral distribution of downstream graphs to the pre-trained knowledge through cluster-augmented prompt tuning, facilitating more effective knowledge transfer.

**Limitation of SCA-GPPT.** Under the transductive setting, our method achieves slightly lower accuracy than the best-performing baseline (by 0.27%) on PubMed. This slight gap may be attributed to the 1-shot setting and the subtle class feature variations in PubMed, which restrict the effective use of structural and feature information.

### 4.3 Ablation Study




(a) Cora→Cora (b) Wisconsin→Wisconsin

(c) Wisconsin→Cornell (d) Citeseer→Squirrel

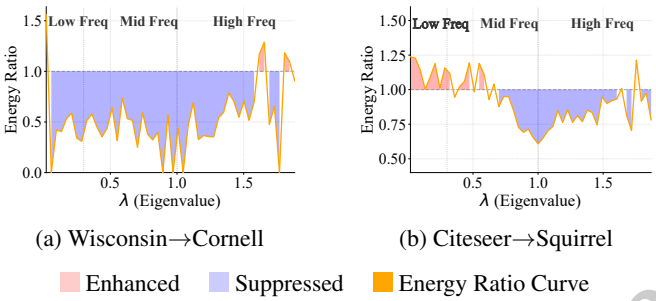
Figure 3: Ablation study on 1-shot transductive and inductive setting.(a)(b)present the ablation results under the transductive setting, while (c)(d) present the ablation results under the inductive setting in in-domain settings and cross-domain setting. The red border indicates the best-performing model in each dataset.

We conduct a systematic ablation study under the 1-shot inductive node classification setting to evaluate the contributions of pre-training, prompt tuning, and spectral modeling in Figures 3. (1) w/o SFA removes smoothness-based feature

alignment. (2) w/o PreTrain replaces the adaptive Chebyshev encoding with a vanilla GCN. (3) w/o Cluster removes the cluster assignments and relies only on the classifier. (4) w/o HighFreq removes high-frequency filter. (5) w/o Prompt disables prompt tuning entirely.

The ablation study shows that all components contribute to the final performance. The clustering head (w/o Cluster) has a more noticeable impact on homophilic graphs (Figure 3a), while removing prompt tuning (w/o Prompt) also causes a moderate decline in performance. Among them, removing spectrum-aware pre-training (w/o PreTrain) leads to the largest performance drop, particularly in the transductive setting (Figure 3c). Removing smoothness-based feature alignment (w/o SFA) or high-frequency modeling (w/o HighFreq) results in consistent performance degradation across both transductive and inductive settings.Overall, the complete model consistently outperforms all ablated variants under both settings.

### 4.4 Spectral Analysis of Prompt tuning

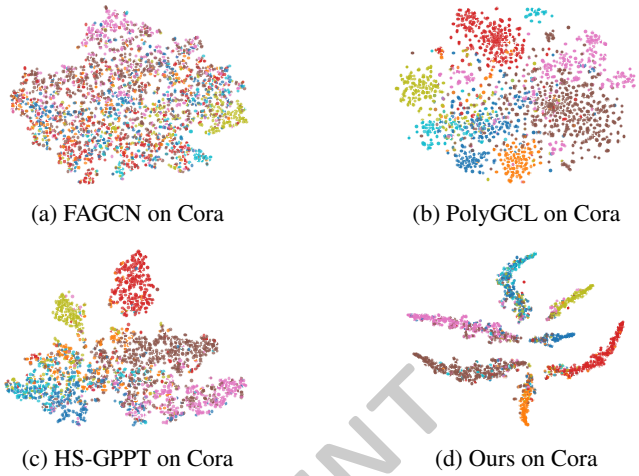



(a) Wisconsin→Cornell (b) Citeseer→Squirrel

Figure 4: Spectral analysis of prompt tuning in 1-shot inductive learning (Wisconsin→Cornell and Citeseer→Squirrel). The orange curve shows the energy ratio curves for pre- and post-prompt tuning: the blue segment indicates energy compression, while the red segment indicates energy enhancement.

As shown in Fig. 4, including the Wisconsin→Cornell (Fig. 4a) and Citeseer→Squirrel (Fig. 4b) inductive settings, the spectral analysis illustrates how the spectral prompt tuning modulates graph signals in the 1-shot transfer learning scenario. The eigenvalue spectrum $\lambda$ is partitioned into low-frequency ($\lambda < 0.3$), mid-frequency ($0.3 \leq \lambda < 1.0$), and high-frequency ($\lambda \geq 1.0$) bands for band-wise spectral energy analysis. Prompt tuning performs targeted spectral shaping by enhancing task-relevant frequencies while attenuating bands associated with noise or domain discrepancies. In the Wisconsin→Cornell transfer (Fig. 4a), spectral energy is primarily suppressed in the low- and mid-frequency bands as well as most of the high-frequency band, with limited enhancement in localized high-frequency regions. In contrast, the Citeseer→Squirrel transfer (Fig. 4b) exhibits broad high-frequency suppression and partial mid-frequency attenuation, accompanied by mild enhancement in the low-frequency region. It is evident that our proposed spetral prompt fine-tuning method can automatically adjust the spectral parameters of the pre-trained filter based on the spectral differences between the source and target graphs, thereby better adapting to the spectral distribution of the downstream target graph.

### 4.5 Experimental Results and Analysis



(a) FAGCN on Cora (b) PolyGCL on Cora

(c) HS-GPPT on Cora (d) Ours on Cora

Figure 5: Comparative 1-shot learning scatter plots on the Cora dataset. (a) FAGCN; (b) PolyGCL; (c) HS-GPPT; (d) Ours.

As shown in Fig. 5, we present the scatter plot of the 1-shot transductive learning on Cora dataset. We have the following observations: On both homophilic and heterophilic graphs, our method effectively captures structural consistency across different graph domains and topological patterns by forming compact and well-separated clusters through the clustering head. Moreover, our approach maintains clear class separability through the classification head, indicating that the proposed spectral filtering and prompt-tuning mechanisms can flexibly adapt to graphs with varying degrees of homophily and enable robust and stable knowledge transfer under challenging few-shot settings.

## 5 Conclusion

Our study examines graph prompt alignment from a spectral perspective, revealing that spectral bias and prompt misalignment constitute the primary obstacles to cross-homophily generalization. We propose SCA-GPPT, unifying spectrum-based pre-training and cluster-augmented prompting for few-shot node classification. Experimental results indicate that in transductive and inductive settings, SCA-GPPT consistently outperforms existing baselines.

**Broader impact.** The framework provides a principled approach to integrating spectral information and prompts in graph learning, offering new insights into few-shot adaptation and cross-domain transfer, with potential applications in social networks, biological networks, and other graph-structured domains.

**Limitations and future work.** This study focuses on eight real-world datasets and few-shot settings. Future work includes exploring other few-shot ratios, diverse graph types, real-world scenarios with spectral misalignment, and alternative spectral representations or prompt designs to enhance adaptation and robustness under extremely weak supervision conditions.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62506308) and Chengdu Science and Technology Program (No.2025-YF12-00030-RC).

## ContributionStatement

Tingting Li and Xiangkun Wang contributed equally to this work.

## References

[Bo *et al.*, 2020] Deyu Bo, Xiao Wang, Chuan Shi, Meiqi Zhu, Emiao Lu, and Peng Cui. Structural deep clustering network. In *Proceedings of the Web Conference 2020*, pages 1400–1410, 2020.

[Bo *et al.*, 2021] Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI cConference on Artificial Intelligence*, volume 35, pages 3950–3957, 2021.

[Bruna *et al.*, 2014] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *In 2th International Conference on Learning Representations, ICLR 2014*, 2014.

[Caron *et al.*, 2018] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 132–149, 2018.

[Chen *et al.*, 2024] Jingyu Chen, Runlin Lei, and Zhewei Wei. Polygcl: Graph contrastive learning via learnable spectral polynomial filters. In *The Twelfth International Conference on Learning Representations*, 2024.

[Chen *et al.*, 2025] Qin Chen, Liang Wang, Bo Zheng, and Guojie Song. Dagprompt: Pushing the limits of graph prompting with a distribution-aware graph prompt tuning approach. In *Proceedings of the ACM on Web Conference 2025*, pages 4346–4358, 2025.

[Golub and Reinsch, 1971] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer, 1971.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017.

[He *et al.*, 2022] Mingguo He, Zhewei Wei, and Ji-Rong Wen. Convolutional neural networks on graphs with chebyshev approximation, revisited. *Advances in Neural Information Processing Systems*, 35:7264–7276, 2022.

[Jiang *et al.*, 2019] Bo Jiang, Ziyan Zhang, Doudou Lin, Jin Tang, and Bin Luo. Semi-supervised learning with graph learning-convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11313–11320, 2019.

[Ju *et al.*, 2025] Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Zhengyang Mao, Hourun Li, Yiyang Gu, Yifang Qin, Nan Yin, Senzhang Wang, et al. A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[Liu *et al.*, 2023] Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM web conference 2023*, pages 417–428, 2023.

[Liu *et al.*, 2024] Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. Arc: A generalist graph anomaly detector with in-context learning. *Advances in Neural Information Processing Systems*, 37:50772–50804, 2024.

[Luo *et al.*, 2025] Haitong Luo, Suhang Wang, Weiyao Zhang, Ruiqi Meng, Xuying Meng, and Yujun Zhang. Generalize across homophily and heterophily: Hybrid spectral graph pre-training and prompt tuning. *arXiv preprint arXiv:2508.11328*, 2025.

[Pearson, 1901] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[Pei *et al.*, 2020] Hongbin Pei, Bingzhe Wei, Kevin Chen Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[Ren *et al.*, 2024] Lingfei Ren, Ruimin Hu, Zheng Wang, Yilin Xiao, Dengshi Li, Junhang Wu, Yilong Zang, Jinzhang Hu, and Zijun Huang. Heterophilic graph invariant learning for out-of-distribution of fraud detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11032–11040, 2024.

[Rozemberczki *et al.*, 2021] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021.

[Sen *et al.*, 2008] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–93, 2008.

[Shchur *et al.*, 2018] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. In *Proceedings of the Relational Representation Learning Workshop (RRL)*, 2018.

[Sun *et al.*, 2022] Mingchen Sun, Kaixiong Zhou, Xin He, Ying Wang, and Xin Wang. Gppt: Graph pre-training and

prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1717–1727, 2022.

[Velickovic *et al.*, 2017] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[Veličković *et al.*, 2019] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep Graph Infomax. In *International Conference on Learning Representations*, 2019.

[Wan *et al.*, 2024] Guancheng Wan, Yijun Tian, Wenke Huang, Nitesh V Chawla, and Mang Ye. S3gcl: Spectral, swift, spatial graph contrastive learning. In *Forty-first International Conference on Machine Learning*, 2024.

[Xie *et al.*, 2016] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International Conference on Machine Learning*, pages 478–487. PMLR, 2016.

[You *et al.*, 2020] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33:5812–5823, 2020.

[Yu *et al.*, 2024a] Xingtong Yu, Yuan Fang, Zemin Liu, and Xinming Zhang. Hgprompt: Bridging homogeneous and heterogeneous graphs for few-shot prompt learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16578–16586, 2024.

[Yu *et al.*, 2024b] Xingtong Yu, Zhenghao Liu, Yuan Fang, Zemin Liu, Sihong Chen, and Xinming Zhang. Generalized graph prompt: Toward a unification of pre-training and downstream tasks on graphs. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6237–6250, 2024.

[Yu *et al.*, 2025a] Xingtong Yu, Zechuan Gong, Chang Zhou, Yuan Fang, and Hui Zhang. Samgpt: Text-free graph foundation model for multi-domain pre-training and cross-domain adaptation. In *Proceedings of the ACM on Web Conference 2025*, pages 1142–1153, 2025.

[Yu *et al.*, 2025b] Xingtong Yu, Jie Zhang, Yuan Fang, and Renhe Jiang. Non-homophilic graph pre-training and prompt learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 1844–1854, 2025.

[Zang *et al.*, 2025] Yilong Zang, Lingfei Ren, Yue Li, Zhikang Wang, David Antony Selby, Zheng Wang, Sebastian Josef Vollmer, Hongzhi Yin, Jiangning Song, and Junhang Wu. Rethinking cancer gene identification through graph anomaly analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13161–13169, 2025.

[Zhao *et al.*, 2024] Haihong Zhao, Aochuan Chen, Xiangguo Sun, Hong Cheng, and Jia Li. All in one and one for all: A simple yet effective method towards cross-domain graph pretraining. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4443–4454, 2024.

[Zheng *et al.*, 2022] Xin Zheng, Yi Wang, Yixin Liu, Ming Li, Miao Zhang, Di Jin, Philip S Yu, and Shirui Pan. Graph neural networks for graphs with heterophily: A survey. *arXiv preprint arXiv:2202.07082*, 2022.

[Zhu *et al.*, 2020] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in Neural Information Processing Systems*, 33:7793–7804, 2020.