

# Improving Expressive Power of Spectral Graph Neural Networks with Eigenvalue Correction

Kangkang Lu<sup>1</sup>, Yanhua Yu<sup>1,\*</sup>, Hao Fei<sup>2</sup>, Xuan Li<sup>1</sup>, Zixuan Yang<sup>1</sup>, Zirui Guo<sup>1</sup>, Meiyu Liang<sup>1</sup>, Mengran Yin<sup>1</sup>, Tat-Seng Chua<sup>2</sup>

<sup>1</sup>Beijing University of Posts and Telecommunications

<sup>2</sup>National University of Singapore

{lukangkang, yuyanhua}@bupt.edu.cn, haofei37@nus.edu.sg, lixuan20000530@bupt.edu.cn, {zxyang.bupt, zrguo.bupt}@qq.com, {meiyu1210, 2021110996}@bupt.edu.cn, dcscts@nus.edu.sg

## Abstract

In recent years, spectral graph neural networks, characterized by polynomial filters, have garnered increasing attention and have achieved remarkable performance in tasks such as node classification. These models typically assume that eigenvalues for the normalized Laplacian matrix are distinct from each other, thus expecting a polynomial filter to have a high fitting ability. However, this paper empirically observes that normalized Laplacian matrices frequently possess repeated eigenvalues. Moreover, we theoretically establish that the number of distinguishable eigenvalues plays a pivotal role in determining the expressive power of spectral graph neural networks. In light of this observation, we propose an eigenvalue correction strategy that can free polynomial filters from the constraints of repeated eigenvalue inputs. Concretely, the proposed eigenvalue correction strategy enhances the uniform distribution of eigenvalues, thus mitigating repeated eigenvalues, and improving the fitting capacity and expressive power of polynomial filters. Extensive experimental results on both synthetic and real-world datasets demonstrate the superiority of our method.

## Introduction

Graph neural networks (GNNs) have recently achieved excellent performance in graph learning tasks such as drug discovery (Jiang et al. 2021; Bongini, Bianchini, and Scarselli 2021), molecular prediction (Gasteiger, Becker, and Günnemann 2021; Wang et al. 2022), and recommendation systems (Wu et al. 2022; Chang et al. 2021). Existing GNNs can be divided into two categories: *spatial GNNs* and *spectral GNNs*. Spatial GNNs often adopt a message-passing mechanism to learn node representations by aggregating neighbor node features. Spectral GNNs map node features to a new desired space by selectively attenuating or amplifying the Fourier coefficients induced by the normalized Laplacian matrix of the graph. This study primarily centers on the realm of spectral GNNs.

Existing spectral filters are commonly approximated using fixed-order (or truncated) polynomials. Specifically, these polynomial filters take the normalized Laplacian eigenvalues as input and map them to varying scalar values

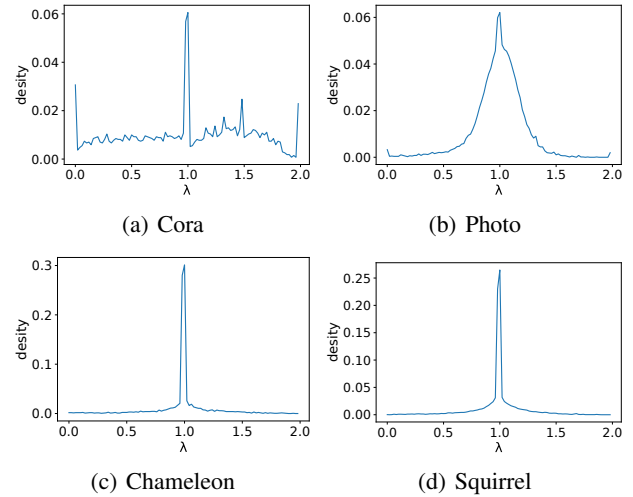


Figure 1: The distribution of eigenvalues of the normalized Laplacian matrix on different datasets. The abscissa represents eigenvalues, and the ordinate represents the probability density.

via a polynomial function. However, our empirical findings unveil that the eigenvalues of normalized Laplacian matrix in real-world graphs tend to exhibit high multiplicity. This phenomenon is visually demonstrated in Figure 1<sup>1</sup>, where a substantial number of eigenvalues cluster around the value 1. Consequently, when two frequency components share the same eigenvalue  $\lambda$ , they will be scaled by the same amount  $h(\lambda)$  (Wang and Zhang 2022). As a result, the coefficients of these frequency components will retain the same proportion in predictions as in the input. And an abundance of repeated eigenvalues inevitably curbs the expressive capacity of polynomial filters. To illustrate, Figure 2 highlights how a filter possessing seven distinct eigenvalues exhibits weaker fitting capabilities than the one endowed with eleven distinct eigenvalues.

Based on the insights gained thus far, it's apparent that the eigenvalue multiplicity of the normalized Laplacian ma-

\*Corresponding author.

<sup>1</sup>Additional distribution histograms for more datasets are provided in Appendix B.

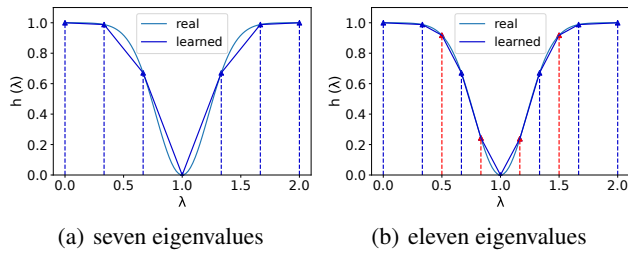


Figure 2: Illustration of two band-rejection filters with different numbers of eigenvalues. The red dotted line in (b) represents more eigenvalues than (a).

trix significantly impedes the polynomial filter’s capacity to effectively fit. More severely, the challenge cannot be readily addressed by simply elevating the polynomial order, because even exceedingly high-order polynomials could fail to differentiate outputs for identical eigenvalues. To tackle this, we propose an **eigenvalue correction strategy** to generate more distinct eigenvalues, which reduces eigenvalue multiplicity and in return enhances the fitting prowess of polynomial filters across intricate and diverse scenarios. In essence, our strategy combines the original eigenvalues and the eigenvalues sampled at equal intervals from  $[0, 2]$ . This amalgamation ensures a more uniform eigenvalue distribution while still retaining the frequency information encapsulated within the eigenvalues. Furthermore, we manage to enhance training efficiency without excessively elongating precomputation time, which allows us to efficiently amplify the order of polynomial spectral GNNs, contributing to a refined and potent modeling capability.

To summarize, our contributions in this work are highlighted as follows:

- We empirically discover that the normalized Laplacian matrix of a majority of real-world graphs exhibits elevated eigenvalue multiplicity. Furthermore, we provide theoretical evidence that the distinctiveness of eigenvalues within the normalized Laplacian matrix critically shapes the efficacy of polynomial filters. Consequently, a surplus of repeated eigenvalues obstructs the potential expressiveness of these filters.
- We introduce an innovative eigenvalue correction strategy capable of substantially diminishing eigenvalue multiplicity. By ingeniously amalgamating original eigenvalues with equidistantly sampled counterparts, our method can seamlessly integrate with existing polynomial filters of GNNs.
- Extensive experiments on synthetic and real-world datasets unequivocally demonstrate the superiority of our proposed approach over state-of-the-art polynomial filters. This substantiates the remarkable enhancement in the fitting prowess and expressive capacity of polynomial filters, achieved through the novel eigenvalue correction strategy we present.

## Related Work

### Spectral GNNs

According to whether the filter can be learned, the spectral GNNs can be divided into *predefined filters* (Kipf and Welling 2017; Gasteiger, Bojchevski, and Günnemann 2019; Zhu et al. 2021) and *learnable filters* (Defferrard, Bresson, and Vandergheynst 2016; Chien et al. 2021; He et al. 2021; Wang and Zhang 2022).

In the category of predefined filters, GCN (Kipf and Welling 2017) uses a simplified first-order Chebyshev polynomial, which is shown to be a low-pass filter. APPNP (Gasteiger, Bojchevski, and Günnemann 2019) utilizes Personalized Page Rank (PPR) to set the weight of the filter. GNN-LF/HF (Zhu et al. 2021) designs filter weights from the perspective of graph optimization functions, which can fit high-pass and low-pass filters.

Another category of spectral GNNs employs learnable filters. ChebNet (Defferrard, Bresson, and Vandergheynst 2016) uses Chebyshev polynomials with learnable coefficients. GPR-GNN (Chien et al. 2021) extends APPNP by directly parameterizing its weights and training them with gradient descent. ARMA (Bianchi et al. 2021) learns a reasonable filter through an autoregressive moving average filter family. BernNet (He et al. 2021) uses Bernstein polynomials to learn filters and forces all coefficients positive. JacobiConv (Wang and Zhang 2022) adopts an orthogonal and flexible Jacobi basis to accommodate a wide range of weight functions. FavardGNN (Guo and Wei 2023) learns an optimal polynomial basis from the space of all possible orthonormal basis.

However, the above filters typically use repeated eigenvalues as input, which hinders their expressiveness. In contrast, we propose a novel eigenvalue correction strategy that can generate more distinct eigenvalues and thus improve the expressive power of polynomial filters.

### Expressive Power of Spectral GNNs

The Weisfeiler-Lehman test (Weisfeiler and Leman 1968) is considered as a classic algorithm for judging graph isomorphism. GIN (Xu et al. 2019) shows that the 1-WL test limits the expressive power of GNNs to distinguish non-isomorphic graphs. Balcilar et al. (2021) theoretically demonstrates some equivalence of graph convolution procedures by bridging the gap between spectral and spatial designs of graph convolutions. JacobiConv (Wang and Zhang 2022) points out that the condition for spectral GNN to generate arbitrary one-dimensional predictions is that the normalized Laplacian matrix has no repeated eigenvalues and no missing frequency components of node features.

### Distribution of Eigenvalues of Normalized Laplacian Matrix

For random graphs, Jiang (2012) proves that the empirical distribution of the eigenvalues of the normalized Laplacian matrix of Erdős-Rényi random graphs converges to the semicircle law. Chung, Lu, and Vu (2003) also proves that the Laplacian spectrum of a random graph with expected degree obeys the semicircle law. Empirically, we observe

datasets	Cora	Citeseer	Pubmed	Computers	Photo	Texas	Cornell	Squirrel	Chameleon	Actor
$\ \mathcal{V}\ $	2708	3327	19717	13752	7650	183	183	5201	2277	7600
$\ \mathcal{E}\ $	5278	4552	44324	245861	119081	279	277	198353	31371	26659
$N_{\text{distinct}}$	2199	1886	7595	13344	7474	106	115	3272	1120	6419
$P_{\text{distinct}}$	81.2%	56.7%	38.5%	97.0%	97.7%	57.9%	62.8%	62.9%	49.2%	84.5%

Table 1: Dataset statistics.  $\|\mathcal{V}\|$  is the number of nodes,  $\|\mathcal{E}\|$  is the number of edges,  $N_{\text{distinct}}$  is the number of distinct eigenvalues, and  $P_{\text{distinct}}$  is the proportion of distinct eigenvalues to all eigenvalues.

that as the average degree of a random graph increases, its eigenvalue distribution becomes more concentrated around 1. Histograms of their distributions can be found in Appendix C.

For real-world graphs, we empirically find that almost all real-world graphs have many repeated eigenvalues, especially around the value 1, which is consistent with the findings of QPGCN (Wu et al. 2023) and SignNet (Lim et al. 2023). Furthermore, some works (Banerjee and Jost 2008; Mehatari and Banerjee 2015) attempt to explain the generation of the multiplicity of eigenvalue 1 by motif doubling and attachment. The number and proportion of distinct eigenvalues for real-world graphs are shown in Table 1. It’s evident that almost all datasets do not have enough distinct eigenvalues, and some are even lower than 50%, which will hinder the fitting ability and expressive power of polynomial filters.

## Proposed Method

### Preliminaries

Assume that we have a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ , where  $\mathcal{V} = \{v_1, \dots, v_n\}$  denotes the vertex set of  $n$  nodes,  $\mathcal{E}$  is the edge set and  $\mathbf{X} \in \mathbb{R}^{n \times d}$  is node feature matrix. The corresponding adjacency matrix is  $\mathbf{A} \in \{0, 1\}^{n \times n}$ , where  $\mathbf{A}_{ij} = 1$  if there is an edge between nodes  $v_i$  and  $v_j$ , and  $\mathbf{A}_{ij} = 0$  otherwise. The degree matrix  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  of  $\mathbf{A}$  is a diagonal matrix with its  $i$ -th diagonal entry as  $d_i = \sum_j \mathbf{A}_{ij}$ . The normalized adjacency matrix is  $\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ . Let  $\mathbf{I}$  denote the identity matrix. The normalized Laplacian matrix  $\hat{\mathbf{L}} = \mathbf{I} - \hat{\mathbf{A}}$ . Let  $\hat{\mathbf{L}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top$  denote the eigen-decomposition of  $\hat{\mathbf{L}}$ , where  $\mathbf{U}$  is the matrix of eigenvectors and  $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda}) = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_n])$  is the diagonal matrix of eigenvalues.

The Fourier transform of a graph signal  $\mathbf{x}$  is defined as  $\hat{\mathbf{x}} = \mathbf{U}^\top \mathbf{x}$ , and its inverse is  $\mathbf{x} = \mathbf{U} \hat{\mathbf{x}}$ . Thus the graph propagation for signal  $\mathbf{x}$  with kernel  $\mathbf{g}$  can be defined as:

$$\mathbf{z} = \mathbf{g} *_G \mathbf{x} = \mathbf{U} ((\mathbf{U}^\top \mathbf{g}) \odot \mathbf{U}^\top \mathbf{x}) = \mathbf{U} \hat{\mathbf{G}} \mathbf{U}^\top \mathbf{x}, \quad (1)$$

where  $\hat{\mathbf{G}} = \text{diag}(\hat{g}_1, \dots, \hat{g}_n)$  denotes the spectral kernel coefficients. In Eq. (1), the signal  $\mathbf{x}$  is first transformed from the spatial domain to the spectral domain, then signals of different frequencies are enhanced or weakened by the filter  $\hat{\mathbf{G}}$ , and finally transformed back to the spatial domain to obtain the filtered signal  $\mathbf{z}$ . To avoid eigen-decomposition, current works with spectral convolution usually use the polynomial functions  $h(\hat{\mathbf{L}})$  to approximate different kernels, we

call them *polynomial spectral GNNs*:

$$\mathbf{Z} = h(\hat{\mathbf{L}}) \mathbf{X} \mathbf{W} = \mathbf{U} h(\mathbf{\Lambda}) \mathbf{U}^\top \mathbf{X} \mathbf{W}, \quad (2)$$

where  $\mathbf{W}$  is a learnable weight matrix for feature mapping and  $\mathbf{Z}$  is the prediction matrix.

### Expressive Power of Existing Polynomial Spectral GNNs

For the expressive power of polynomial spectral GNNs, The following lemma is given in JacobiConv:

**Lemma 1** *polynomial spectral GNNs can produce any one-dimensional prediction if  $\hat{\mathbf{L}}$  has no repeated eigenvalues and  $\mathbf{X}$  contains all frequency components.*

The proof of this lemma can be found in Appendix B.1 of JacobiConv (Wang and Zhang 2022). The lemma shows that the expressive power of polynomial spectral GNNs is related to the repeated eigenvalues.

To focus on exploring the fitting ability of the filter  $h(\mathbf{\Lambda})$  and avoid the influence of  $\mathbf{W}$ , we fix the weight matrix  $\mathbf{W}$  and assume that all elements of  $(\mathbf{U}^\top \mathbf{X} \mathbf{W})$  are non-zero. Thus, for the relationship between the expressive power of polynomial spectral GNNs and the distinct eigenvalues, we have the following theorem:

**Theorem 1** *When there are only  $k$  distinct eigenvalues of the normalized Laplacian matrix, polynomial spectral GNNs can produce at most  $k$  different filter coefficients, and thus can only generate one-dimensional predictions with a maximum of  $k$  arbitrary elements.*

We provide the proof of this theorem in Appendix A.1. Theorem 1 strictly shows that when the eigenvalues are more distinct, the polynomial spectral GNNs will be more expressive. In other words, the more repeated eigenvalues, the less expressive polynomial spectral GNNs are. Thus, the number of distinct eigenvalues is the key to the expressive power of polynomial spectral GNNs. Therefore, we need to reduce the repeated eigenvalues of the normalized Laplacian matrix as much as possible, given that real-world graphs often have many repeated eigenvalues.

### Eigenvalue Correction

We now elaborate on the proposed eigenvalue correction strategy. In order to reduce repeated eigenvalues, a natural solution is that we can sample equally spaced from  $[0, 2]$  as new eigenvalues  $\mathbf{v}$ . Thus, the  $i$ -th value of  $\mathbf{v}$  is defined as follows:

$$v_i = \frac{2i}{n-1}, \quad i \in \{0, 1, 2, \dots, n-1\}, \quad (3)$$

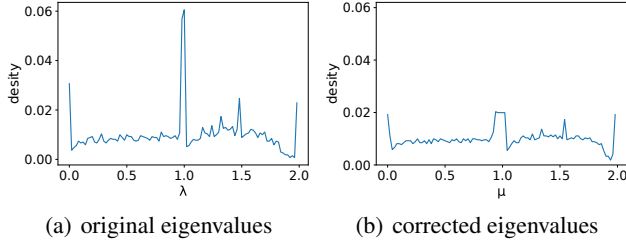


Figure 3: Histogram of the probability density distribution of the original eigenvalues and the corrected eigenvalues.

where  $n$  is the number of nodes, which is also the number of eigenvalues. We serialize the vector  $\mathbf{v}$  into  $\{v_i\}$ . Obviously, the sequence  $\{v_i\}$  is a strictly monotonically increasing arithmetic sequence with 0 as the first item and  $\frac{2}{n-1}$  as the common difference. We call  $\mathbf{v}$  equidistant eigenvalues.

However, the original eigenvalue  $\lambda$  of the normalized Laplacian matrix describes the frequency information and represents the speed of the signal change. Specifically, the magnitude of the eigenvalue represents the difference information of the signal between the neighbors of the node. Therefore, if we directly use the  $\mathbf{v}$  in Eq. (3) as the new eigenvalues, polynomial spectral GNNs will not be able to capture the frequency information represented by each frequency component. Hence, it is still necessary to include the original eigenvalues in the input. A simple way is to amalgamate them linearly to obtain the new eigenvalues  $\mu$ :

$$\mu_i = \beta\lambda_i + (1 - \beta)v_i, \quad i \in \{0, 1, 2, \dots, n-1\}, \quad (4)$$

where  $\beta \in [0, 1]$  is a tunable hyperparameter, and  $\{\lambda_i\}$  is a sequence of original eigenvalues  $\lambda$  in ascending order. Thus, we have the following theorem:

**Theorem 2** *When  $\beta \in [0, 1]$ , the sequence  $\{\mu_i\}$  is strictly monotonically increasing, that is, all  $n$  elements of the sequence  $\{\mu_i\}$  are different from each other.*

The proof of this theorem is given in Appendix A.2. Theorem 2 indicates that the new eigenvalues  $\mu = \{\mu_1, \dots, \mu_n\}$  have no repeated elements when  $\beta \in [0, 1]$ , thus satisfying the condition of Lemma 1. It should be noted that if the spacing between two eigenvalues is very small, it is difficult for a polynomial spectral GNN to generate different filter coefficients for them. Therefore, the spacing between eigenvalues is crucial to guarantee the expressive power of polynomial spectral GNNs.

It is worth noting that  $\beta$  is a key hyperparameter in Eq. (4), which can be used to trade off between  $\{\lambda_i\}$  and  $\{v_i\}$ . The smaller  $\beta$ , the closer  $\{\mu_i\}$  is to  $\{v_i\}$ , and the more uniform the spacing between the new eigenvalues  $\{\mu_i\}$ , but the frequency information may not be captured. On the contrary, the larger  $\beta$ , the closer  $\{\mu_i\}$  is to  $\{\lambda_i\}$ , and the closer the new eigenvalue is to the frequency it represents, but the spacing between the eigenvalues may be uneven. When  $\beta = 1$ ,  $\{\mu_i\}$  degenerates into the original eigenvalues  $\{\lambda_i\}$  of the Laplacian matrix. When  $\beta = 0$ ,  $\{\mu_i\}$  degenerates to the equidistantly sampled  $\{v_i\}$  in Eq. (3), whose spacing is absolutely uniform. In short, the hyperparameter  $\beta$  needs to

be set to an appropriate value so that the spacing between eigenvalues is relatively uniform and not too far away from the original eigenvalues.

For example, the Figure 3 shows the distribution histogram of the original eigenvalues  $\{\lambda_i\}$  and the corrected eigenvalues  $\{\mu_i\}$  on Cora dataset when  $\beta=0.5$ . We can see that, compared to Figure 3(a), Figure 3(b) has no sharp bumps, and the distribution of eigenvalues is more uniform. Hence, it is expected to fit more complex and diverse filters.

### Combination of Corrected Eigenvalues with Existing Polynomial Filters

In order to fully take advantage of polynomial spectral GNNs, after getting the corrected eigenvalues, We are required to integrate them with a polynomial filter. First, we set  $\mathbf{E} = \mathbf{U} \text{diag}(\mathbf{v})\mathbf{U}^\top$ ,  $\mathbf{H} = \mathbf{U} \text{diag}(\mu)\mathbf{U}^\top$ . Combined with Eq. (4) and  $\hat{\mathbf{L}} = \mathbf{U} \text{diag}(\lambda)\mathbf{U}^\top$ , we have:

$$\mathbf{H} = \beta\hat{\mathbf{L}} + (1 - \beta)\mathbf{E}. \quad (5)$$

Then, a polynomial  $h(\mathbf{H})$  of order  $K$  can be written as:

$$h(\mathbf{H}) = \sum_{k=0}^K \alpha_k \mathbf{H}^k = \sum_{k=0}^K \alpha_k (\beta\hat{\mathbf{L}} + (1 - \beta)\mathbf{E})^k, \quad (6)$$

where  $h(x) = \sum_{k=0}^K \alpha_k x^k$ . The difference between the proposed method and the existing polynomial filter is that  $\hat{\mathbf{L}}$  is replaced by  $\mathbf{H}$ . Therefore, our method can be quite flexibly integrated into existing polynomial filters without excessive modifications. However, when the degree of the polynomial is high, the matrix multiplication will be computationally expensive, especially for dense graphs. Fortunately, we can compute  $h(\mathbf{H})$  by directly using polynomials of the eigenvalues instead of polynomials of the matrix:

$$h(\mathbf{H}) = h(\mathbf{U} \text{diag}(\mu)\mathbf{U}^\top) = \mathbf{U} \sum_{k=0}^K \text{diag}(\alpha_k \mu^k) \mathbf{U}^\top. \quad (7)$$

We find that using Eq. (7) to calculate  $h(\mathbf{H})$  is much faster than Eq. (6), resulting in improved training efficiency. Moreover, we can increase the order  $K$  to improve the fitting ability without increasing too much computational cost.

The complete filtering process can be expressed as:

$$\mathbf{Z} = h(\mathbf{H})\mathbf{X}\mathbf{W} = \mathbf{U} \sum_{k=0}^K \text{diag}(\alpha_k \mu^k) \mathbf{U}^\top \mathbf{X}\mathbf{W}. \quad (8)$$

Next, we present the details of the proposed eigenvalue correction strategy integrated into existing polynomial filters.

**Insertion into GPR-GNN.** GPR-GNN (Chien et al. 2021) directly assigns a learnable coefficient to each order of the normalized adjacency matrix  $\hat{\mathbf{A}}$ , and its polynomial filter is defined as:

$$\sum_{k=0}^K \gamma_k \hat{\mathbf{A}}^k = \mathbf{U} g_{\gamma, K}(\Lambda) \mathbf{U}^\top, \quad (9)$$

where  $g_{\gamma, K}(\Lambda)$  is an element-wise operation, and  $g_{\gamma, K}(x) = \sum_{k=0}^K \gamma_k x^k$ . It is worth noting that GPR-GNN uses a polynomial of the normalized adjacency

matrix instead of the Laplacian matrix. Fortunately, they are constrained via  $\hat{\mathbf{A}} = \mathbf{I} - \hat{\mathbf{L}}$ . Therefore, we integrate the corrected eigenvalues  $\boldsymbol{\mu}$  into GPR-GNN to obtain a new polynomial filter:

$$\sum_{k=0}^K \gamma_k \hat{\mathbf{A}}^k = \mathbf{U} \sum_{k=0}^K \text{diag}(\gamma_k (1 - \boldsymbol{\mu})^k) \mathbf{U}^T. \quad (10)$$

**Insertion into BernNet.** BernNet (He et al. 2021) expresses the filtering operation with Bernstein polynomials and forces all coefficients to be positive, and its filter is defined as:

$$\sum_{k=0}^K \theta_k \frac{1}{2^K} \binom{K}{k} (2\mathbf{I} - \mathbf{L})^{K-k} \mathbf{L}^k. \quad (11)$$

Similarly, we have:

$$\mathbf{U} \sum_{k=0}^K \text{diag}(\theta_k \frac{1}{2^K} \binom{K}{k} (2 - \boldsymbol{\mu})^{K-k} \odot \boldsymbol{\mu}^k) \mathbf{U}^T. \quad (12)$$

**Insertion into JacobiConv.** JacobiConv (Wang and Zhang 2022) proposes a Jacobi basis to adapt a wide range of weight functions due to its orthogonality and flexibility. The iterative process of Jacobi basis can be defined as:

$$\begin{aligned} P_0^{a,b}(x) &= 1, \\ P_1^{a,b}(x) &= 0.5a - 0.5b + (0.5a + 0.5b + 1)x, \\ P_k^{a,b}(x) &= (2k + a + b - 1) \\ &\cdot \frac{(2k + a + b)(2k + a + b - 2)x + a^2 - b^2}{2k(k + a + b)(2k + a + b - 2)} P_{k-1}^{a,b}(x) \\ &- \frac{(k + a - 1)(k + b - 1)(2k + a + b)}{k(k + a + b)(2k + a + b - 2)} P_{k-2}^{a,b}(x), \end{aligned} \quad (13)$$

where  $a$  and  $b$  are tunable hyperparameters. Unlike GPR-GNN and BernNet, JacobiConv adopts an individual filter function for each output dimension  $l$ :

$$\mathbf{Z}_{:l} = \sum_{k=0}^K \alpha_{kl} P_k^{a,b}(\hat{\mathbf{A}})(\mathbf{XW})_{:l}. \quad (14)$$

Then, we integrate the corrected eigenvalues  $\boldsymbol{\mu}$  into the JacobiConv:

$$\mathbf{Z}_{:l} = \mathbf{U} \sum_{k=0}^K \text{diag}(\alpha_{kl} \cdot P_k^{a,b}(1 - \boldsymbol{\mu})) \mathbf{U}^T (\mathbf{XW})_{:l}. \quad (15)$$

## Experiment

In this section, we conduct a series of comprehensive experiments to demonstrate the effectiveness of the proposed method. First, we integrate the eigenvalue correction strategy into the three methods of GPR-GNN, BernNet, and JacobiConv to complete the learning of filter functions on synthetic datasets and node classification tasks on real-world datasets. Then, we report experiments on ablation analysis and sensitivity analysis of the hyperparameter  $\beta$ . Finally, we show the evaluation of the time cost. All experiments are conducted on a machine with 3 NVIDIA A5000 24GB GPUs and Intel(R) Xeon(R) Silver 4310 2.10GHz CPU.

	Low	High	Band	Reject	Comb
GCN	3.4799	67.6635	25.8755	21.0747	50.5120
GAT	2.3574	21.9618	14.4326	12.6384	23.1813
ARMA	1.8478	1.8632	7.6922	8.2732	15.1214
ChebyNet	0.8220	0.7867	2.2722	2.5296	4.0735
GPR-GNN	0.4169	0.0943	3.5121	3.7917	4.6549
<b>EC-GPR</b>	<b>0.2703</b>	<b>0.0656</b>	<b>0.2920</b>	<b>1.1266</b>	<b>1.0681</b>
BernNet	0.0314	<b>0.0113</b>	0.0411	0.9313	0.9982
<b>EC-Bern</b>	<b>0.0277</b>	<b>0.0113</b>	<b>0.0058</b>	<b>0.2391</b>	<b>0.3302</b>
JacobiConv	<b>0.0003</b>	<b>0.0011</b>	0.0213	0.0156	0.2933
<b>EC-Jacobi</b>	<b>0.0003</b>	<b>0.0011</b>	<b>0.0088</b>	<b>0.0018</b>	<b>0.0370</b>

Table 2: Average loss in learning filters experiments.

## Evaluation on Learning Filters

Following previous works (He et al. 2021; Wang and Zhang 2022), we transform 50 real images to 2D regular 4-neighbor grid graphs. The nodes are pixels, and every two neighboring pixels have an edge between the corresponding nodes. Five predefined graph filters, i.e., low ( $e^{-10\lambda^2}$ ), high ( $1 - e^{-10\lambda^2}$ ), band ( $e^{-10(\lambda-1)^2}$ ), reject ( $1 - e^{-10(\lambda-1)^2}$ ), and comb ( $|\sin \pi \lambda|$ ) are used to generate ground truth graph signals. We use the mean squared error of 50 images as the evaluation metric.

To verify the effectiveness of the proposed Eigenvalue Correction (EC) strategy, we integrate it into three representative spectral GNNs with trainable polynomial filters, i.e., GPR-GNN (Chien et al. 2021), BernNet (He et al. 2021), and JacobiConv (Wang and Zhang 2022). Thus, we have three variants: EC-GPR, EC-Bern, and EC-Jacobi. For fairness, all experimental settings are the same as the base model, including dataset partition, hyperparameters, random seeds, etc. We just tune the  $\beta \in \{0, 0.1, \dots, 0.9\}$  hyperparameter to select the best model. In addition, for a more comprehensive comparison, we also consider four other baseline methods, including GCN (Kipf and Welling 2017), GAT (Veličković et al. 2019), ARMA (Bianchi et al. 2021), and ChebNet (Defferrard, Bresson, and Vandergheynst 2016).

The quantitative experiment results are shown in Table 2, revealing that our methods all achieve the best performance compared to the base models. For example, on low-pass and high-pass filters, our method outperforms the base model by an average of 12.9%, while on band-pass, comb-pass, and reject-pass filters, our method outperforms the base model by an average of 77.9%. Since complex filters require more frequency components to fit, our method improves more on band-pass, comb-pass, and reject-pass filters. Moreover, GCN and GAT only perform better on low-pass filters applied to homophilic graphs, illustrating the limited fitting ability of fixed filters. In contrast, polynomial-based GNNs perform better due to the ability to learn arbitrary filters. Nevertheless, they all take a large number of repeated eigenvalues as input, therefore, their expressive power is still weaker than our methods.

## Evaluation on Real-World Datasets

We also evaluate the proposed method on the node classification task on real-world datasets. Following Jacobi-



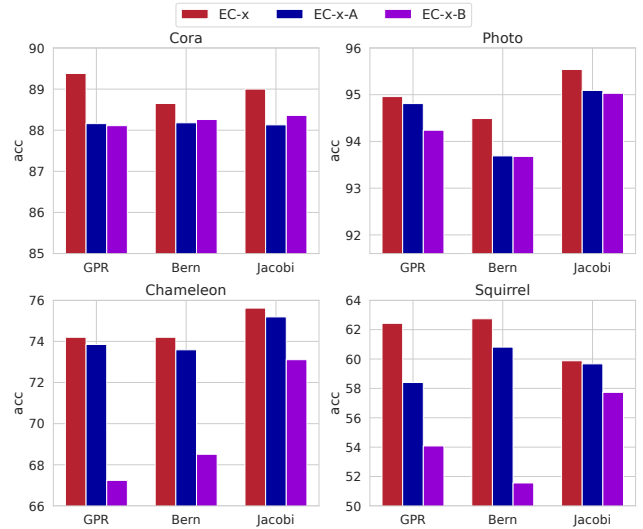
Datasets	GCN	APPNP	Chebynet	GPR-GNN	EC-GPR	BernNet	EC-Bern	JacobiConv	EC-Jacobi
Cora	87.14 $\pm$ 1.01	88.14 $\pm$ 0.73	86.67 $\pm$ 0.82	88.57 $\pm$ 0.69	<b>89.38<math>\pm</math>0.95</b>	88.52 $\pm$ 0.95	<b>88.65<math>\pm</math>0.87</b>	88.98 $\pm$ 0.46	<b>89.00<math>\pm</math>0.67</b>
Citeseer	79.86 $\pm$ 0.67	80.47 $\pm$ 0.74	79.11 $\pm$ 0.75	80.12 $\pm$ 0.83	<b>80.66<math>\pm</math>0.85</b>	80.09 $\pm$ 0.79	<b>80.26<math>\pm</math>0.60</b>	80.78 $\pm$ 0.79	<b>81.23<math>\pm</math>1.00</b>
Pubmed	86.74 $\pm$ 0.27	88.12 $\pm$ 0.31	87.95 $\pm$ 0.28	88.46 $\pm$ 0.33	<b>89.63<math>\pm</math>0.31</b>	88.48 $\pm$ 0.41	<b>89.10<math>\pm</math>0.34</b>	89.62 $\pm$ 0.41	<b>89.64<math>\pm</math>0.37</b>
Computers	83.32 $\pm$ 0.33	85.32 $\pm$ 0.37	87.54 $\pm$ 0.43	86.85 $\pm$ 0.25	<b>89.89<math>\pm</math>0.43</b>	87.64 $\pm$ 0.44	<b>88.36<math>\pm</math>0.48</b>	<b>90.39<math>\pm</math>0.29</b>	90.33 $\pm$ 0.28
Photo	88.26 $\pm$ 0.73	88.51 $\pm$ 0.31	93.77 $\pm$ 0.32	93.85 $\pm$ 0.28	<b>94.76<math>\pm</math>0.94</b>	93.63 $\pm$ 0.35	<b>94.49<math>\pm</math>0.30</b>	95.43 $\pm$ 0.23	<b>95.54<math>\pm</math>0.34</b>
Chameleon	59.61 $\pm$ 2.21	51.84 $\pm$ 1.82	59.28 $\pm$ 1.25	67.28 $\pm$ 1.09	<b>74.20<math>\pm</math>1.07</b>	68.29 $\pm$ 1.58	<b>74.20<math>\pm</math>1.31</b>	74.20 $\pm$ 1.03	<b>75.62<math>\pm</math>1.51</b>
Actor	33.23 $\pm$ 1.16	39.66 $\pm$ 0.55	37.61 $\pm$ 0.89	39.92 $\pm$ 0.67	<b>40.43<math>\pm</math>0.84</b>	41.79 $\pm$ 1.01	<b>41.88<math>\pm</math>1.05</b>	<b>41.17<math>\pm</math>0.64</b>	40.99 $\pm$ 0.70
Squirrel	46.78 $\pm$ 0.87	34.71 $\pm$ 0.57	40.55 $\pm$ 0.42	50.15 $\pm$ 1.92	<b>62.43<math>\pm</math>0.67</b>	51.35 $\pm$ 0.73	<b>62.74<math>\pm</math>1.04</b>	57.38 $\pm$ 1.25	<b>59.88<math>\pm</math>0.86</b>
Texas	77.38 $\pm$ 3.28	90.98 $\pm$ 1.64	86.22 $\pm$ 2.45	<b>92.95<math>\pm</math>1.31</b>	92.30 $\pm$ 1.80	93.12 $\pm$ 0.65	<b>94.43<math>\pm</math>1.31</b>	93.44 $\pm$ 2.13	<b>93.44<math>\pm</math>1.48</b>
Cornell	65.90 $\pm$ 4.43	91.81 $\pm$ 1.96	83.93 $\pm$ 2.13	<b>91.37<math>\pm</math>1.81</b>	90.66 $\pm$ 2.13	92.13 $\pm$ 1.64	<b>93.77<math>\pm</math>0.98</b>	92.95 $\pm$ 2.46	<b>93.28<math>\pm</math>2.30</b>

Table 3: Results on real-world datasets: Mean accuracy (%)  $\pm$  95% confidence interval.

Conv (Wang and Zhang 2022), for the homophilic graphs, we employ three citation datasets Cora, CiteSeer and PubMed (Yang, Cohen, and Salakhudinov 2016), and two co-purchase datasets Computers and Photo (Shchur et al. 2018). For heterophilic graphs, we use Wikipedia graphs Chameleon and Squirrel (Rozemberczki, Allen, and Sarkar 2021), the Actor co-occurrence graph, and the webpage graph Texas and Cornell from WebKB3 (Pei et al. 2020). Their statistics are listed in Table 1.

Following previous work (Wang and Zhang 2022), we use the full-supervised split, i.e., 60% for training, 20% for validation, and 20% for testing. Mean accuracy and 95% confidence intervals over 10 runs are reported. We integrate the eigenvalue correction strategy into GPR-GNN (Chien et al. 2021), BernNet (He et al. 2021), and JacobiConv (Wang and Zhang 2022). Moreover, we also add the results of GCN (Kipf and Welling 2017), APPNP (Gasteiger, Bojchevski, and Günnemann 2019), and ChebyNet (Defferrard, Bresson, and Vandergheynst 2016). For a fair comparison, we only tune the hyperparameter  $\beta \in \{0, 0.01, \dots, 0.99\}$ , and the other experimental settings remain the same as the base model. All hyperparameter  $\beta$  settings can be found in Appendix D.

Table 3 presents the experimental results on real-world datasets, from which we have the following observations. 1) GNNs with learnable filters tend to perform better than GNNs with fixed filters since fixed filters, such as GCN and APPNP, cannot learn complex filter responses. In contrast, GPR-GNN, BernNet, and JacobiConv can learn different types of filters from both homophilic graphs and heterophilic graphs. 2) The proposed method outperforms the base models in at least eight of the ten datasets, especially on the Squirrel dataset, with an average improvement of 17% over the three base models. This is because heterophilic graphs require more frequency components than homophilic graphs to fit the patterns. In general, we achieve a 4.2% improvement in GPR-GNN, 3.7% improvement in BernNet, and 0.68% improvement in JacobiConv, demonstrating that our proposed eigenvalue correction strategy can improve the expressive power of polynomial spectral GNNs. 3) With our eigenvalue correction strategy, GPR-GNN and BernNet achieve competitive performance with JacobiConv, which shows that the expressive power of different polynomial filters with the same order is the same.

Figure 4: Ablation study of proposed method EC- $x$  on four datasets with our variants EC- $x$ -A and EC- $x$ -B for all  $x \in \{\text{GPR, Bern, Jacobi}\}$ .

## Ablation Analysis

This subsection aims to validate our design through ablation studies. Previously, we believe that neither the original eigenvalues nor the equally spaced eigenvalues are optimal, while only combining them in a certain proportion can achieve the optimal performance. Therefore, we design two variants EC- $x$ -A and EC- $x$ -B to verify our conjecture, where  $x \in \{\text{GPR, Bern, Jacobi}\}$ . EC- $x$ -A directly uses the equidistant eigenvalues  $v$  as input, that is,  $\beta$  in Eq. (4) equals to 0. In contrast, EC- $x$ -B directly uses the original eigenvalues  $\lambda$  as input, that is,  $\beta$  in Eq. (4) equals to 1.

Figure 4 shows the results of the ablation experiments on two homophilic graphs, Cora and Photo, and two heterophilic graphs, Chameleon and Squirrel. From Figure 4, we have the following findings: 1) The utilization of solely the original eigenvalues and equidistant eigenvalues as input demonstrates a decrease in performance, corroborating our hypothesis necessitating their amalgamation. 2) EC- $x$ -B of the heterophilic graphs drops more than the homophilic graphs, which indicates that the heterophilic graphs need

Datasets	GPR-GNN	EC-GPR	BernNet	EC-Bern	JacobiConv	EC-Jacobi	Decomposition
Cora	6.73/1.40	4.09/0.84	22.25/6.06	4.73/1.28	9.50/5.02	9.29/3.74	0.62
Citeseer	6.90/1.46	4.09/0.93	22.12/6.46	4.39/1.29	9.39/4.71	9.45/3.07	0.82
Pubmed	7.18/1.46	9.15/1.86	20.45/8.70	9.73/4.15	9.90/7.63	13.79/11.26	58.59
Computers	8.20/1.88	7.83/1.09	29.40/8.32	8.49/2.31	9.81/5.58	11.91/7.92	27.53
Photo	7.09/1.56	4.54/1.23	21.04/8.74	5.26/2.21	9.23/6.66	9.97/8.18	3.86
Chameleon	7.31/1.65	4.12/1.00	22.61/5.12	5.01/1.18	9.32/7.47	9.62/7.10	0.27
Actor	7.17/2.54	4.53/1.49	18.74/6.68	5.44/2.02	9.37/4.25	9.97/3.86	3.56
Squirrel	7.23/3.65	3.96/1.75	22.66/7.54	2.42/0.63	9.44/8.16	9.60/7.86	1.61
Texas	6.83/1.41	3.80/0.79	22.42/4.76	4.50/0.95	9.58/6.47	9.33/6.48	0.03
Cornell	7.17/1.48	3.89/0.80	22.34/4.66	4.49/0.93	9.10/5.30	9.53/3.18	0.02

Table 4: Per-epoch time (ms)/total training time (s) and eigen-decomposition time (s).

more frequency component information to fit various responses.

### Parameter Sensitivity Analysis

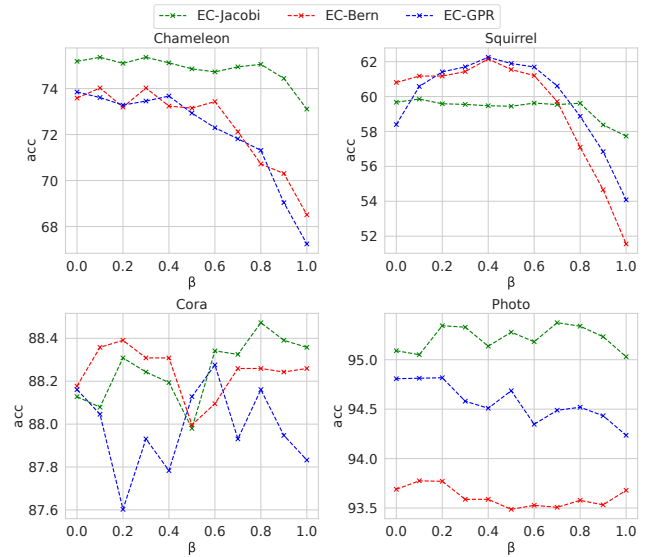
As discussed above, we add a key hyperparameter  $\beta$  to adjust the ratio of the original eigenvalue  $\lambda$  and the equidistant eigenvalue  $v$ . Now we analyze the impact of this hyperparameter on the performance of homophilic graphs and heterophilic graphs. Similar to the ablation experiments, we select two homophilic datasets Cora and Photo, and two heterophilic datasets Chameleon and Squirrel. Figure 5 shows the effect of the hyperparameter  $\beta$  on the performance of the three methods EC-GPR, EC-Bern, and EC-Jacobi.

Figure 5 indicates that on the Cora dataset, the general trend of model performance is to decrease with the increase of  $\beta$ , and then increase, which shows that both the equidistant eigenvalues and original eigenvalues are useful. The trend on the Photo dataset is not prominent, possibly due to its limited occurrence of repeated eigenvalues. In comparison, for heterophilic graphs, the model performance is maintained at a high level when  $\beta$  is less than 0.6, and when  $\beta$  is greater than 0.6, the model performance drops significantly. This suggests that equidistant eigenvalues contribute more to heterophilic graphs, which is consistent with our previous analysis.

### Evaluation on Time Cost

Table 4 shows our time cost compared with the base models GPR-GNN, BernNet and JacobiConv. We can conclude that our time spent with the JacobiConv model is similar, 30.9% less than GPR-GNN and 75.5% less than BernNet. In particular, we significantly improve the training efficiency of BernNet, which is a quadratic time complexity related to the order  $K$  of its polynomial. This is because we change the time-consuming polynomial of the matrix to the polynomial of the eigenvalues, thus reducing the calculation time.

Since obtaining eigenvalues and eigenvectors requires eigen-decomposition, we also add the time overhead of eigen-decomposition in Table 4. We can see from Table 4, the eigen-decomposition time for most datasets is less than 5 seconds, even for a large graph with nearly 20,000 nodes like Pubmed, the time for eigen-decomposition is less than one minute. For larger graphs, we can follow previous works (Lim et al. 2023; Bo et al. 2023; Lin, Chen,

Figure 5: Effect of hyperparameter  $\beta$  on model performance.

and Wang 2023; Yao, Yu, and Jiao 2021) and only take the largest and smallest  $k$  eigenvalues and eigenvectors without a full eigen-decomposition. Moreover, the eigen-decomposition only needs to be precomputed once and can be reused in training afterward. In summary, the cost of eigen-decomposition is acceptable and negligible.

### Conclusions

In this paper, we point out that the current polynomial spectral GNNs have insufficient expressive power due to a large number of repeated eigenvalues, so we propose a novel eigenvalue correction strategy that amalgamates the original eigenvalues with equidistant eigenvalues counterparts to make the distribution of eigenvalues more uniform, and not to deviate too far from the original frequency information. Moreover, we improve the training efficiency without increasing the precomputation time too much, which allows us to efficiently increase the order of polynomial spectral GNNs. Extensive experiments on synthetic and real-world datasets demonstrate that the proposed eigenvalue correction strategy outperforms the base model.

## Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant No. U22B2019), National key research and development plan (Grant No. 2020YFB2104503), BUPT Excellent Ph.D. Students Foundation(CX2023128) and CCF-Baidu Open Fund.

## References

- Balcilar, M.; Guillaume, R.; Héroux, P.; Gaüzère, B.; Adam, S.; and Honeine, P. 2021. Analyzing the expressive power of graph neural networks in a spectral perspective. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Banerjee, A.; and Jost, J. 2008. On the spectrum of the normalized graph Laplacian. *Linear algebra and its applications*, 428(11-12): 3015–3022.
- Bianchi, F. M.; Grattarola, D.; Livi, L.; and Alippi, C. 2021. Graph neural networks with convolutional arma filters. *IEEE transactions on pattern analysis and machine intelligence*, 44(7): 3496–3507.
- Bo, D.; Shi, C.; Wang, L.; and Liao, R. 2023. Specformer: Spectral Graph Neural Networks Meet Transformers. In *The Eleventh International Conference on Learning Representations*.
- Bongini, P.; Bianchini, M.; and Scarselli, F. 2021. Molecular generative graph neural networks for drug discovery. *Neurocomputing*, 450: 242–252.
- Chang, J.; Gao, C.; Zheng, Y.; Hui, Y.; Niu, Y.; Song, Y.; Jin, D.; and Li, Y. 2021. Sequential recommendation with graph neural networks. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 378–387.
- Chien, E.; Peng, J.; Li, P.; and Milenkovic, O. 2021. Adaptive Universal Generalized PageRank Graph Neural Network. In *International Conference on Learning Representations*.
- Chung, F.; Lu, L.; and Vu, V. 2003. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 100(11): 6313–6318.
- Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29.
- Gasteiger, J.; Becker, F.; and Günnemann, S. 2021. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34: 6790–6802.
- Gasteiger, J.; Bojchevski, A.; and Günnemann, S. 2019. Predict then Propagate: Graph Neural Networks meet Personalized PageRank. In *International Conference on Learning Representations*.
- Guo, Y.; and Wei, Z. 2023. Graph Neural Networks with Learnable and Optimal Polynomial Bases. *arXiv preprint arXiv:2302.12432*.
- He, M.; Wei, Z.; Xu, H.; et al. 2021. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. *Advances in Neural Information Processing Systems*, 34: 14239–14251.
- Jiang, D.; Wu, Z.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; and Hou, T. 2021. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of cheminformatics*, 13(1): 1–23.
- Jiang, T. 2012. Empirical distributions of Laplacian matrices of large dilute random graphs. *Random Matrices: Theory and Applications*, 1(03): 1250004.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.
- Lim, D.; Robinson, J. D.; Zhao, L.; Smidt, T.; Sra, S.; Maron, H.; and Jegelka, S. 2023. Sign and Basis Invariant Networks for Spectral Graph Representation Learning. In *The Eleventh International Conference on Learning Representations*.
- Lin, L.; Chen, J.; and Wang, H. 2023. Spectral Augmentation for Self-Supervised Learning on Graphs. In *The Eleventh International Conference on Learning Representations*.
- Mehatari, R.; and Banerjee, A. 2015. Effect on normalized graph Laplacian spectrum by motif attachment and duplication. *Applied Mathematics and Computation*, 261: 382–387.
- Pei, H.; Wei, B.; Chang, K. C.-C.; Lei, Y.; and Yang, B. 2020. Geom-GCN: Geometric Graph Convolutional Networks. In *International Conference on Learning Representations*.
- Rozemberczki, B.; Allen, C.; and Sarkar, R. 2021. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2): cnab014.
- Shchur, O.; Mumme, M.; Bojchevski, A.; and Günnemann, S. 2018. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2019. Graph Attention Networks. In *International Conference on Learning Representations*.
- Wang, X.; and Zhang, M. 2022. How powerful are spectral graph neural networks. In *International Conference on Machine Learning*, 23341–23362. PMLR.
- Wang, Y.; Wang, J.; Cao, Z.; and Barati Farimani, A. 2022. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3): 279–287.
- Weisfeiler, B.; and Leman, A. 1968. The reduction of a graph to canonical form and the algebra which appears therein. *nti, Series*, 2(9): 12–16.
- Wu, G.; Lin, S.; Shao, X.; Zhang, P.; and Qiao, J. 2023. QPGCN: Graph convolutional network with a quadratic polynomial filter for overcoming over-smoothing. *Applied Intelligence*, 53(6): 7216–7231.



- Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2022. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5): 1–37.
- Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations*.
- Yang, Z.; Cohen, W.; and Salakhudinov, R. 2016. Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning*, 40–48. PMLR.
- Yao, S.; Yu, D.; and Jiao, X. 2021. Perturbing eigenvalues with residual learning in graph convolutional neural networks. In *Asian Conference on Machine Learning*, 1569–1584. PMLR.
- 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 2021*, 1215–1226.