

Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering

Jingwei Guo, Kaizhu Huang*, Xinping Yi, Zixian Su, and Rui Zhang

Abstract—Whilst spectral Graph Neural Networks (GNNs) are theoretically well-founded in the spectral domain, their practical reliance on polynomial approximation implies a profound linkage to the spatial domain. As previous studies rarely examine spectral GNNs from the spatial perspective, their spatial-domain interpretability remains elusive, *e.g.*, what information is essentially encoded by spectral GNNs in the spatial domain? In this paper, to answer this question, we investigate the theoretical connection between spectral filtering and spatial aggregation, unveiling an intrinsic interaction that spectral filtering implicitly leads the original graph to an adapted new graph, explicitly computed for spatial aggregation. Both theoretical and empirical investigations reveal that the adapted new graph not only exhibits non-locality but also accommodates signed edge weights to reflect label consistency among nodes. These findings highlight the interpretable role of spectral GNNs in the spatial domain and inspire us to rethink graph spectral filters beyond the fixed-order polynomials, which limit the effective propagation range and hinder their ability to capture long-range dependencies. Built upon the theoretical findings, we revisit the state-of-the-art spectral GNNs and propose a novel Spatially Adaptive Filtering (SAF) framework, which leverages the adapted new graph by spectral filtering for an auxiliary non-local aggregation. Notably, our SAF comprehensively models both node similarity and dissimilarity from a global perspective, therefore alleviating persistent deficiencies of GNNs related to long-range dependencies and graph heterophily. Extensive experiments over 13 node classification benchmarks demonstrate the superiority of our proposed framework to the state-of-the-art methods.

Index Terms—Graph Neural Networks, Spectral Filtering, Long-range Dependency, Graph Heterophily

I. INTRODUCTION

Graph Neural Networks (GNNs) have shown remarkable abilities to uncover the intricate dependencies within graph-structured data, and achieved tremendous success in graph machine learning [1]–[6]. Spectral GNNs are a class of GNNs rooted in spectral graph theory [7], [8], implementing graph convolutions via spectral filters [9], [10]. Whilst various spectral filtering strategies [11]–[19] have been proposed for spectral GNNs, their practical implementations invariably resort to approximating graph filters with fixed-order polynomials for computational efficiency [15], [16]. This truncated approach essentially relies on the direct extraction of spatial features

from the local regions of nodes. As such, the spatial domain of a graph, albeit loosely connected to spectral GNNs in theory, still plays a crucial role in effectively learning node representations.

However, there is a notable lack of research examining spectral GNNs from the spatial perspective. Though recent studies analyze both spectral and spatial GNNs to elucidate their similarities in model formulations [20]–[22], outcomes [23], [24], and expressiveness [16], [25]–[27], they neglect the interpretability that could arise mutually from the other domain. Specifically, while most spectral GNNs have well explained their learned filters in the spectral domain [13], [15]–[17], understandings from the spatial viewpoint are merely limited to fusing multi-scale graph information [28]; this unfortunately lacks a deeper level of interpretability in the vertex domain. Therefore, a natural question arises: *what information is essentially encoded by spectral GNNs in the spatial domain?*

In this work, we attempt to answer this question by exploring the connection between spectral filtering and spatial aggregation. The former is the key component in spectral GNNs, while the latter is closely associated with spatial GNNs utilizing recursive neighborhood aggregation. In existing GNN frameworks, these two approaches rarely interact each other at the risk of domain information trade-offs due to uncertainty principles [29]–[31]. Recognizing the spatial significance in spectral filtering, [13] have recently considered non-negative constraints as part of a generalized graph optimization problem. Notably, however, spatial aggregation meanwhile resembles the optimizing trajectory of the same optimization problem through iterative steps, which may be easily overlooked. Inspired by such observation, we examine, for the first time, the theoretical interaction between spectral filtering and spatial aggregation. This exploration has led us to uncover an intriguing theoretical interplay, *i.e.*, spectral filtering implicitly modifies the original graph, transforming it into a new one that explicitly functions as a computation graph for spatial aggregation. Delving deeper, we discover that the adapted new graph enjoys some desirable properties, enabling a direct link among nodes that originally require multiple hops to do so, thereby exhibiting nice non-locality. Moreover, we find that the new graph edges allow signed weights, which turns out capable of distinguishing between label agreement and disagreement of the connected nodes.

Overall, these findings underscore the interpretable role and significance of spectral GNNs in the spatial domain, inspiring us to rethink graph spectral filters beyond fixed-order polynomials, which, albeit efficient, limit models’ effective propagation range and hinder their ability to capture long-

J. Guo is with University of Liverpool, Liverpool, UK and is also with Xi’an Jiaotong-Liverpool University, Suzhou, China (E-mail: jingweigu01@outlook.com); K. Huang is with Duke Kunshan University, Suzhou, China (E-mail: kaizhu.huang@dukekunshan.edu.cn); X. Yi is with Southeast University, Nanjing, China (E-mail: xyi@seu.edu.cn); Z. Su is with University of Liverpool, Liverpool, UK and is also with Xi’an Jiaotong-Liverpool University, Suzhou, China (E-mail: zixian.su@liverpool.ac.uk); R. Zhang is with Xi’an Jiaotong-Liverpool University, Suzhou, China (E-mail: rui.zhang02@xjtlu.edu.cn); *Corresponding author: Kaizhu Huang.

range dependencies. Concretely, we propose a novel Spatially Adaptive Filtering (SAF) framework to fully explore spectral GNNs in the spatial domain. SAF leverages the adapted new graph by spectral filtering for auxiliary spatial aggregation and allows individual nodes to flexibly balance between spectral and spatial features. By performing non-local aggregation with signed edge weights, SAF adeptly overcomes the limitations of truncated polynomials, enabling the model to capture both node similarity and dissimilarity at a global scale. As a benefit, it can mitigate persistent deficiencies of GNNs regarding long-range dependencies and graph heterophily. The contributions are summarized as follows:

- Our investigation into spectral GNNs in the spatial domain reveals that they fundamentally alters the original graph, introducing non-locality and signed edge weights to discern node label consistency.
- We propose Spatially Adaptive Filtering (SAF) framework, a paradigm-shifting approach to spectral GNNs that jointly leverages graph learning in both spatial and spectral domains, making it a powerful tool for capturing long-range dependencies and handling graph heterophily.
- Extensive experiments in node classification exhibit notable improvements of up to 15.37%, and show that SAF beats the best-performing spectral GNNs on average.

II. MOTIVATION AND RELATED WORKS

This section outlines the motivation behind our research, derived from thoughtful consideration of the existing related works, and elucidates how our work diverges from and contributes to the current body of research.

A. Graph Neural Networks

GNNs can be broadly divided into spatial-based and spectral-based methods. Spatial GNNs leverage the spatial connections among nodes to perform message passing, also known as spatial aggregation [32], [33] (readers are directed to works [2], [34] for a thorough review). Spectral GNNs leverage the graph’s spectral domain for convolution or, alternatively, spectral filtering [10], [35]. Prevailing approaches focus on developing polynomial graph filters, by either learning polynomial coefficients, such as GPR-GNN [12], BernNet [13], ChebNetII [15], and JacobiConv [16], or optimizing the polynomial basis for better adaption, as seen in models like LON-GNN [18] and OptBasisGNN [19]. Diverging from this trend, ARMA [36] employs rational filter functions while still approximating them with polynomials. Although these methods are theoretically grounded in the spectral domain, their practical reliance on polynomial approximation hints at a profound linkage to the spatial domain. However, the spatial-domain interpretation of spectral GNNs is rarely examined. To this end, we delve into in this paper the intrinsic information spectral GNNs convey within the spatial context.

B. Unified Viewpoints for GNNs.

Several works have explored the nuances between spatial and spectral GNNs. Early studies by [20] and [21] examined

their similarities in model formulations. [25] proved their spatial GNN’s anti-oversmoothing ability via spectral analysis. [23] and [24] utilized the graph signal denoising problem to integrate both GNN types, and their expressiveness equivalence is further explored in works [16], [26]. Recently, [27] have highlighted the feature space constraints of both spatial and spectral GNNs, while [22] tend to emphasize their relationship via residual connection. Though these studies effectively bridge spectral and spatial GNNs, they remain focused on congruencies. Unlike them, our work represents the first endeavor to delve into the interpretability of spectral GNNs in the spatial domain, emphasizing the theoretical synergy between spectral filtering and spatial aggregation. The empirical success of our method (as compared to unified GNNs in Tables II and III), stem from this in-depth analysis, further underscoring our practical contributions to the literature.

C. Long-range Dependencies

While substantial efforts have been directed towards capturing long-range dependencies in spatial GNNs [25], [37]–[42], the exploration of the same challenge in spectral GNNs remains under-studied. Specifically, most spectral GNNs approximate graph filters with fixed-order polynomials, which, albeit efficient, limits the effective propagation range and hinder their ability to capture the long-range dependencies. To fill this gap, we propose SAF framework that emerges as a valuable consequence of analyzing spectral GNNs in the spatial domain. Aligned with our objective, Specformer [43] is introduced to addresses long-range dependencies for spectral GNNs, using a Transformer based set-to-set spectral filter. However, it lacks spatial-domain interpretability and introduce more trainable parameters. In contrast, our approach creates a non-local new graph without learning additional parameters, meanwhile elucidating the spatial implications of spectral GNNs. Similarly, a recent approach, FLODE [44], also produces matrices capable of capturing long-range dependencies by utilizing the fractional graph Laplacian [45]. However, this work mainly generalizes existing concepts from undirected to directed graphs to mitigate graph oversmoothing. Our study, conversely, focuses on exploring the fundamental issues surrounding undirected graphs to delve deeper into the intrinsic significance of spectral GNNs.

D. Graph Heterophily

Graph heterophily [39], [46], where different labeled nodes connect, challenges GNNs operating under the homophily assumption [47]. Although many GNNs have been crafted to manage heterophilic connections [12], [48]–[55], our method offers a distinct solution. Specifically, SAF innovatively conducts an auxiliary non-local aggregation using signed edge weights, emphasizing both intra-class similarity and inter-class difference on a global scale. One should note that there are two recent works [56], [57] also employ signed edge weights, introducing GloGNN and LRGNN, respectively. GloGNN aims to capture global homophily but is limited to K-hop neighborhood information, while LRGNN extends this by using low-rank properties to approximate the true global relationships

between node labels. Despite their success, both methods focus primarily on fitting node label relationships without fully exploring the fundamental GNN mechanisms, leading to a heavy reliance on label supervision (Section VI-I). Differently, our SAF leverages cross-domain insights into GNNs, ensuring the theoretical soundness of non-local learning, as proven in Section IV-B. This allows SAF to consistently perform well under both sparse and dense supervision.

III. NOTATIONS AND PRELIMINARIES

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with node set \mathcal{V} and edge set \mathcal{E} , where the number of nodes is denoted by N . The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ defines the edge weights $A_{i,j}$ between nodes v_i and v_j , and the degree matrix \mathbf{D} can be obtained by summing the rows of \mathbf{A} into a diagonal matrix. We denote the graph Laplacian matrix as \mathbf{L} , which is often normalized into $\hat{\mathbf{L}} = \mathbf{I} - \hat{\mathbf{A}}$ with $\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ and an identity matrix \mathbf{I} . Let $\hat{\mathbf{L}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ be the Laplacian eigendecomposition, where \mathbf{U} is eigenvector matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ consists of eigenvalues within $[0, 2]$. For node classification, we usually have a feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$ with F being raw feature dimensions, and each node is assigned a one-hot label vector $\mathbf{y}_i \in \mathbb{R}^C$ where $C \leq N$ is class number.

A. Spectral Filtering

Spectral filtering is essential in spectral GNNs. It selectively shrinks or amplifies the Fourier coefficients of node features [9] for learning tasks and usually take the form as

$$\mathbf{Z} = g_\psi(\hat{\mathbf{L}}) \mathbf{X} = \mathbf{U} g_\psi(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{X}. \quad (1)$$

Here, $g_\psi : [0, 2] \rightarrow \mathbb{R}$ defines a graph filter function, which are often approximated by a K -order polynomial in practice. Specifically, we have $g_\psi(\lambda) = \sum_{k=0}^K \psi_k P_k(\lambda) = \sum_{k=0}^K \omega_k \lambda^k$ where $P_k : [0, 2] \rightarrow \mathbb{R}$ refers to a polynomial basis and both ψ_k and ω_k denote the polynomial coefficient.

B. Spatial Aggregation

Spatial aggregation is a central component of spatial GNNs, facilitating the propagation of node information along edges and its subsequent aggregation within node neighborhood. To provide a better illustration, let's consider a popular spatial GNN, APPNP [37]. This model begins with a feature transformation $-\mathbf{Z}^{(0)} = f(\mathbf{X})$, and then perform the spatial aggregation as:

$$\mathbf{Z}^{(k)} = (1 - \eta) \mathbf{Z}^{(0)} + \eta \hat{\mathbf{A}} \mathbf{Z}^{(k-1)}, \quad k = 1, 2, \dots, K, \quad (2)$$

where $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and η refers to the update rate.

IV. RETHINKING SPECTRAL GRAPH NEURAL NETWORKS FROM THE SPATIAL PERSPECTIVE

In this section, we provide both theoretical and empirical analyses to examine spectral GNNs from the spatial perspective and answer the question, *i.e.*, what information is essentially encoded by spectral GNNs in the spatial domain?

A. Interplay of Spectral and Spatial Domains through the Lens of Graph Optimization

The graph signal denoising problem [8] was initially leveraged in [23], [24] as a means to interpret GNNs with smoothness assumption, which yet does not always hold in certain real-world graph scenarios such as heterophily [46]. Without loss of generality, in this work, we consider a more generalized graph optimization problem¹

$$\arg \min_{\mathbf{Z}} \mathcal{L} = \alpha \|\mathbf{X} - \mathbf{Z}\|_F^2 + (1 - \alpha) \cdot \text{tr}(\mathbf{Z}^T \gamma_\theta(\hat{\mathbf{L}}) \mathbf{Z}) \quad (3)$$

where $\mathbf{Z} \in \mathbb{R}^{N \times d}$ refers to node representations, $\gamma_\theta(\hat{\mathbf{L}})$ determines the rate of propagation [58] by operating on the graph spectrum, *i.e.*, $\gamma_\theta(\hat{\mathbf{L}}) = \mathbf{U} \gamma_\theta(\mathbf{\Lambda}) \mathbf{U}^T$, and $\alpha \in (0, 1)$ is a trade-off coefficient. In case of setting $\gamma_\theta(\hat{\mathbf{L}}) = \hat{\mathbf{L}}$, Eq. (3) turns into the well-known graph signal denoising problem. To ensure the convexity of the objective in Eq. (3), a positive semi-definite constraint is imposed on $\gamma_\theta(\hat{\mathbf{L}})$, *i.e.*, $\gamma_\theta(\lambda) \geq 0$ for $\lambda \in [0, 2]$. Then, one can address this minimization problem through either closed-form or iterative solutions.

1) *Closed-form Solution*: The closed-form solution can be obtained by setting the derivative of the objective function \mathcal{L} to 0, *i.e.*, $\frac{\partial \mathcal{L}}{\partial \mathbf{Z}} = 2\alpha(\mathbf{Z} - \mathbf{X}) + 2(1 - \alpha)\gamma_\theta(\hat{\mathbf{L}})\mathbf{Z} = 0$. Let $g_\psi(\lambda) = (1 + \frac{1-\alpha}{\alpha}\gamma_\theta(\lambda))^{-1}$, we can observe that the closed-form solution in Eq. (4) is equivalent to the spectral filtering in Eq. (1).

$$\mathbf{Z}^* = (\mathbf{I} + \frac{1-\alpha}{\alpha}\gamma_\theta(\hat{\mathbf{L}}))^{-1} \mathbf{X} = g_\psi(\hat{\mathbf{L}}) \mathbf{X} \quad (4)$$

As $\gamma_\theta(\lambda) \geq 0$, this establishes a more stringent constraint for the graph filter in spectral GNNs, *i.e.*, $0 < g_\psi(\lambda) \leq \frac{\alpha}{\alpha + (1-\alpha) \cdot 0} = 1$, which is termed as a non-negative constraint in this paper.

2) *Iterative Solution*: Alternatively, we can take an iterative gradient descent method such that $\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)} - b \frac{\partial \mathcal{L}}{\partial \mathbf{Z}}|_{\mathbf{Z}=\mathbf{Z}^{(k-1)}}$ with a step size $b = \frac{1}{2}$, which yields a concise iterative solution in Eq. (5) with $\hat{\mathbf{A}}^{\text{new}} = \mathbf{I} - \gamma_\theta(\hat{\mathbf{L}})$. Notably, by taking $\hat{\mathbf{A}}^{\text{new}}$ as a new computation graph, this solution closely mirrors the spatial aggregation in Eq. (2).

$$\mathbf{Z}^{(k)} = \alpha \mathbf{X} + (1 - \alpha) \hat{\mathbf{A}}^{\text{new}} \mathbf{Z}^{(k-1)}, \quad k = 1, 2, \dots, K \quad (5)$$

3) *Theoretical Interaction — the Adapted New Graph*: With the non-negative constraint, it is evident that both spectral filtering and spatial aggregation effectively address the generalized graph optimization problem in Eq. (3), despite their distinctive forms and operation domains. Upon closer examination, we discover a compelling relationship between the graph filter $g_\psi(\lambda)$ in Eq. (4) and the new graph $\hat{\mathbf{A}}^{\text{new}}$ in Eq. (5), given $g_\psi(\lambda) = (1 + \frac{1-\alpha}{\alpha}\gamma_\theta(\lambda))^{-1}$,

$$\hat{\mathbf{A}}^{\text{new}} = \mathbf{I} - \gamma_\theta(\hat{\mathbf{L}}) = \mathbf{I} - \frac{\alpha}{1-\alpha} (g_\psi(\hat{\mathbf{L}})^{-1} - \mathbf{I}) \quad (6)$$

which unveils an intrinsic inter-play, *i.e.*, spectral filtering implicitly leads the original graph to an adapted new graph, explicitly computed for spatial aggregation.

¹This problem was first introduced in [13] for theoretically grounded graph filters. However, in this study, we repurpose it as a bridge between spectral filtering and spatial aggregation.

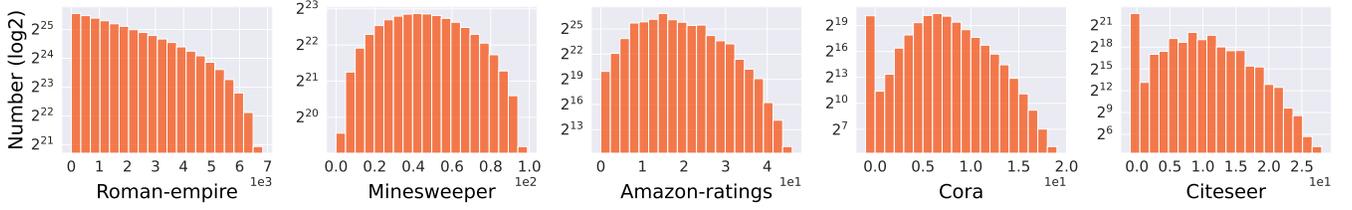


Fig. 1. Distributions of connected nodes in the new graph based on their geodesic/shortest-path distance (as $\Delta_{i,j}$) in the original graph. Nodes, distant in the original graph ($\Delta_{i,j} > 1$ in x-axis), can be linked in the new graph (Number > 0 in y-axis).

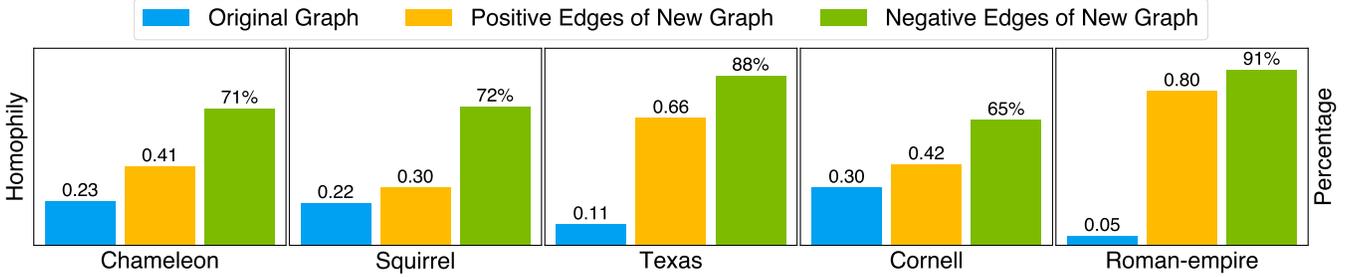


Fig. 2. Left y-axis: Homophily comparison between original and new graphs, considering only positive edges (blue and yellow bars). Right y-axis: Percentage of negative edges in the new graph that connect nodes from different classes (green bar).

4) What are the differences between $\hat{\mathbf{A}}^{\text{new}}$ and $g_\psi(\hat{\mathbf{L}})$:

Whereas the former as the uncovered new graph elucidates the inherent spatial node relationships, the latter is a operation in spectral GNNs that processes graph features within spectral domain. It is crucial to understand that $g_\psi(\hat{\mathbf{L}})$ may not result in a dense matrix, especially with fixed-order polynomial approximation. This is because it captures up to only a K -hop neighborhood, i.e., $g_\psi(\hat{\mathbf{L}}) = \sum_{k=0}^K \psi_k P_k(\hat{\mathbf{L}}) = \sum_{k=0}^K \omega_k \hat{\mathbf{A}}^k$, practically limiting spectral GNNs' effective propagation range. In contrast, our newfound graph $\hat{\mathbf{A}}^{\text{new}}$ intrinsically enjoys a non-local property, as confirmed in the following section. Building upon this discovery, we further devise a framework to transcend domain barriers, overcoming the limitations of current spectral GNNs due to truncated polynomials (see details in Section V).

B. In-depth Analysis of the Adapted New Graph

To deepen our understanding of the interpretability produced by spectral GNNs in the spatial domain, we embark upon a blend of theoretical and empirical inquiries into the adapted new graph.

1) *Non-locality*: Our examination of the adapted new graph illuminates its non-local nature, particularly evident in the infinite series expansion of the original graph's adjacency matrix. To elucidate, we first introduce an pivotal mathematical construct, the Neumann series, in the following lemma.

Lemma 1. Let $\mathbf{M} \in \mathbb{R}^{N \times N}$ be a matrix with eigenvalues λ_n , if $|\lambda_n| < 1$ for all $n = 1, 2, \dots, N$, then $(\mathbf{I} - \mathbf{M})^{-1}$ exists and can be expanded as an infinite series, i.e., $(\mathbf{I} - \mathbf{M})^{-1} = \sum_{t=0}^{\infty} \mathbf{M}^t$, which is known as Neumann series.

With the established non-negative constraint on graph filters, specifically $0 < g_\psi(\lambda) \leq 1$, it becomes evident that the eigenvalues of $\mathbf{I} - g_\psi(\hat{\mathbf{L}})$ falls into the interval permitting Neumann

series expansion, as shown in lemma 1 [59]. Building on this observation, we present a non-trivial property of the new graph in the following proposition.

Proposition 1. Given adjacency matrix $\hat{\mathbf{A}}^{\text{new}}$ formulated in Eq. (6), the adapted new graph exhibits non-locality. Specifically, $\hat{\mathbf{A}}^{\text{new}}$ is expressible as an infinite series expansion of the original graph's adjacency matrix $\hat{\mathbf{A}}$. Formally, we have $\hat{\mathbf{A}}^{\text{new}} = \mathbf{I} - \frac{\alpha}{1-\alpha} \sum_{t=1}^{\infty} (\mathbf{I} - \sum_{k=0}^K \pi_k \hat{\mathbf{A}}^k)^t = \sum_{t=0}^{\infty} \phi_t \hat{\mathbf{A}}^t$ where π_k and ϕ_t refer to the constant coefficients computed from $\{\psi_0, \psi_1, \dots, \psi_K\}$ in distinct ways.

Proof. We begin with the assertion that the eigenvalues of $\mathbf{I} - g_\psi(\hat{\mathbf{L}})$ are positive and strictly less than 1, which fulfills the necessary condition for the Neumann series expansion stated in Lemma 1. As such, we can deduce $g_\psi(\hat{\mathbf{L}})^{-1} = (\mathbf{I} - (\mathbf{I} - g_\psi(\hat{\mathbf{L}})))^{-1} = \sum_{t=0}^{\infty} (\mathbf{I} - g_\psi(\hat{\mathbf{L}}))^t$. Owing to the prevalent polynomial approximation, we are eligible to express $g_\psi(\hat{\mathbf{L}})$ w.r.t. adjacency matrix $\hat{\mathbf{A}}$, i.e., $g_\psi(\hat{\mathbf{L}}) = g_\psi(\mathbf{I} - \hat{\mathbf{A}}) = \sum_{k=0}^K \pi_k \hat{\mathbf{A}}^k$ where π_k refers to the new coefficients made of up $\{\psi_m\}_{m=0}^K$. Substituting this polynomial representation into our Neumann expansion, we obtain $g_\psi(\hat{\mathbf{L}})^{-1} = \sum_{n=0}^{\infty} (\mathbf{I} - \sum_{k=0}^K \pi_k \hat{\mathbf{A}}^k)^t$. Now, revisiting $\hat{\mathbf{A}}^{\text{new}}$ in Eq. (6), we have $\hat{\mathbf{A}}^{\text{new}} = \mathbf{I} - \frac{\alpha}{1-\alpha} (g_\psi(\hat{\mathbf{L}})^{-1} - \mathbf{I}) = \mathbf{I} - \frac{\alpha}{1-\alpha} \sum_{t=1}^{\infty} (\mathbf{I} - \sum_{k=0}^K \pi_k \hat{\mathbf{A}}^k)^t = \sum_{t=0}^{\infty} \phi_t \hat{\mathbf{A}}^t$ where ϕ_t is a constant coefficient made up of $\{\pi_m\}_{m=0}^K$. \square

This proposition implies that the new graph engenders immediate links between nodes that originally necessitate multiple hops for connection. To further underpin this theoretical claim, we analyze the general connection status on the new graph by BernNet [13], a spectral GNN adhering to the non-negative constraint. From Fig. 1, it is apparent that nodes originally separated by multiple hops achieve direct connections in the new graph.

2) Signed Edge Weights — Discerning Label Consistency:

Upon further scrutinizing the adapted new graph, we make a notable discovery that it readily accommodates both positive and negative edge weights. A more granular analysis in Fig. 2 reveals that a considerable portion of positive edge weights are assigned to the same-class node pairs, enhancing graph homophily (exemplified by edge homophily ratio [46]). Conversely, edges parameterized with negative weights tend to bridge nodes with different labels. These findings demonstrate the newfound graph’s adeptness in discerning label consistency among nodes. To theoretically explain this phenomenon, we further present the proposition below:

Proposition 2. *Let \mathbf{Z}^* be the node representations optimized by Eq. (3). For \mathbf{Z}^* to be effective in label prediction, it is a necessary condition that $\hat{\mathbf{A}}^{\text{new}}$ accommodates both positive and negative edge weights s.t. for any node pairs $v_i, v_j \in \mathcal{V}$, $\hat{A}_{i,j}^{\text{new}} > 0$ if $\mathbf{y}_i = \mathbf{y}_j$ and $\hat{A}_{i,j}^{\text{new}} < 0$ if $\mathbf{y}_i \neq \mathbf{y}_j$.*

Proof. Let us commence the proof by contradiction. Let \mathcal{C} denote the condition described in proposition 2. Assume, for the sake of contradiction, that \mathcal{C} is not requisite for the optimal node representations \mathbf{Z}^* to be predictive of node labels. Under this assumption, there are node pairs $v_i, v_j \in \mathcal{V}$ such that: (1) if $\mathbf{y}_i = \mathbf{y}_j$, $\hat{A}_{i,j}^{\text{new}} < 0$; (2) if $\mathbf{y}_i \neq \mathbf{y}_j$, $\hat{A}_{i,j}^{\text{new}} > 0$. Without loss of generality, given the non-locality as proved in proposition 1, we exclude cases where $\hat{A}_{i,j}^{\text{new}} = 0$ from our consideration. Now, consider the second objective term $\text{tr}(\mathbf{Z}^T \gamma_\theta(\hat{\mathbf{L}}) \mathbf{Z})$ in Eq. (3). Using the relationship $\gamma_\theta(\hat{\mathbf{L}}) = \mathbf{I} - \hat{\mathbf{A}}^{\text{new}}$, we can expand this term into $\sum_{v_i, v_j \in \mathcal{V}} \hat{A}_{i,j}^{\text{new}} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2$. Under (1), for same-class nodes v_i, v_j with $\hat{A}_{i,j}^{\text{new}} < 0$, minimizing the objective term pulls \mathbf{Z}_i and \mathbf{Z}_j apart in the latent space. This behavior violates the canonical understanding that nodes from the same class should exhibit similar representations. Under (2), for different-class nodes v_i, v_j with $\hat{A}_{i,j}^{\text{new}} > 0$, the optimization encourages \mathbf{Z}_i and \mathbf{Z}_j to be more similar. This is in direct opposition to the basic classification principle that nodes from different classes should have distinct representations. Given these contradictions stemming from the mathematical implications in optimization, we must reject assumptions (1) and (2), affirming the necessary condition \mathcal{C} for accurate label prediction by \mathbf{Z}^* . \square

Proposition 2 provides a theoretical foundation of our empirical findings on the new graph. The essence lies in the objective in Eq. (3), particularly the trace term $\text{tr}(\mathbf{Z}^T \gamma_\theta(\hat{\mathbf{L}}) \mathbf{Z})$. For clarity, let us reinterpret this trace term as $\text{tr}(\bar{\mathbf{Z}}^T (\mathbf{D}^{\text{new}} - \mathbf{A}^{\text{new}}) \bar{\mathbf{Z}})$, where \mathbf{D}^{new} denotes the related degree matrix and $\bar{\mathbf{Z}}$ is derived from rescaling \mathbf{Z} . Clearly, this term evaluates label smoothness among adjacent nodes in the new graph, which, given its non-local nature, includes both intra-class ($=$) and inter-class (\neq) node connections such that $\mathbf{A}^{\text{new}} = \mathbf{A}_{=}^{\text{new}} + \mathbf{A}_{\neq}^{\text{new}}$. Drawing from proposition 2, we can further dissect the original trace term, splitting it into $\text{tr}(\bar{\mathbf{Z}}^T (\mathbf{D}_{=}^{\text{new}} - \mathbf{A}_{=}^{\text{new}}) \bar{\mathbf{Z}}) - \text{tr}(\bar{\mathbf{Z}}^T |(\mathbf{D}_{\neq}^{\text{new}} - \mathbf{A}_{\neq}^{\text{new}})| \bar{\mathbf{Z}})$ where the $|\cdot|$ operation denotes absolute values. As such, it becomes evident that minimizing this trace term not only enhances the representational proximity for same-class node pairs but also strengthens the distinctiveness for different-class nodes pairs. Such nuanced behaviors, inherent

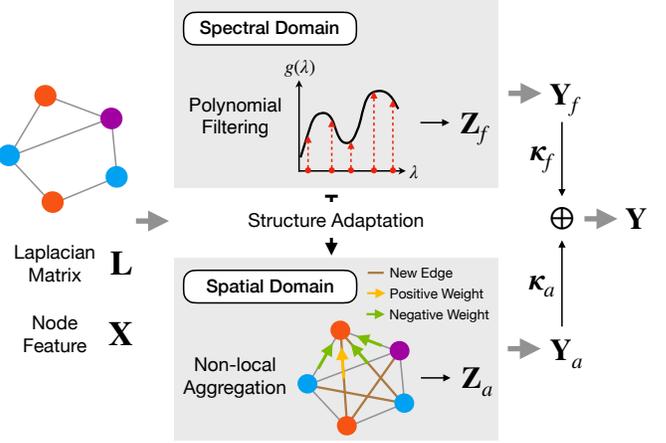


Fig. 3. Illustration of SAF framework where varying node colors represent different node labels. SAF leverages the adapted new graph by spectral filtering for auxiliary non-local aggregation in the spatial domain and allows individual nodes to flexibly balance between spectral and spatial features.

to the optimization in Eq. (3), are necessary for GNN models to achieve accurate label predictions.

3) *Summarization:* Our investigation into spectral GNNs in the spatial domain reveals that they fundamentally alters the original graph, imbuing it with non-locality and signed edge weights that capture label relationships among nodes. These findings highlight the interpretable role of spectral GNNs in the spatial domain and prompt us to rethink current models beyond the truncated polynomial filters.

V. SPATIALLY ADAPTIVE FILTERING FRAMEWORK

Building on our discoveries, we re-evaluate the state-of-the-art spectral GNNs and put forth a paradigm-shifting framework, Spatially Adaptive Filtering (SAF), for joint exploitation of graph-structured data across both spectral and spatial domains. SAF leverages the adapted new graph by spectral filtering for an auxiliary non-local aggregation, addressing enduring challenges in GNNs related to long-range dependencies and graph heterophily. See the overall pipeline in Fig. 3.

1) *Non-negative Spectral Filtering:* The proposed SAF requires explicit computation of the newfound graph, as outlined in Eq. (6). This further necessitates the graph filter $g_\psi : [0, 2] \rightarrow \mathbb{R}$ to satisfy the non-negative constraint from Eq. (3): $0 \leq g_\psi(\lambda) \leq 1$. However, not all extant graph filters fulfill this prerequisite. For instance, the filter use by GCN [10], $g_\psi(\lambda) = 1 - \lambda$, takes negative values when $\lambda > 1$. In this research, we approximate the graph filter using Bernstein polynomials [60], which are known for their non-negative traits [61] and are essential in a preeminent spectral GNN, BernNet [13]. For $g_\psi(\lambda) \leq 1$ part, we rescale Bernstein polynomials with the following proposition.

Proposition 3. *Let $B_{k,K}(x)$ denote the Bernstein polynomial basis of index k and order K , which is defined as $B_{k,K}(x) = \binom{K}{k} (1-x)^{K-k} x^k$ for $x \in [0, 1]$. Let ψ_k denote the k -th coefficient of a polynomial $p(x)$ of order K , where $p(x) = \sum_{k=0}^K \psi_k B_{k,K}(x)$ with $\psi_k \geq 0$ for all k . Then for all $x \in [0, 1]$, we have $g_\psi(x) \leq \max\{\psi_k\}_{k=0}^K$.*

Proof. We denote $p(x) = \sum_{k=0}^K \psi_k \binom{K}{k} (1-x)^{K-k} x^k$ as a Bernstein polynomial with $\psi_k \geq 0$ for all k and $\psi_{\max} = \max\{\psi_k\}_{k=0}^K$. Given $x \in [0, 1]$, we can derive the following inequality as

$$\begin{aligned} p(x) &= \sum_{k=0}^K \psi_k \binom{K}{k} (1-x)^{K-k} x^k \\ &\leq \psi_{\max} \sum_{k=0}^K \binom{K}{k} (1-x)^{K-k} x^k \\ &= \psi_{\max} (1-x+x)^K = \psi_{\max}. \end{aligned}$$

Therefore, we have $p(x) \leq \max\{\psi_k\}_{k=0}^K$ for all $x \in [0, 1]$. \square

Proposition 3 suggests that the Bernstein polynomial function attains its maximum value in $\psi_{\max} = \max\{\psi_k\}_{k=0}^K$. Therefore, $g_\psi(\lambda)$ can be rescaled within $[0, 1]$ by $\hat{g}_\psi(\lambda) = \frac{1}{\psi_{\max}} \sum_{k=0}^K \psi_k B_{k,K}(\frac{\lambda}{2})$, which enables us to formulate the spectral filtering in SAF as $\mathbf{Z}_f = \hat{g}_\psi(\hat{\mathbf{L}}) f_\varphi(\mathbf{X}) = \frac{1}{\psi_{\max}} \sum_{k=0}^K \psi_k \frac{1}{2^k} \binom{K}{k} (2\mathbf{I} - \hat{\mathbf{L}})^{K-k} \hat{\mathbf{L}}^k f_\varphi(\mathbf{X})$ where $f_\varphi(\cdot)$, a two-layer MLP, maps \mathbf{X} from F to C dimensions using 64 hidden units, and $\{\psi_k\}_{k=1}^K$ are non-negative learnable parameters. Note that SAF also permits alternative implementations such as using Chebyshev polynomials [9], [62] for graph filter learning, enhancing models like ChebNetII [15] (see details in Section VI-H).

2) *Non-local Spatial Aggregation:* Once acquiring a suitable spectral filter $\hat{g}_\psi(\lambda)$, we compute the adapted new graph as $\hat{\mathbf{A}}^{\text{new}} = \mathbf{I} - \tau(\mathbf{U}_m g_\psi(\mathbf{\Lambda}_m)^{-1} \mathbf{U}_m^T - \mathbf{I})$ by Eq. (6) where $\tau = \frac{\alpha}{1-\alpha}$ is a scaling parameter and a partial eigendecomposition [63] can be employed to produce a low-rank, robust structure for $\hat{\mathbf{A}}^{\text{new}}$ using only m extremal eigenvalues. Equipped with this newfound graph, we proceed to perform non-local aggregation: $\mathbf{Z}^{(l)} = (1-\eta)\mathbf{Z}^{(0)} + \eta\hat{\mathbf{A}}^{\text{new}}\mathbf{Z}^{(l-1)}$, where η refers to the update rate and $\mathbf{Z}^{(0)} = f_\varphi(\mathbf{X})$ and $l = 1, 2, \dots, L$ denote layer number. The iteratively aggregated results are denoted as \mathbf{Z}_a . Recognizing the potential noise from the non-local nature of $\hat{\mathbf{A}}^{\text{new}}$, we apply a sparsification technique, leveraging a positive threshold ϵ , and retain only essential elements outside the $[-\epsilon, \epsilon]$ interval. For clarity, this refined model is referred to as SAF- ϵ .

3) *Node-wise Prediction Amalgamation:* To leverage information from different graph domains, we employ an attention mechanism, allowing nodes to determine the importance of each space. This mechanism produces pairwise weights for a nuanced amalgamation during prediction. Specifically, the weight pair is computed as $\kappa_f = \text{Sigmoid}(\mathcal{P}_f(\mathbf{Z}_f))$, $\kappa_a = \text{Sigmoid}(\mathcal{P}_a(\mathbf{Z}_a))$ where $\kappa_f, \kappa_a \in \mathbb{R}^N$ contain the weights for each node, and $\mathcal{P}_f(\cdot)$ and $\mathcal{P}_a(\cdot)$ are two different mappings from \mathbb{R}^C to \mathbb{R} . For simplicity, we implement them using two one-layer MLPs. Given domain predictions $\mathbf{Y}_f, \mathbf{Y}_a \in \mathbb{R}^C$, the final model prediction is attained as $\mathbf{Y} = \text{diag}(\kappa_f) \cdot \mathbf{Y}_f + \text{diag}(\kappa_a) \cdot \mathbf{Y}_a$ where a normalization $[\kappa_f, \kappa_a] \leftarrow \frac{[\kappa_f, \kappa_a]}{\max\{\|\kappa_f, \kappa_a\|_1, \delta\}}$ is performed beforehand to maintain $\kappa_f + \kappa_a = \mathbf{1}$ with small value δ preventing zero division. Similar schemes can be founded in works [64]–[66].

TABLE I

STATISTICS OF REAL-WORLD DATASETS. F AND C DENOTES THE NUMBER OF FEATURES AND CLASSES. Δ REPRESENTS GRAPH DIAMETER REFERRING TO THE LONGEST GEODESIC DISTANCE BETWEEN NODES ON THE GRAPH; FOR PENN94, DUE TO MULTIPLE SUBGRAPHS, WE REPORT Δ OF THE LARGEST CONNECTED COMPONENT. WE ASSESS GRAPH HOMOPHILY USING THREE METRICS - EDGE HOMOPHILY [46] \mathcal{H} , CLASS HOMOPHILY [67] $\mathcal{H}_{\text{CLASS}}$, AND ADJUSTED HOMOPHILY [68] $\mathcal{H}_{\text{ADJUSTED}}$ - WHICH RANGES FROM 0 (HIGH HETEROPHILY) TO 1 (HIGH HOMOPHILY).

Dataset	$ \mathcal{V} $	$ \mathcal{E} $	F	C	Δ	\mathcal{H}	$\mathcal{H}_{\text{class}}$	$\mathcal{H}_{\text{adjusted}}$
Chameleon	2,227	36,101	2,325	5	11	0.23	0.06	0.03
Squirrel	5,201	217,073	2,089	5	10	0.22	0.03	0.01
Texas	183	309	1,703	5	8	0.11	0.00	-0.23
Cornell	183	295	1,703	5	8	0.30	0.05	-0.08
Actor	7,600	33,544	931	5	12	0.22	0.01	0.00
Cora	2,708	5,429	1,433	7	19	0.81	0.77	0.77
Citeseer	3,327	4,732	3,703	6	28	0.74	0.63	0.67
Pubmed	19,717	44,338	500	3	18	0.80	0.66	0.69
Minesweeper	10,000	39,402	7	2	99	0.68	0.01	0.01
Tolokers	11,758	519,000	10	2	11	0.59	0.18	0.09
Amazon-ratings	24,492	93,050	300	5	46	0.38	0.13	0.14
Roman-empire	22,662	32,927	300	18	6,824	0.05	0.02	-0.05
Penn94	41,554	1,362,229	5	2	8	0.47	0.05	0.02

4) *Complexity Analysis:* SAF augments spectral GNNs with non-local aggregation and node-wise amalgamation. The first part entails creating a new graph and information propagation. In SAF- ϵ , these two steps are separated, culminating in $\mathcal{O}(N^3 + N^2 + nnz(\hat{\mathbf{A}}^{\text{new}})d)$ complexity, where nnz denotes non-zero element count. Conversely, SAF, viewing non-local aggregation holistically, can reduce complexity to $\mathcal{O}(2dN^2 + dN)$ when $d \ll N$. For node-wise amalgamation, its parallelizable nature ensures computational efficiency. Our method also requires eigendecomposition precomputation, which, though naively complex at $\mathcal{O}(N^3)$, can be reduced to $\mathcal{O}(m^2 + nnz(\hat{\mathbf{L}})m)$ using Lanczos method [63] with $m \ll N$ iterative steps. The results are also reusable for both training and inference. We present empirical studies on both time and space overheads in Section VI-J.

5) *Practical Implications of Eigendecomposition:* Eigendecomposition breaks down a matrix into its eigenvalues and eigenvectors, offering insights into matrix properties, especially for the graph Laplacian. Despite computational demands, this technique has attracted surging interest in the graph learning community due to its theoretical richness, and it can be practically expedited for larger graphs using Lanczos [63] and Sparse Generalized Eigenvalue [69] algorithms. Recent advancements also underscore its value in various applications such as graph positional encoding [70]–[72], spectral graph convolution [28], graph domain adaptation [73], graph robustness [74], graph expressivity [27], [51], [75], [76], and graph transformers [43], [77]–[79]. In line with these developments, our SAF also utilizes eigendecomposition to explicitly create a new graph, enabling efficient non-local aggregation with signed weights to tackle long-range dependency and graph heterophily.

VI. EXPERIMENTS

A. Datasets and Experimental Setup

1) *Datasets:* We conduct experiments on 13 real-world datasets with detailed statistics provided in Table I. In certain

TABLE II
SEMI-SUPERVISED NODE CLASSIFICATION ACCURACY (%) \pm 95% CONFIDENCE INTERVAL.

Method	Cham.	Squi.	Texas	Corn.	Actor	Cora	Cite.	Pubm.
MLP	26.36 \pm 2.85	21.42 \pm 1.50	32.42 \pm 9.91	36.53 \pm 7.92	29.75 \pm 0.95	57.17 \pm 1.34	56.75 \pm 1.55	70.52 \pm 2.01
GCN	38.15 \pm 3.77	31.18 \pm 0.93	34.68 \pm 9.07	32.36 \pm 8.55	22.74 \pm 2.37	79.19 \pm 1.37	69.71 \pm 1.32	78.81 \pm 0.84
APPNP	32.73 \pm 2.31	24.50 \pm 0.89	34.79 \pm 10.11	34.85 \pm 9.71	29.74 \pm 1.04	82.39 \pm 0.68	69.79 \pm 0.92	79.97 \pm 1.58
ARMA	37.42 \pm 1.72	24.15 \pm 0.93	39.65 \pm 8.09	28.90 \pm 10.07	27.02 \pm 2.31	79.14 \pm 1.07	69.35 \pm 1.44	78.31 \pm 1.33
GPR-GNN	33.03 \pm 1.92	24.36 \pm 1.52	33.98 \pm 11.90	38.95 \pm 12.36	28.58 \pm 1.01	82.37 \pm 0.91	69.22 \pm 1.27	79.28 \pm 2.25
BernNet	27.32 \pm 4.04	22.37 \pm 0.98	43.01 \pm 7.45	39.42 \pm 9.59	29.87 \pm 0.78	82.17 \pm 0.86	69.44 \pm 0.97	79.48 \pm 1.47
ChebNetII	43.42\pm3.54	33.96\pm1.22	46.58 \pm 7.68	42.19 \pm 11.61	30.18 \pm 0.81	82.42 \pm 0.64	69.89 \pm 1.21	79.51 \pm 1.03
JacobiConv	36.67 \pm 1.63	29.38 \pm 0.71	48.50 \pm 5.90	43.01 \pm 11.92	31.69 \pm 0.71	82.93 \pm 0.55	70.25 \pm 1.02	79.53 \pm 1.28
Specformer	36.05 \pm 3.47	29.64 \pm 0.88	50.00 \pm 8.33	43.76 \pm 5.84	31.45 \pm 0.68	81.44 \pm 0.63	66.11 \pm 0.98	78.05 \pm 1.03
LON-GNN	35.17 \pm 1.85	30.25 \pm 1.04	45.38 \pm 7.92	<u>35.32\pm8.09</u>	31.51 \pm 1.23	81.93 \pm 0.74	70.41 \pm 1.10	79.57 \pm 1.08
OptBasisGNN	35.56 \pm 2.86	31.25 \pm 1.06	37.11 \pm 5.09	32.31 \pm 7.11	31.73 \pm 0.50	78.69 \pm 0.86	63.46 \pm 1.30	77.38 \pm 0.98
FLODE	40.20 \pm 1.02	31.99 \pm 1.05	<u>50.29\pm4.74</u>	42.89 \pm 7.69	<u>32.18\pm1.10</u>	79.90 \pm 1.07	69.89 \pm 2.03	77.78 \pm 1.47
GNN-LF	26.49 \pm 2.00	22.01 \pm 1.04	39.02 \pm 6.24	36.65 \pm 9.60	28.28 \pm 0.71	81.96 \pm 0.92	69.80 \pm 1.36	79.50 \pm 1.28
GNN-HF	35.57 \pm 2.26	22.36 \pm 1.26	44.80 \pm 5.67	38.79 \pm 11.62	29.15 \pm 0.78	81.15 \pm 0.78	69.68 \pm 0.73	79.10 \pm 1.19
ADA-UGNN	39.39 \pm 2.02	25.65 \pm 0.49	47.86 \pm 6.65	42.89 \pm 8.09	30.78 \pm 1.00	82.52 \pm 1.04	70.18 \pm 1.40	79.78 \pm 1.32
FE-GNN	38.23 \pm 1.66	31.67 \pm 1.60	47.40 \pm 5.90	41.21 \pm 8.96	26.20 \pm 0.76	77.00 \pm 0.74	61.24 \pm 1.26	75.63 \pm 1.33
SAF	41.82 \pm 1.74	31.77 \pm 0.69	58.04 \pm 3.76	52.49 \pm 8.56	33.50 \pm 0.55	83.57 \pm 0.66	71.07 \pm 1.08	79.51 \pm 1.12
SAF- ϵ	<u>41.88\pm2.04</u>	<u>32.05\pm0.40</u>	58.38\pm3.47	53.41\pm5.55	33.84\pm0.58	83.79\pm0.71	71.30\pm0.93	80.16\pm1.25
Improv. ³	14.56%	9.68%	15.37%	13.99%	3.97%	1.62%	1.86%	0.68%

compact sections of this paper, we use four-letter abbreviations for dataset names.

- Chameleon and Squirrel are Wikipedia networks collected by [80] where nodes are web pages connected by mutual links. We utilize the labels from [39], dividing nodes into five classes based on their average monthly traffic.
- Texas and Cornell are webpage datasets from WebKB project, containing nodes from five classes (student, project, course, staff, and faculty) and connected by hyperlinks. This study uses the preprocessed version by [39].
- Actor is an actor co-occurrence network induced from the film-director-actor-writer network [81] by [39]. In this network, nodes are actors divided into five classes, and edges represent actor co-occurrence on Wikipedia pages.
- Cora, Citeseer, and Pubmed are citation network benchmarks [82], [83] with nodes as scientific papers and edges as undirected citations. Each node is assigned with a class label by research topic and bag-of-word features.
- Minesweeper, Tolokers, Amazon-ratings, and Roman-empire are recently proposed benchmarks by [84] for specifically evaluating GNN models under heterophily.
- Penn94 is a large-scale friendship network [85] from the Facebook 100 networks, where nodes denote students labeled by reported genders and possess features such as major, second major/minor, dorm/house, year, and high school. This work uses the version preprocessed by [67].

2) *Baselines*: We compare SAF with 21 models: (1) MLP; (2) Basic GNNs: GCN [10] and APPNP [37]; (3) Spectral GNNs: ARMA [36], GPR-GNN [12], BernNet [13], ChebNetII [15], JacobiConv [16], Specformer [43], LON-GNN [18], OptBasisGNN [19] and FLODE [44]; (4) Spatial GNNs: GCNII [25], PDE-GCN [38], NodeFormer [41], GloGNN++ [56] and LRGNN [86]; (5) Unified GNNs: GNN-LF [24], GNN-HF [24], ADA-UGNN [23] and FE-GNN [27].

3) *Implementation Details*: To follow [13], [15], [16], we fix $K = 10$. For each dataset, we perform a grid search to tune the hyper-parameters of all models. With the best hyper-parameters, we train models with Adam optimizer [87] in 1,000 epochs using early-stopping strategy and a patience of 200 epochs, and report the mean classification accuracies with a 95% confidence interval on 10 random data splits. As [15] have made a thorough evaluation and share the same experimental protocol with us, we leverage their results for models: MLP, GCN, APPNP, ARMA, GPR-GNN, BernNet, ChebNetII, GCNII and PDE-GCN on datasets – Chameleon, Squirrel, Texas, Cornell, Actor, Cora, Citeseer, and Pubmed. For JacobiConv, LON-GNN, and OptBasisGNN, we also report the results from their papers [16], [18], [19]. Our codes will be made available if the paper could be accepted.

4) *Hyper-parameters Setting*: We perform a grid search on the hyper-parameters of all models for each dataset using Optuna [88] To accommodate extensive experiments across diverse datasets in both semi- and full-supervised setting, we define a broad searching space as: learning rate $lr \sim \{1e-3, 5e-3, 1e-2, 5e-2, 0.1\}$, weight decay $L_2 \sim \{0.0, 1e-6, 5e-6, 1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2\}$, dropout $\sim \{0.0, 0.1, \dots, 0.8\}$ with step 0.1, non-local aggregation step $L \sim \{1, 2, \dots, 10\}$ with step 1, scaling parameter $\tau \sim \{0.1, 0.2, \dots, 1.0\}$ with step 0.1, update rate $\eta \sim \{0.1, 0.2, \dots, 1.0\}$ with step 0.1, and threshold $\epsilon \sim \{0.0, 1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2\}$. For other parameters specific to different base models, we strictly follow their instructions in the original papers.

B. Semi-supervised Node Classification.

In this task, we follow the experimental protocol established by [15] and compare our models with MLP, two basic GNNs, eight popular polynomial spectral GNNs, and five unified GNNs. For data splitting on homophilic graphs (Cora, Citeseer, and Pubmed), we apply the standard division [89] with 20

TABLE III
FULL-SUPERVISED NODE CLASSIFICATION ACCURACY (%) \pm 95% CONFIDENCE INTERVAL.

Method	Cham.	Squi.	Texas	Corn.	Actor	Cora	Cite.	Pubm.
MLP	46.59 \pm 1.84	31.01 \pm 1.18	86.81 \pm 2.24	84.15 \pm 3.05	40.18 \pm 0.55	76.89 \pm 0.97	76.52 \pm 0.89	86.14 \pm 0.25
GCN	60.81 \pm 2.95	45.87 \pm 0.88	76.97 \pm 3.97	65.78 \pm 4.16	33.26 \pm 1.15	87.18 \pm 1.12	79.85 \pm 0.78	86.79 \pm 0.31
APNP	52.15 \pm 1.79	35.71 \pm 0.78	90.64 \pm 1.70	91.52 \pm 1.81	39.76 \pm 0.49	88.16 \pm 0.74	80.47 \pm 0.73	88.13 \pm 0.33
ARMA	60.21 \pm 1.00	36.27 \pm 0.62	83.97 \pm 3.77	85.62 \pm 2.13	37.67 \pm 0.54	87.13 \pm 0.80	80.04 \pm 0.55	86.93 \pm 0.24
GPR-GNN	67.49 \pm 1.38	50.43 \pm 1.89	92.91 \pm 1.32	91.57 \pm 1.96	39.91 \pm 0.62	88.54 \pm 0.67	80.13 \pm 0.84	88.46 \pm 0.31
BernNet	68.53 \pm 1.68	51.39 \pm 0.92	92.62 \pm 1.37	92.13 \pm 1.64	41.71 \pm 1.12	88.51 \pm 0.92	80.08 \pm 0.75	88.51 \pm 0.39
ChebNetII	71.37 \pm 1.01	57.72 \pm 0.59	93.28 \pm 1.47	92.30 \pm 1.48	41.75 \pm 1.07	88.71 \pm 0.93	80.53 \pm 0.79	88.93 \pm 0.29
JacobiConv	74.20 \pm 1.03	57.38 \pm 1.25	93.44 \pm 2.13	92.95 \pm 2.46	41.17 \pm 0.64	88.98 \pm 0.46	80.78 \pm 0.79	89.62 \pm 0.41
Specformer	75.06 \pm 1.10	65.05\pm0.96	90.33 \pm 3.12	90.00 \pm 2.79	42.55 \pm 0.67	88.85 \pm 0.46	80.68 \pm 0.90	91.25 \pm 0.31
LON-GNN	73.00 \pm 2.20	60.61 \pm 1.69	87.54 \pm 3.45	84.47 \pm 3.45	39.10 \pm 1.59	89.44 \pm 1.12	81.41 \pm 1.15	90.98 \pm 0.64
OptBasisGNN	74.26 \pm 0.74	63.62 \pm 0.76	91.15 \pm 1.97	89.84 \pm 2.46	42.39 \pm 0.52	87.96 \pm 0.71	80.58 \pm 0.82	90.30 \pm 0.19
FLODE	74.38 \pm 0.92	63.09 \pm 1.04	89.34 \pm 1.32	89.02 \pm 2.95	<u>42.75\pm1.10</u>	88.28 \pm 1.02	80.66 \pm 1.16	90.59 \pm 0.55
GCNII	63.44 \pm 0.85	41.96 \pm 1.02	80.46 \pm 5.91	84.26 \pm 2.13	36.89 \pm 0.95	88.46 \pm 0.82	79.97 \pm 0.65	89.94 \pm 0.31
PDE-GCN	66.01 \pm 1.56	48.73 \pm 1.06	93.24 \pm 2.03	89.73 \pm 1.35	39.76 \pm 0.74	88.62 \pm 1.03	79.98 \pm 0.97	89.92 \pm 0.38
NodeFormer	53.02 \pm 1.58	34.25 \pm 1.96	87.71 \pm 2.13	90.00 \pm 3.45	41.74 \pm 0.61	86.93 \pm 1.22	79.58 \pm 0.85	91.27 \pm 0.39
GloGNN++	72.36 \pm 0.85	60.60 \pm 1.04	91.48 \pm 1.48	89.84 \pm 3.62	41.87 \pm 1.02	87.21 \pm 0.59	79.89 \pm 0.61	<u>86.89\pm0.33</u>
LRGNN	75.01 \pm 0.64	63.32 \pm 1.27	91.80 \pm 2.46	90.33 \pm 1.15	41.16 \pm 0.91	88.77 \pm 0.94	79.85 \pm 0.83	90.73 \pm 0.47
GNN-LF	53.74 \pm 1.29	36.15 \pm 0.86	76.07 \pm 2.62	78.36 \pm 2.46	38.39 \pm 0.81	88.51 \pm 0.89	79.84 \pm 0.56	89.86 \pm 0.23
GNN-HF	55.97 \pm 1.05	35.29 \pm 0.72	81.15 \pm 2.62	85.41 \pm 3.12	38.96 \pm 0.77	88.28 \pm 0.64	80.04 \pm 0.93	90.35 \pm 0.30
ADA-UGNN	61.09 \pm 1.51	42.02 \pm 1.26	84.92 \pm 3.12	83.61 \pm 3.44	41.10 \pm 0.62	88.74 \pm 0.85	79.81 \pm 1.11	90.61 \pm 0.44
FE-GNN	73.00 \pm 1.31	63.28 \pm 0.81	88.03 \pm 1.80	86.07 \pm 3.12	41.74 \pm 0.67	89.21 \pm 0.71	80.26 \pm 1.06	90.80 \pm 0.30
SAF	75.30\pm0.96	63.63 \pm 0.81	94.10 \pm 1.48	92.95 \pm 1.97	42.93 \pm 0.79	89.80 \pm 0.69	80.61 \pm 0.81	91.49 \pm 0.29
SAF- ϵ	74.84 \pm 0.99	<u>64.00\pm0.83</u>	94.75\pm1.64	93.28\pm1.80	42.98\pm0.61	89.87\pm0.51	81.45\pm0.59	91.52\pm0.30
Improv. ³	6.77%	12.61%	2.13%	1.15%	1.27%	1.36%	1.37%	3.01%

nodes per class for training, 500 nodes for validation, and 1,000 nodes for testing. On the other five heterophilic graphs, we leverage the sparse splitting [12] with 2.5%/2.5%/95% samples respectively for training/validation/testing. The results are reported in Table II, where the best results are bold and the underlined letters denote the second highest accuracy. We first observe that both SAF and SAF- ϵ substantially boosts its base model, BernNet, with gains reaching a notable 15.37%. This impressive enhancement is credited to their capacity to effectively exploit the task-beneficial information, which is implicitly encoded by spectral filtering in the spatial domain. This ability is particularly advantageous in contexts with limited supervision, where it allows effective leveraging of extra prior knowledge during training. Generally, our models outperform competitors on all datasets except for Chameleon and Squirrel, where SAF maintains a second-place rank with considerable improvements on BernNet by 14.56% and 9.68%. In these cases, ChebNetII initially surpasses our model, yet, with more training samples, our SAF manages to beats it by margins of 3.93% and 6.28% (see Table III). Moreover, SAF- ϵ averagely delivers better results than SAF by using thresholding sparsity to reduce non-local noise. However, this enhancement also incurs higher computational costs, as illustrated in both Section V-4 and Section VI-J.

C. Full-supervised Node Classification.

To bolster our evaluation, we expand the previously compared baselines to include five cutting-edge spatial GNN models: GCNII & PDE-GCN capturing long-range dependency, and NodeFormer, GloGNN++ & LRGNN, which go

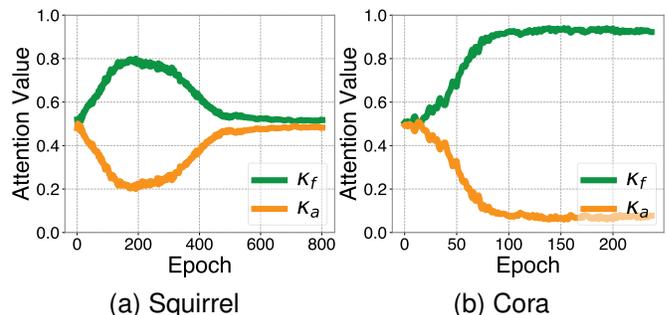


Fig. 4. Attention changing trends w.r.t. training epochs.

further by not only capturing long-range dependencies but also effectively addressing graph heterophily. For all datasets, we randomly divide them into 60%/20%/20% for training/validation/testing by following [13], [15]. Table III summarizes the mean classification accuracies. Our methods demonstrate superior performance across most datasets, with an exception on Squirrel where they achieve comparable results to Specformer. This notable performance is primarily attributed to our SAF’s effective non-local aggregation, utilizing signed edge weights to model global label relationships. This enables our methods to outperform GNNs that are specifically tailored for long-range dependency and/or graph heterophily.

³Improv. indicates the relative improvement of our SAF over its base model, BernNet [13]. For alternative implementation using ChebNetII [15] as backbone, please refer to Section VI-H.

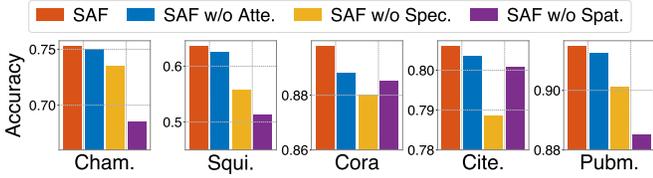


Fig. 5. Ablation study of SAF framework on five datasets.

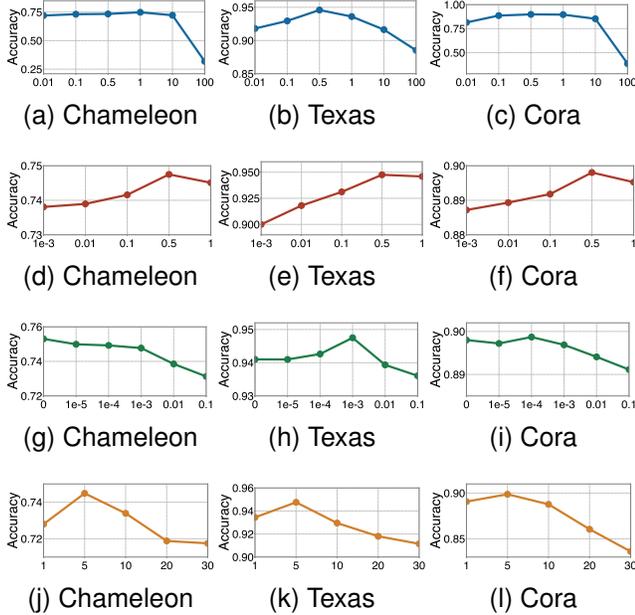


Fig. 6. Sensitivity analysis of hyper-parameters: τ , η , ϵ , and L from top to bottom rows.

D. Analysis of Attention Trends

We analyze the changing trends of the pair-wise attention weights during training SAF on Squirrel and Cora datasets. From Fig. 4(b-c), the average weights for filtering and aggregation start similarly but diverge throughout training, showing different trends in heterophilic and homophilic graphs. On the heterophilic graph Squirrel, both weights converge to similar values, demonstrating their mutual importance in modeling complex connectivity. Conversely, κ_f becomes dominant on the homophilic graph Cora due to the sufficiency of node proximity information for label prediction, thereby diminishing the relevance of κ_a and non-local aggregation.

E. Ablation Study

This section aims to validate our designs by comparing SAF with its three ablated variants – SAF w/o Atte., SAF w/o Spec., and SAF w/o Spat. – in full-supervised node classification. Specifically, Atte., Spec., and Spat. respectively refers to: attention mechanism in “Node-wise Prediction Amalgamation”, “Non-negative Spectral Filtering”, and “Non-local Spatial Aggregation”. For SAF w/o Atte., we remove the attention mechanism and equally blend predictions from different domains. SAF w/o Spec. abandons the spectral filtering phase, practically setting $\kappa_f = \mathbf{0}$, $\kappa_a = \mathbf{1}$. As the SAF w/o Spat. configuration is equivalent to BernNet

TABLE IV
EVALUATIONS ON NEW HETEROPHILIC GRAPH DATASETS.

Method	Mine.	Tolo.	Amaz.	Roma.	Penn94
MLP	50.61±0.87	74.58±0.69	45.50±0.38	66.11±0.33	74.58±0.37
GCN	72.25±0.60	76.56±0.85	48.06±0.39	53.49±0.33	82.47±0.27
APPNP	68.48±1.20	74.13±0.62	48.12±0.37	72.99±0.46	75.29±0.27
GPR-GNN	89.76±0.53	75.82±0.50	49.06±0.25	73.19±0.24	81.38±0.16
ChebNetII	83.62±1.51	78.95±0.49	49.76±0.36	74.52±0.54	83.12±0.22
JacobiConv	89.88±0.33	77.24±0.39	43.89±0.28	74.30±0.50	83.35±0.11
NodeFormer	89.89±0.46	80.31±0.75	43.67±1.54	73.59±0.60	69.66±0.83
GloGNN++	72.59±1.54	79.01±0.84	50.03±0.29	66.10±0.26	73.15±0.59
FE-GNN	84.68±0.36	79.31±0.59	49.46±0.29	74.50±0.30	82.30±0.54
BernNet	77.75±0.61	75.35±0.63	49.84±0.52	74.56±0.74	82.47±0.21
SAF	90.54±0.30	79.38±0.58	50.49±0.28	74.87±0.22	83.86±0.26
Improv.	12.79%	4.03%	0.65%	0.31%	1.39%

model, the corresponding results are posted directly. From Fig. 5, we can draw several insights: **1)** The impact of Atte. module on our SAF varies by datasets, e.g., on Chameleon and Squirrel, showing a slight performance reduction upon its removal. This observation aligns with our observation that their optimal attention values are close to an even split, as suggested in Figures 4(a). Conversely, Cora dataset exhibits a notable drop, due to its optimal attention weights being far from even, as depicted in Fig. 4(b); **2)** Spectral filtering (Spec. module) remains vital for discriminative node representation learning. Specifically, the quality of the adapted new graph fundamentally hinges on the graph spectral filters’ training, as underscored by their theoretical interaction in Eq. (6). Practically, the absence of spectral filtering markedly reduces model accuracy, confirming its importance in SAF; **3)** This visualization not only reaffirms the pivotal role of the non-local aggregation (Spat. module), but also underscores its position as the most crucial component in advancing spectral GNNs within the SAF framework.

F. Parameter Study

This section presents the sensitivity analysis of hyper-parameters including τ , η , ϵ , and L . Fig. 6 visualizes how varying these parameters within a broad range influences learning performance, showcasing our model’s robust stability over diverse settings. Beyond empirical observation, we also provides deeper insights into parameter understanding and rationalizes the chosen ranges for parameter searching: **1)** The scaling parameter $\tau = \frac{\alpha}{1-\alpha}$, crucial in new graph construction in Eq. (6), stems from the trade-off parameter $\alpha \in (0, 1)$ within the graph optimization problem in Eq. (3). While theoretically we have $0 = \frac{0}{1-0} < \tau < \frac{1}{1-1} = \infty$, practical considerations for extracting structural information suggest a larger penalty on the trace objective term $\text{tr}(\mathbf{Z}^T \gamma_\theta(\hat{\mathbf{L}})\mathbf{Z})$, i.e., keeping $\alpha < 0.5$, thereby limiting $\tau < \frac{0.5}{1-0.5} = 1$. This rationale substantiates our selection of τ within the set $\{0.1, 0.2, \dots, 1\}$ as stated in Section VI-A4, aligning with the observed optimal performance in Figures 6(a)-(e). When addressing graphs with noisy structure, we may adjust the upper limit of α to $t \in (0, 1)$, setting τ ’s maximum possible value to $\frac{t}{1-t}$. For graph benchmarking evaluations in this work, where extracting structural information is important, we practically set $t = 0.5$; **2)** For the non-local aggregation

TABLE V
FULL-SUPERVISED NODE CLASSIFICATION ACCURACY (%) WHILE IMPLEMENTING SAF UPON CHEBNETII.

Method	Cham.	Squi.	Texas	Corn.	Actor	Cora	Cite.	Pubm.
ChebNetII	71.37±1.01	57.72±0.59	93.28±1.47	92.30±1.48	41.75±1.07	88.71±0.93	80.53±0.79	88.93±0.29
SAF-Cheb	74.97±0.66	64.06±0.59	94.43±1.81	92.62±2.13	42.65±1.01	89.56±0.64	80.68±0.68	91.27±0.34
SAF-Cheb- ϵ	75.25±0.96	64.42±0.82	94.26±1.64	93.12±1.64	42.79±1.04	89.61±0.71	81.08±0.68	91.73±0.18
Improv.	3.88%	6.70%	1.15%	0.82%	1.04%	0.90%	0.15%	2.80%

layer number L , a noticeable decline in model performance is observed when L exceeds 10. This is attributed to the non-local nature of our new graph, which facilitates efficient information exchange between nodes. Exceeding a certain number of layers may potentially lead to oversmoothing, where there is an overemphasis on global information, thus degrading model performance. However, choosing the number of layers within a reasonable range generally ensures consistent and impressive model performance, as verified in Figures 6(j)-(l).

G. New Benchmarks for Graph Heterophily

For a more extensive evaluation across various domains, we also test SAF on five recently introduced datasets, including Mine., Tolo., Amaz., Roma., and Penn94 [67], [84]. In this context, we draw comparisons solely with MLP, GCN, APPNP, along with seven GNN models that have previously shown promising results in prior tasks, namely GPR-GNN, BernNet, ChebNetII, JacobiConv, NodeFormer, GloGNN++ and FE-GNN. Table IV lists the average classification accuracies, obtained over random splits provided by [67], [84], with a distribution of 50%/25%/25% for training/validation/testing. In summary, SAF achieves significant performance gains of 12.79% and 4.03% on Minesweeper and Tolokers datasets, respectively, while maintaining competitiveness on the others.

H. SAF with ChebNetII as Base Model

To expand the versatility of our SAF framework, we introduced ChebNetII [15] as an alternative base model, chosen for its adherence to the non-negative constraint, critical in our model design as stated in Section V. The rationale behind this choice is ChebNetII’s use of Chebyshev interpolation for learning Chebyshev polynomials, where the constraint can be ensured by keeping its learnable parameters $\{\gamma_j\}_{j=0}^K$ non-negative. Our experiments, as shown in Table V, confirm that SAF can significantly enhance ChebNetII’s performance, underscoring the framework’s flexibility with different spectral filters. Interestingly, we observed that SAF, utilizing Bernstein polynomials (SAF-Bern), slightly surpasses its performance with Chebyshev polynomials (SAF-Cheb) in most datasets. The margin of improvement over the base model is also more pronounced with SAF-Bern. This phenomenon could be attributed to the $g_\phi(\lambda) \leq 1$ constraint within SAF (refer to Section V), necessitating the rescaling of filter functions by their maximum values. For Bernstein polynomials, this maximum is readily obtained as the largest polynomial coefficient $\max\{\phi_k\}_{k=0}^K$, as per Proposition 3. However, for Chebyshev polynomials, the best theoretical upper bound is

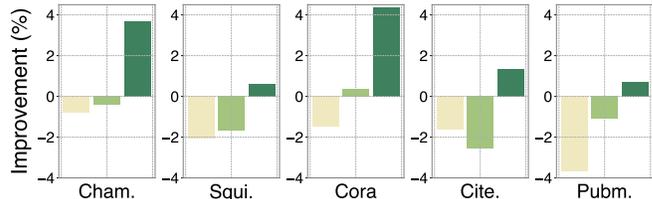


Fig. 7. Improved classification accuracy by GloGNN++, LRGNN and our SAF (from left to right) over the vanilla GCN under semi-supervised setting (2.5%/2.5%/95%).

the sum of absolute coefficients, $\sum_{k=0}^K |\phi_k|$, which is comparatively less precise. This difference may impact the quality of graph construction and, subsequently, compromise the model’s performance. Exploring these nuances will be a focal point of our future research.

I. Supervision Dependence in GloGNN++ and LRGNN

This subsection empirically verifies the heavy reliance on label supervision for both GloGNN++ [56] and LRGNN [86] models. In Fig. 7, we compare the improved classification accuracy of these models and our SAF upon the vanilla GCN under a semi-supervised setting. The figure illustrates that, despite competitive results under dense supervision (Table III), the performance of both GloGNN++ and LRGNN deteriorates, demonstrating negative optimization on many datasets, when the label rate drops approximately from 60% to 2.5%. This degradation highlights their dependency on high label rates for effective learning. In contrast, our SAF constantly delivers good performance regardless of the supervision level, showing its effectiveness in both high and low label rate scenarios.

J. Time and Space Overheads

1) *Eigendecomposition*: Our SAF framework pre-computes eigendecomposition once per graph and reuses it in Eq. (6) in both training and inference. This aspect is crucial, as the forward-pass cost in model training often exceeds the preprocessing expense of eigendecomposition. To empirically validate this, we compare the time overheads of eigendecomposition with the training times of various models in Table VI. Overall, we have following observations: **1)** For datasets with a small number of nodes, the time consumed by decomposition is significantly less than the time required for model training; **2)** For medium-sized graphs such as Pubmed, while the full decomposition time exceeds that of BernNet, it still maintains efficiency against more advanced GNNs such as ChebNetII,

TABLE VI
TIME OVERHEADS (S).

Method	Cham.	Squi.	Texas	Corn.	Actor	Cora	Cite.	Pubm.	Penn94
# Nodes	2,277	5,201	183	183	7,600	2,708	3,327	19,717	41,554
# Edges	36,101	217,073	309	295	33,544	5,429	4,732	44,338	1,362,229
BernNet	8.36	13.74	3.92	4.16	4.88	5.24	5.52	6.06	24.05
ChebNetII	22.82	30.73	11.47	9.64	14.88	19.96	16.14	36.91	41.67
FE-GNN	3.99	45.89	1.03	0.84	0.98	2.54	2.06	6.78	7.33
NodeFormer	58.96	79.66	14.29	18.89	66.20	19.25	32.00	68.57	122.91
GloGNN++	35.63	68.31	4.47	3.00	73.13	32.68	12.35	5266.53	3614.37
SAF	11.55	18.78	4.38	4.70	5.36	6.04	6.12	18.43	23.49
Decomposition	0.58	1.59	0.02	0.02	3.93	1.00	0.77	21.34	4.76

TABLE VII
SPACE OVERHEADS (MB).

Method	Cham.	Squi.	Texas	Corn.	Actor	Cora	Cite.	Pubm.	Penn94
BernNet	72	232	5	5	292	64	152	1546	1902
ChebNetII	72	231	5	5	291	63	152	1584	1850
FE-GNN	1337	6919	23	10	416	302	740	2213	5854
NodeFormer	1522	3965	15	37	775	480	764	2119	3056
GloGNN++	290	1525	5	5	2471	331	607	17892	25260
SAF	112	440	5	5	733	120	237	4515	8491
Decomposition	141	540	1	1	1206	140	239	7641	4

NodeFormer and GloGNN++; **3)** Moving to the large-scale graph, Penn94 (with 41,554 nodes and 1,362,229 edges), where only partial eigendecomposition with 100 extremal eigenvalues is considered, the computation time is markedly reduced compared to all the models. For space overheads in Table VII, similar patterns can be observed.

2) *Model Comparison:* In Table VI, we compare the running times of our SAF against two notable spectral GNNs (BernNet, ChebNetII), two non-local GNNs (NodeFormer, GloGNN++), and one unified GNN model (FE-GNN). Generally, one can observe that SAF, while slightly slower than its base model, BernNet, due to the integration of non-local spatial aggregation, remains more efficient than or comparable to other SOTA methods, particularly those also employing non-local approaches such as NodeFormer and GloGNN++.

VII. CONCLUSION

This study presents a cross-domain analysis on GNN models, offering a fresh perspective by rethinking spectral GNNs from a spatial lens. We reveal that spectral GNNs fundamentally leads the original graph to an adapted new one, which exhibits non-locality and accommodates signed edge weights to reflect label consistency among nodes. This insight leads to our proposed Spatially Adaptive Filtering (SAF) framework, enhancing spectral GNNs for more effective and versatile graph representation learning. While SAF adeptly captures long-range dependencies and addresses graph heterophily, we acknowledge two limitations of this work: 1) the non-negative constraints of the proposed SAF on graph filters might limit model expressiveness, indicating room for theoretical refinement; 2) although this study focuses on a

node-level investigation, it raises intriguing questions about the implications of spectral GNNs at the graph-level in the spatial domain. Future work could expand this examination by relaxing theoretical constraints or exploring the cross-domain interplay from a broader graph-level viewpoint.

ACKNOWLEDGMENTS

The work was supported by the following: National Natural Science Foundation of China under No. 92370119, and 62376113.

REFERENCES

- [1] U. S. Shanthamallu, J. J. Thiagarajan, H. Song, and A. Spanias, "Gramme: Semisupervised learning using multilayered graph attention models," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3977–3988, 2019.
- [2] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.
- [3] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, 2021.
- [4] I. Chami, S. Abu-El-Haija, B. Perozzi, C. Ré, and K. Murphy, "Machine learning on graphs: A model and comprehensive taxonomy," *Journal of Machine Learning Research*, vol. 23, no. 89, pp. 1–64, 2022.
- [5] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He *et al.*, "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Transactions on Recommender Systems*, vol. 1, no. 1, pp. 1–51, 2023.
- [6] A. Gravina and D. Bacciu, "Deep learning for dynamic graphs: models and benchmarks," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [7] F. R. K. Chung, "Spectral graph theory," 1996.

- [8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
- [9] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017.
- [11] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, "Cayleynets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 97–109, 2018.
- [12] E. Chien, J. Peng, P. Li, and O. Milenkovic, "Adaptive universal generalized pagerank graph neural network," in *International Conference on Learning Representations*, 2021.
- [13] M. He, Z. Wei, H. Xu *et al.*, "Bernnet: Learning arbitrary graph spectral filters via bernstein approximation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 239–14 251, 2021.
- [14] M. Ju, S. Hou, Y. Fan, J. Zhao, Y. Ye, and L. Zhao, "Adaptive kernel graph neural network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 6, 2022, pp. 7051–7058.
- [15] M. He, Z. Wei, and J.-R. Wen, "Convolutional neural networks on graphs with chebyshev approximation, revisited," in *Advances in Neural Information Processing Systems*, 2022.
- [16] X. Wang and M. Zhang, "How powerful are spectral graph neural networks," in *International Conference on Machine Learning*. PMLR, 2022, pp. 23 341–23 362.
- [17] J. Guo, K. Huang, X. Yi, and R. Zhang, "Graph neural networks with diverse spectral filtering," in *Proceedings of the ACM Web Conference 2023*, 2023, pp. 306–316.
- [18] Q. Tao, Z. Wang, W. Yu, Y. Li, and Z. Wei, "Longnn: Spectral gnns with learnable orthonormal basis," *arXiv preprint arXiv:2303.13750*, 2023.
- [19] Y. Guo and Z. Wei, "Graph neural networks with learnable and optimal polynomial bases," *arXiv preprint arXiv:2302.12432*, 2023.
- [20] M. Balcilar, G. Renton, P. Héroux, B. Gauzere, S. Adam, and P. Honeine, "Bridging the gap between spectral and spatial domains in graph neural networks," *arXiv preprint arXiv:2003.11702*, 2020.
- [21] Z. Chen, F. Chen, L. Zhang, T. Ji, K. Fu, L. Zhao, F. Chen, L. Wu, C. Aggarwal, and C.-T. Lu, "Bridging the gap between spatial and spectral domains: A unified framework for graph neural networks," *arXiv preprint arXiv:2107.10234*, 2021.
- [22] Y. Guo and Z. Wei, "Clenshaw graph neural networks," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 614–625.
- [23] Y. Ma, X. Liu, T. Zhao, Y. Liu, J. Tang, and N. Shah, "A unified view on graph neural networks as graph signal denoising," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2021, pp. 1202–1211.
- [24] M. Zhu, X. Wang, C. Shi, H. Ji, and P. Cui, "Interpreting and unifying graph neural networks with an optimization framework," in *Proceedings of the Web Conference 2021*, 2021, pp. 1215–1226.
- [25] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and deep graph convolutional networks," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1725–1735.
- [26] M. Balcilar, R. Guillaume, P. Héroux, B. Gaüzère, S. Adam, and P. Honeine, "Analyzing the expressive power of graph neural networks in a spectral perspective," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [27] J. Sun, L. Zhang, G. Chen, P. Xu, K. Zhang, and Y. Yang, "Feature expansion for graph neural networks," in *International Conference on Machine Learning*. PMLR, 2023, pp. 33 156–33 176.
- [28] R. Liao, Z. Zhao, R. Urtasun, and R. S. Zemel, "Lanczosnet: Multi-scale deep graph convolutional networks," *arXiv preprint arXiv:1901.01484*, 2019.
- [29] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Zeitschrift für Physik*, vol. 43, no. 3-4, pp. 172–198, 1927.
- [30] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," *Journal of Fourier analysis and applications*, vol. 3, pp. 207–238, 1997.
- [31] A. Agaskar and Y. M. Lu, "A spectral graph uncertainty principle," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4338–4356, 2013.
- [32] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1263–1272.
- [33] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [34] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.
- [35] Y. Dong, K. Ding, B. Jalaian, S. Ji, and J. Li, "Adagmn: Graph neural networks with adaptive frequency response filter," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2021, pp. 392–401.
- [36] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, "Graph neural networks with convolutional arma filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [37] J. Gasteiger, A. Bojchevski, and S. Günnemann, "Predict then propagate: Graph neural networks meet personalized pagerank," in *International Conference on Learning Representations*, 2019.
- [38] M. Eliasof, E. Haber, and E. Treister, "Pde-gcn: novel architectures for graph neural networks motivated by partial differential equations," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3836–3849, 2021.
- [39] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, "Geom-gcn: Geometric graph convolutional networks," *arXiv preprint arXiv:2002.05287*, 2020.
- [40] M. Liu, Z. Wang, and S. Ji, "Non-local graph neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 10 270–10 276, 2021.
- [41] Q. Wu, W. Zhao, Z. Li, D. P. Wipf, and J. Yan, "Nodeformer: A scalable graph structure learning transformer for node classification," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 387–27 401, 2022.
- [42] K. Ding, E. Nouri, G. Zheng, H. Liu, and R. White, "Toward robust graph semi-supervised learning against extreme data scarcity," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [43] D. Bo, C. Shi, L. Wang, and R. Liao, "Specformer: Spectral graph neural networks meet transformers," *arXiv preprint arXiv:2303.01028*, 2023.
- [44] S. Maskey, R. Paolino, A. Bacho, and G. Kutyniok, "A fractional graph laplacian approach to oversmoothing," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
- [45] M. Benzi, D. Bertaccini, F. Durastante, and I. Simunec, "Non-local network dynamics via fractional graph laplacians," *Journal of Complex Networks*, vol. 8, no. 3, p. cnaa017, 2020.
- [46] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra, "Beyond homophily in graph neural networks: Current limitations and effective designs," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7793–7804, 2020.
- [47] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.
- [48] D. Bo, X. Wang, C. Shi, and H. Shen, "Beyond low-frequency information in graph convolutional networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 3950–3957.
- [49] R. Wang, S. Mou, X. Wang, W. Xiao, Q. Ju, C. Shi, and X. Xie, "Graph structure estimation neural networks," in *Proceedings of the Web Conference 2021*, 2021, pp. 342–353.
- [50] J. Guo, K. Huang, R. Zhang, and X. Yi, "Es-gnn: Generalizing graph neural networks beyond homophily with edge splitting," *arXiv preprint arXiv:2205.13700*, 2022.
- [51] Y. Yan, Y. Chen, H. Chen, M. Xu, M. Das, H. Yang, and H. Tong, "From trainable negative depth to edge heterophily in graphs," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [52] J. Chen, S. Chen, J. Gao, Z. Huang, J. Zhang, and J. Pu, "Exploiting neighbor effect: Conv-agnostic gnn framework for graphs with heterophily," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [53] J. Yoo, M.-C. Lee, S. Shekhar, and C. Faloutsos, "Less is more: Slim for accurate, robust, and interpretable graph mining," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 3128–3139.
- [54] G. Cui and Z. Wei, "Mggn: Graph neural networks inspired by distance geometry problem," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 335–347.

- [55] J. Li, R. Zheng, H. Feng, M. Li, and X. Zhuang, "Permutation equivariant graph framelets for heterophilous graph learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [56] X. Li, R. Zhu, Y. Cheng, C. Shan, S. Luo, D. Li, and W. Qian, "Finding global homophily in graph neural networks when meeting heterophily," in *International Conference on Machine Learning*. PMLR, 2022, pp. 13 242–13 256.
- [57] L. Liang, X. Hu, Z. Xu, Z. Song, and I. King, "Predicting global label relationship matrix for graph neural networks under heterophily," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- [58] D. Spielman, "Spectral graph theory," *Combinatorial Scientific Computing*, vol. 18, p. 18, 2012.
- [59] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.
- [60] R. T. Farouki, "The bernstein polynomial basis: A centennial retrospective," *Computer Aided Geometric Design*, vol. 29, no. 6, pp. 379–419, 2012.
- [61] V. Powers and B. Reznick, "Polynomials that are positive on an interval," *Transactions of the American Mathematical Society*, vol. 352, no. 10, pp. 4677–4692, 2000.
- [62] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [63] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," 1950.
- [64] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.
- [65] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei, "Am-gcn: Adaptive multi-channel graph convolutional networks," in *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2020, pp. 1243–1253.
- [66] S. Zhu, S. Pan, C. Zhou, J. Wu, Y. Cao, and B. Wang, "Graph geometry interaction learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7548–7558, 2020.
- [67] D. Lim, F. Hohne, X. Li, S. L. Huang, V. Gupta, O. Bhalerao, and S. N. Lim, "Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20 887–20 902, 2021.
- [68] O. Platonov, D. Kuznedelev, A. Babenko, and L. Prokhorenkova, "Characterizing graph datasets for node classification: Homophily-heterophily dichotomy and beyond," in *The Second Learning on Graphs Conference*, 2023.
- [69] Y. Cai, G. Fang, and P. Li, "A note on sparse generalized eigenvalue problem," *Advances in Neural Information Processing Systems*, vol. 34, pp. 23 036–23 048, 2021.
- [70] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [71] V. P. Dwivedi, C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking graph neural networks," *arXiv preprint arXiv:2003.00982*, 2020.
- [72] H. Wang, H. Yin, M. Zhang, and P. Li, "Equivariant and stable positional encoding for more powerful graph neural networks," in *International Conference on Learning Representations*, 2022.
- [73] Y. You, T. Chen, Z. Wang, and Y. Shen, "Graph domain adaptation via theory-grounded spectral regularization," in *The Eleventh International Conference on Learning Representations*, 2022.
- [74] H. Chang, Y. Rong, T. Xu, Y. Bian, S. Zhou, X. Wang, J. Huang, and W. Zhu, "Not all low-pass filters are robust in graph convolutional networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25 058–25 071, 2021.
- [75] D. Lim, J. Robinson, L. Zhao, T. Smidt, S. Sra, H. Maron, and S. Jegelka, "Sign and basis invariant networks for spectral graph representation learning," *arXiv preprint arXiv:2202.13013*, 2022.
- [76] D. Lim, J. Robinson, S. Jegelka, and H. Maron, "Expressive sign equivariant networks for spectral geometric learning," *arXiv preprint arXiv:2312.02339*, 2023.
- [77] D. Kreuzer, D. Beaini, W. Hamilton, V. Létourneau, and P. Tossou, "Rethinking graph transformers with spectral attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 618–21 629, 2021.
- [78] J. Kim, D. Nguyen, S. Min, S. Cho, M. Lee, H. Lee, and S. Hong, "Pure transformers are powerful graph learners," *Advances in Neural Information Processing Systems*, vol. 35, pp. 14 582–14 595, 2022.
- [79] L. Rampásek, M. Galkin, V. P. Dwivedi, A. T. Luu, G. Wolf, and D. Beaini, "Recipe for a general, powerful, scalable graph transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 14 501–14 515, 2022.
- [80] B. Rozemberczki, C. Allen, and R. Sarkar, "Multi-scale attributed node embedding," *Journal of Complex Networks*, vol. 9, no. 2, p. cnab014, 2021.
- [81] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 807–816.
- [82] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data," *AI Magazine*, vol. 29, no. 3, pp. 93–93, 2008.
- [83] G. Namata, B. London, L. Getoor, B. Huang, and U. Edu, "Query-driven active surveying for collective classification," in *10th International Workshop on Mining and Learning with Graphs*, vol. 8, 2012, p. 1.
- [84] O. Platonov, D. Kuznedelev, M. Diskin, A. Babenko, and L. Prokhorenkova, "A critical look at the evaluation of gnns under heterophily: are we really making progress?" *arXiv preprint arXiv:2302.11640*, 2023.
- [85] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of facebook networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 16, pp. 4165–4180, 2012.
- [86] L. Liang, X. Hu, Z. Xu, Z. Song, and I. King, "Predicting global label relationship matrix for graph neural networks under heterophily," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
- [87] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [88] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2623–2631.
- [89] Z. Yang, W. Cohen, and R. Salakhudinov, "Revisiting semi-supervised learning with graph embeddings," in *International Conference on Machine Learning*. PMLR, 2016, pp. 40–48.