A Unified Spectral-Spatial Framework for GNNs: Balancing Over-Smoothing and Over-Squashing

Xinya Qin[1], **Lu Bai**[1*], **Lixin Cui**[2], **Ming Li**[3], **Hangyuan Du**[4] and **Jing Li**[5]
[1]School of Artificial Intelligence, Beijing Normal University, Beijing, China
[2]School of Information, Central University of Finance and Economics, Beijing, China
[3]Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China
[4]School of Computer and Information Technology, Shanxi University, Taiyuan, China
[5]School of Geospatial Artificial Intelligence, East China Normal University, Shanghai, China.
XinyaQin@mail.bnu.edu.cn, bailu@bnu.edu.cn, cuilixin@cufe.edu.cn.

Abstract

Over-smoothing (OSM) and over-squashing (OSQ) are two fundamental phenomena that limit the performance of Graph Neural Networks (GNNs), yet a unified spectral-spatial understanding of these phenomena remains underexplored. In this paper, we adopt polynomial spectral filters as an analytical tool to establish a unified spectral-spatial framework for graph convolution and systematically characterize the effect of the polynomial order k on information propagation in GNNs. Within this framework, we reveal an intrinsic trade-off induced by the polynomial order. Specifically, higher-order filters enhance spectral expressiveness and alleviate OSM caused by the dominant low-frequency components. However, they also expand the spatial receptive field, thereby intensifying information compression and increasing the risk of OSQ. Based on this analysis, we provide a principled guideline for selecting the polynomial order and propose a Quadratic Spectral Graph Convolution Network (QS-GCN) for graph classification. Experiments demonstrate the effectiveness and robustness of the proposed method.

1 Introduction

Graph-structured data has gained increasing attention in the field of machine learning and has been widely applied in various real-world domains, including bioinformatics [Yan *et al.*, 2023], social networks [Min *et al.*, 2021], and recommendation systems [Wang *et al.*, 2023]. However, most classical deep learning methods are originally designed for grid-structured data such as images or sequences, which cannot directly accommodate the irregular graph data from the non-Euclidean space. To address this limitation, Graph Neural Networks (GNNs) have emerged as powerful frameworks for learning on graphs, achieving notable success in tasks such as node classification, graph classification, and link prediction.

*Corresponding Author: Lu Bai

Most existing GNNs generally follow two dominant perspectives including spectral filtering and spatial message passing. Spectral methods are grounded in the eigendecomposition of the graph Laplacian and the graph Fourier transform, where the graph convolution is formulated as the spectral filtering operation. Although this formulation provides a solid theoretical foundation, it typically incurs high computational costs due to the eigendecomposition [Bruna *et al.*, 2013]. To alleviate this issue, various spectral variants based on polynomial approximations or simplified filters have been proposed, such as the ChebNet [Defferrard *et al.*, 2016], vanilla Graph Convolutional Network (GCN) [Kipf and Welling, 2017], and JacobiConv [Wang and Zhang, 2022]. Among them, the first-order approximation adopted by the vanilla GCN enables spectral filtering to be efficiently implemented in the spatial domain, where a node updates its representation by aggregating the representations of its neighbors. This process facilitates the propagation of information across nodes, offering improved interpretability. This mechanism has further driven the rapid development of spatial GNNs, such as the GraphSAGE [Hamilton *et al.*, 2017], GIN [Xu *et al.*, 2019], and DGCNN [Zhang *et al.*, 2019].

Despite these advances in modeling flexibility and computational efficiency, GNNs still suffer from two fundamental challenges in practice, namely over-smoothing (OSM) [Rusch *et al.*, 2023; Li *et al.*, 2018; NT and Maehara, 2019; Chen *et al.*, 2020a] and over-squashing (OSQ) [Dai *et al.*, 2023; Singh, 2025]. As the network depth or propagation distance increases, node representations tend to become increasingly homogeneous. This homogenization degrades the discriminability of the node representations and is commonly referred to as OSM. Meanwhile, OSQ arises from the rapid expansion of the receptive field, where the information from a large number of distant nodes must be compressed into fixed-dimensional embeddings, leading to severe information loss and creating an information bottleneck.

However, most existing studies treat OSM and OSQ as separate phenomena, and only a limited number of works have begun to explore the trade-off between them. For instance, FoSR [Karhadkar *et al.*, 2023] and ProxyDelete [Jamadandi *et al.*, 2024] investigate spectral reconnection strategies from

the perspectives of the spectral gap and the Cheeger constant, aiming to improve graph structures and alleviate excessive smoothing and compression during the information propagation. From a geometric viewpoint, BORF [Nguyen *et al.*, 2023] analyzes the relationship between edges with the positive and negative curvature as well as the OSM/OSQ phenomena, and proposes a curvature-driven graph reconnection method that can mitigate both issues to some extent. Nevertheless, these approaches largely rely on a single analytical perspective, focusing either on spectral properties or geometric characteristics. Consequently, they do not capture the interplay between spectral and spatial factors that jointly influence OSM and OSQ. A unified framework that systematically characterizes this interplay from both spectral and spatial perspectives is essential to understand the intrinsic trade-off, explain feature homogenization and information compression in deep GNNs, and guide the design of architectures that balance these phenomena effectively.

Motivated by the aforementioned observations, we revisit classical polynomial spectral filters and establish a unified spectral-spatial framework for graph convolution, aiming to systematically investigate the trade-off between OSM and OSQ in a principled manner. The main contributions of this paper are summarized as follows.

First, we establish a unified spectral-spatial analysis framework based on the polynomial spectral filters, which explicitly and rigorously characterizes the intrinsic relationship between the polynomial order k, spectral frequency selectivity, and the spatial aggregation range.

Second, within the proposed framework, we provide a systematic theoretical analysis of how the polynomial order k influences the OSM and OSQ issues in GNNs. We show that the increasing polynomial order k can enhance the spectral expressiveness and alleviate OSM caused by dominant low-frequency components, but also simultaneously expands the spatial receptive field and further exacerbates the information compression, thereby increasing the risk of OSQ.

Third, motivated by the above analysis, we argue that a moderate polynomial order is critical for achieving a better balance between OSM and OSQ. Following this principle, we propose a new Quadratic Spectral Graph Convolution Network (QS-GCN) for graph classification, comprising a Quadratic Spectral Graph Convolution (QS-Conv) and an alignment-based readout module. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness of the proposed QS-GCN.

This paper is organized as follows. Section 2 reviews the OSM and OSQ problems. Section 3 presents a unified theoretical framework, and Section 4 analyzes OSM and OSQ through this framework. Section 5 introduces the proposed trade-off strategy QS-GCN. Section 6 reports the experimental evaluation. Section 7 concludes this paper.

2 Related Works

2.1 The Over-Smoothing (OSM) Problem

The OSM problem is a fundamental challenge in GNNs, where repeated message passing causes node representations to become increasingly indistinguishable, ultimately degrading the performance on both node-level and graph-level tasks. Several studies have investigated its underlying mechanisms from different perspectives. For example, [Li *et al.*, 2018] model GCN layers as Laplacian smoothing processes and demonstrate that deeper networks result in indistinguishable node embeddings. From a spectral perspective, [NT and Maehara, 2019] and [Balcilar *et al.*, 2021] interpret GCNs as low-pass filters that suppress high-frequency signals. Although such filtering can effectively reduce noise, the excessive smoothing inevitably blurs discriminative information. To address the OSM problem, various solutions have been proposed, including PairNorm [Zhao and Akoglu, 2019], DropEdge [Rong *et al.*, 2020], GCNII [Chen *et al.*, 2020b], GraphCON [Rusch *et al.*, 2022a], G^2 [Rusch *et al.*, 2022b], and Res-GCN [Li *et al.*, 2023].

2.2 The Over-Squashing (OSQ) Problem

The OSQ issue is another challenge for GNNs [Dai *et al.*, 2023; Singh, 2025]. It refers to the phenomenon where the information from a large number of nodes within an expanding receptive field is excessively compressed into fixed-dimensional node vectors, resulting in ineffective message propagation from distant nodes. Recently, several approaches have been proposed to alleviate OSQ. For instance, FA [Alon and Yahav, 2021] connects all nodes of the final layer to reduce the compression. SDRF [Topping *et al.*, 2022] improves the message flow by modifying negatively curved edges. FoSR [Karhadkar *et al.*, 2023] alleviates OSQ by optimizing the spectral gap. GTR [Black *et al.*, 2023] rewires edges to reduce the effective resistance between nodes.

3 A Unified Spectral-Spatial Framework

This section establishes a unified spectral-spatial framework for graph convolution based on polynomial spectral filters. By explicitly connecting spectral filtering behavior with spatial neighborhood aggregation, the framework provides a principled foundation for analyzing the intrinsic trade-off between OSM and OSQ arising in GNNs.

Let $G = (V, E)$ be an undirected graph with n nodes, represented by an adjacency matrix $A \in \{0,1\}^{n\times n}$ and a node feature matrix $X \in \mathbb{R}^{n\times f}$. To incorporate self-information into the message propagation, self-loops are added, yielding $\tilde{A} = A + I$. Let $\tilde{D}$ denote the degree matrix, whose diagonal entries are given by $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. The symmetrically normalized adjacency matrix is $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$. Accordingly, the normalized graph Laplacian is $\hat{L} = I - \hat{A}$. Since $\hat{L}$ is symmetric and positive semi-definite, it admits the eigendecomposition $\hat{L} = U\,\mathrm{diag}(\boldsymbol{\lambda})\,U^\top$, where $U \in \mathbb{R}^{n\times n}$ is the matrix of eigenvectors, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ denotes the corresponding eigenvalues.

A general spectral graph convolution is defined as

$$X' = \sigma\Big(U\,\mathrm{diag}\big(g(\boldsymbol{\lambda})\big)\,U^\top XW\Big), \tag{1}$$

where $\sigma(\cdot)$ is a nonlinear activation function, W is a learnable weight matrix, and $g(\boldsymbol{\lambda})$ denotes the element-wise evaluation

of the spectral response function $g(\lambda)$ on the eigenvalue vector $\boldsymbol{\lambda}$, i.e., $g(\boldsymbol{\lambda}) = \big(g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)\big)$. In this paper, we focus on the k-th order polynomial spectral filter

$$g_k(\lambda) = \sum_{t=0}^{k} \alpha_t \lambda^t, \tag{2}$$

where $\{\alpha_t\}_{t=0}^{k}$ are polynomial coefficients. Substituting $g_k(\boldsymbol{\lambda})$ into Eq. (1), we have

$$\begin{aligned} X' &= \sigma\Big(U \operatorname{diag}\big(g_k(\boldsymbol{\lambda})\big)\, U^\top XW\Big) \\ &= \sigma\Big(U \operatorname{diag}\Big(\sum_{t=0}^{k} \alpha_t \lambda_1^t, \ldots, \sum_{t=0}^{k} \alpha_t \lambda_n^t\Big)\, U^\top XW\Big) \\ &= \sigma\Big(\sum_{t=0}^{k} \alpha_t\, U \operatorname{diag}(\lambda_1^t, \ldots, \lambda_n^t)\, U^\top XW\Big) \\ &= \sigma\Big(\sum_{t=0}^{k} \alpha_t \hat{L}^t XW\Big). \end{aligned} \tag{3}$$

This derivation shows that polynomial spectral filtering is equivalent to a linear combination of powers of the normalized graph Laplacian applied to node features.

From the spatial perspective, aggregating the information from k-hop neighborhoods can be written as

$$X' = \sigma\Big(\hat{A}^k XW\Big). \tag{4}$$

Since the normalized Laplacian satisfies $\hat{L} = I - \hat{A}$, the above polynomial spectral convolution is equivalent to

$$X' = \sigma\Big(\sum_{t=0}^{k} \alpha_t \hat{L}^t XW\Big) = \sigma\Big(\sum_{t=0}^{k} \alpha_t (I - \hat{A})^t XW\Big). \tag{5}$$

Applying the binomial expansion to each term, we have

$$(I - \hat{A})^t = \sum_{s=0}^{t} \binom{t}{s} (-1)^s \hat{A}^s. \tag{6}$$

Substituting this expansion into the convolution and rearranging the sums, we have

$$\begin{aligned} X' &= \sigma\Big(\sum_{t=0}^{k} \alpha_t \sum_{s=0}^{t} \binom{t}{s} (-1)^s \hat{A}^s XW\Big) \\ &= \sigma\Big(\sum_{s=0}^{k} \sum_{t=s}^{k} \alpha_t \binom{t}{s} (-1)^s \hat{A}^s XW\Big). \end{aligned} \tag{7}$$

Defining

$$\beta_s = \sum_{t=s}^{k} \alpha_t \binom{t}{s} (-1)^s, \tag{8}$$

we can rewrite the above convolution compactly as

$$X' = \sigma\Big(\sum_{s=0}^{k} \beta_s \hat{A}^s XW\Big). \tag{9}$$

This derivation demonstrates that a k-th order polynomial spectral filter is spatially equivalent to a weighted aggregation over neighborhoods up to k hops. Thus, the polynomial order k governs both the spectral expressiveness and the spatial receptive field of the convolution.

4 Theoretical Analysis of OSM and OSQ

The unified spectral-spatial framework indicates that increasing polynomial order k allows the convolution to aggregate more information over larger neighborhoods while realizing richer spectral responses. This dual perspective provides a solid theoretical foundation for analyzing the information propagation in GNNs in relation to OSM and OSQ. Based on this foundation, we systematically investigate the impact of the polynomial order k on the OSM and OSQ issues.

4.1 Analyzing OSM with Polynomial Order k

Generally, first-order spectral filters often yield simple monotonic frequency responses under common parameterizations, whereas higher-order polynomial filters enhance spectral expressiveness, mitigating the low-pass bias responsible for OSM. Based on this observation, we first provide a formal definition of OSM and then present a rigorous analysis to characterize its relationship with the polynomial order k.

Definition 1 (OSM). Let $h_p^{(l)} \in \mathbb{R}^{f'}$ denote the representation of node p at layer l. The OSM phenomenon occurs when the network depth increases, and all node representations tend to converge to the same vector, i.e.,

$$\lim_{l \to \infty} \|h_p^{(l)} - h_q^{(l)}\|_2 = 0, \quad \forall\, p, q \in V. \tag{10}$$

In this case, node representations discard their discriminative power and become indistinguishable.

Theorem 1. *Increasing the polynomial order k of a spectral graph filter can help mitigate OSM under the following mild assumptions: (i) the normalized graph Laplacian $\hat{L}$ has a finite spectrum; and (ii) there exists at least one discriminative spectral component whose contribution induces a nonzero node-wise separation.*

Proof. OSM fundamentally arises from the selective suppression of discriminative spectral components by graph convolutional filtering, and stacking layers further amplifies this effect. To reveal the underlying mechanism, we analyze a single spectral filtering operation and investigate how the polynomial order k affects the preservation of discriminative information. For analytical simplicity, we omit the nonlinear activation function and the feature transformation matrix W, and consider a single-channel input signal $x \in \mathbb{R}^n$. Let $\hat{x} = U^\top x$. Then Eq. (1) reduces to

$$h = U \operatorname{diag}\big(g_k(\boldsymbol{\lambda})\big)\, U^\top x = \sum_{i=1}^{n} g_k(\lambda_i)\, \hat{x}_i\, u^{(i)}, \tag{11}$$

where $u^{(i)}$ is the i-th eigenvector of $\hat{L}$. Let $\mathcal{S} \subseteq \{1, \ldots, n\}$ denote the set of discriminative spectral components, and define the ideal target response r_i as below

$$r_i = \begin{cases} 1, & i \in \mathcal{S}, \\ 0, & i \notin \mathcal{S}. \end{cases} \tag{12}$$

For any pair of nodes $p, q \in V$, define M and E as

$$M_{p,q} = \sum_{i \in \mathcal{S}} \hat{x}_i \big(u_p^{(i)} - u_q^{(i)}\big), \tag{13}$$

$$E_{p,q}^{(k)} = \sum_{i=1}^{n} \big(g_k(\lambda_i) - r_i\big) \hat{x}_i \big(u_p^{(i)} - u_q^{(i)}\big), \tag{14}$$

so that

$$h_p - h_q = M_{p,q} + E^{(k)}_{p,q}. \tag{15}$$

By the non-degeneracy assumption, there exists at least one pair of nodes p, q such that $M_{p,q} \neq 0$. Define $\Delta_{p,q} := |M_{p,q}| > 0$. For the error term, let $\delta_k = \max_{1\le i\le n} |g_k(\lambda_i) - r_i|$. Since the eigenvectors are normalized, we have $|u_p^{(i)} - u_q^{(i)}| \le 2$ for any i. Therefore,

$$|E^{(k)}_{p,q}| \le 2\delta_k \sum_{i=1}^{n} |\hat{x}_i| \le 2\delta_k \sqrt{n}\, \|x\|_2. \tag{16}$$

Since the spectrum of $\hat{L}$ contains only finitely many eigenvalues, polynomial interpolation on the discrete spectral set guarantees that there exists a polynomial filter g_k of sufficiently large order k such that $\delta_k \le \varepsilon$ for any $\varepsilon > 0$. So

$$|h_p - h_q| \ge \Delta_{p,q} - 2\varepsilon\sqrt{n}\, \|x\|_2. \tag{17}$$

In particular, once k is large enough such that $2\varepsilon\sqrt{n}\, \|x\|_2 < \Delta_{p,q}$, we obtain $|h_p - h_q| > 0$. This result indicates that sufficiently high-order polynomial spectral filters preserve discriminative spectral components and prevent node representations from collapsing to identical embeddings after graph convolution. Therefore, increasing the polynomial order k helps mitigate OSM. □

4.2 Analyzing OSQ with Polynomial Order k

Within the unified spectral-spatial framework, the polynomial order k determines the number of node hops whose features are aggregated during graph convolution. As k increases, the convolution aggregates information from more distant neighbors, thereby expanding the spatial receptive field of each node. However, under fixed-dimensional node embeddings, this expansion can intensify information compression and lead to OSQ. In this part, we first provide a formal definition of OSQ and then analyze how the polynomial order affects it.

Definition 2 (OSQ). Let $h_p \in \mathbb{R}^{f'}$ be the representation of node p after graph convolution, measuring the sensitivity of h_p to the input feature of node q by the node-wise Jacobian

$$J_{p\leftarrow q} = \left\| \frac{\partial h_p}{\partial x_q} \right\|_2, \tag{18}$$

OSQ occurs when $J_{p\leftarrow q}$ is close to zero for a non-negligible fraction of node pairs $(p, q) \in V \times V$, indicating that information from node q is excessively compressed and thus fails to meaningfully influence the representation of node p.

Theorem 2. *Increasing the polynomial order k enlarges the spatial receptive field and can exacerbate the OSQ issue.*

Proof. To isolate the intrinsic compression mechanism of OSQ, we omit the nonlinear activation and analyze a single graph convolution layer. For theoretical clarity, we consider a connected non-bipartite c-regular graph, which is commonly used in prior theoretical studies on OSQ [Nguyen *et al.*, 2023; Karhadkar *et al.*, 2023; Dai *et al.*, 2023]. After adding self-loops, the augmented graph becomes $(c+1)$-regular, so $\tilde{D} = (c+1)I$ and $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} = \tilde{D}^{-1}\tilde{A} = P$, where P is the random-walk transition matrix on the augmented graph. From the spatial form in Eq. (9), we have

$$h = \sum_{s=0}^{k} \beta_s \hat{A}^s X W. \tag{19}$$

For any pair of nodes (p, q), let $\mathrm{dist}(p, q) = d$ be the shortest-path distance. Since $[\hat{A}^s]_{pq} = 0$ for all $s < d$, we obtain

$$J_{p\leftarrow q} = \left\| \frac{\partial h_p}{\partial x_q} \right\|_2 \le \|W\|_2 \sum_{s=d}^{k} |\beta_s|\, |[\hat{A}^s]_{pq}|. \tag{20}$$

By the classical spectral mixing theorem for random walks [Chung, 1997], there exist constants $C_0 > 0$ and $\rho \in (0, 1)$ such that

$$\left| [\hat{A}^s]_{pq} - 1/n \right| = |[P^s]_{pq} - 1/n| \le C_0 \rho^s. \tag{21}$$

The stationary term $1/n$ is identical for all node pairs and is therefore non-discriminative. Further assume that the polynomial coefficients are bounded, i.e., $|\beta_s| \le B$ for all $s = 0, \ldots, k$. Substituting the above bound into Eq. (20) yields

$$\left\| \frac{\partial h_p}{\partial x_q} - \frac{1}{n}\Big(\sum_{s=d}^{k} \beta_s\Big) W \right\|_2 \le \|W\|_2 B C_0 \sum_{s=d}^{k} \rho^s \le C' \rho^d, \tag{22}$$

where $C' = \frac{\|W\|_2 B C_0}{1-\rho}$. This shows that, after removing the stationary averaging component, the node-specific sensitivity decays exponentially with the shortest-path distance d.

We now examine the spatial effect induced by increasing k. As k increases, progressively farther distance shells are incorporated into the receptive field. In a connected non-bipartite c-regular graph, the number of nodes in the distance-d shell grows on the order of c^d, whereas the above bound shows that the discriminative sensitivity of these nodes decays exponentially with d. Therefore, although higher-order convolution introduces exponentially more distant nodes, the node-specific information carried by each newly added node becomes exponentially weaker. As a result, increasing k rapidly enlarges the set of low-sensitivity node pairs, making many Jacobian entries $J_{p\leftarrow q}$ nearly negligible. In other words, the expansion of the receptive field does not yield a proportional gain in effective information; instead, it compresses an increasing number of weak signals into the same fixed-dimensional representation h_p, thereby intensifying information mixing and aggravating the OSQ issue. □

4.3 Trade-Off Insights between OSM and OSQ

To illustrate the impact of the polynomial order k on node representations, Figure 1 provides a schematic example. Here, $k = 1, 2, 3$ are used to represent low, medium, and high polynomial orders. The upper row depicts the spectral responses of polynomial filters at different orders. As k increases, the filter exhibits greater spectral expressiveness, which helps mitigate OSM, consistent with Theorem 1. The lower row shows the corresponding spatial receptive fields. Higher-order polynomials enlarge the receptive field, allowing information aggregation from more distant neighbors and potentially aggravating OSQ, as established by Theorem 2.

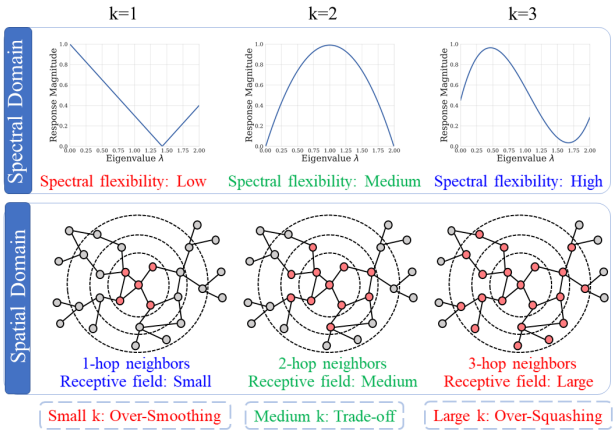


Figure 1: Illustration of the trade-off induced by polynomial order between spectral expressiveness and spatial aggregation.

This analysis reveals the dual role of k in governing OSM and OSQ. Specifically, a small k limits spectral expressiveness, causing node representations to become overly similar and leading to OSM. In contrast, an excessively large k overly expands the receptive field, forcing weak long-range signals to be compressed into fixed-dimensional representations and thereby aggravating OSQ. Therefore, a moderate polynomial order enables a better balance between spectral discriminability and spatial locality, effectively alleviating both phenomena. In summary, selecting an appropriate k is crucial for balancing spectral filtering and spatial neighborhood aggregation. To validate this trade-off, we present experiments on the effect of polynomial order in Section 6.1.

5 The Proposed QS-GCN Model

In this section, we propose a Quadratic Spectral Graph Convolutional Network (QS-GCN) to balance OSM and OSQ, thereby enhancing the discriminative power of graph representations. Motivated by the theoretical analysis in Section 4.3 and later empirically validated in Section 6.1, we set the polynomial order to $k = 2$ in the proposed model. This choice balances spectral expressiveness and spatial locality, mitigating OSM while avoiding excessive OSQ. Specifically, QS-GCN consists of two main components: the Quadratic Spectral Graph Convolution (QS-Conv) and the Alignment-based Readout Module. The former provides node-level feature extraction with an interpretable trade-off between spectral expressiveness and spatial locality. The latter aligns node features into a fixed cluster order and maps the aligned representation to the final graph-level output via a shared Multi-Layer Perceptron (MLP).

5.1 The Quadratic Spectral Graph Convolution

The QS-Conv performs convolution on graph signals using a quadratic spectral polynomial. This design provides a flexible frequency response, which helps alleviate OSM, while restricting the convolutional scope to at most two hops and thus preventing excessive expansion of the receptive field. The operation is defined as

$$H = \sigma\Big((\alpha_0 I + \alpha_1 \hat{L} + \alpha_2 \hat{L}^2)XW\Big), \tag{23}$$

where $(\alpha_0, \alpha_1, \alpha_2)$ are learnable coefficients for the quadratic spectral terms.

k	MUTAG	PTCMR	PROTEINS	IMDB-B
0	86.81	60.25	74.66	72.79
1	85.58	60.55	74.57	71.30
2	**90.91**	**63.42**	**75.15**	**73.46**
3	82.89	61.92	74.03	73.26
4	85.34	57.65	73.49	72.94
5	84.21	54.28	72.32	70.29

Table 1: The graph classification accuracies (%) under different polynomial orders k.

5.2 The Alignment-based Readout Module

Although OSM and OSQ are typically discussed at the node level, they also directly affect graph-level tasks. Traditional readout functions, such as sum, mean, or max, are permutation-invariant but do not explicitly preserve structural correspondence across graphs. To address this issue, we propose an alignment-based readout module that reduces indiscriminate aggregation, preserves local structural distinctions, and produces order-consistent graph representations.

Specifically, let $H \in \mathbb{R}^{n\times d}$ denote the node feature matrix after the last graph convolution layer. We softly assign nodes to s clusters using a shared assignment function parameterized by $W_S \in \mathbb{R}^{d\times s}$, which produces an assignment probability matrix $S \in \mathbb{R}^{n\times s}$ as

$$S = \mathrm{softmax}(\hat{A}HW_S), \tag{24}$$

where softmax is applied row-wise so that each row of S forms a valid probability distribution over clusters. Using S, we align the node feature matrix H into fixed-order cluster representations as

$$\tilde{H} = S^\top H \in \mathbb{R}^{s\times d}. \tag{25}$$

Since $\tilde{H}$ is order-consistent across graph samples, we flatten it into a vector of dimension sd and feed it into an MLP to assign different importance weights to each cluster, producing the final graph-level representation as

$$y = \mathrm{MLP}(\mathrm{flatten}(\tilde{H})). \tag{26}$$

6 Experiments

In this section, we first explore the impact of the polynomial order k on graph classification performance to validate the theoretical analysis presented earlier. We then compare the effectiveness of our QS-GCN with representative graph kernels and GNNs on benchmark datasets.

6.1 Impact of Polynomial Order k

We investigate the effect of the polynomial order k on graph classification performance within the QS-GCN framework, where k is varied from 0 to 5. Experiments are conducted

	MUTAG	PTC_MR	PROTEINS	DD	IMDB-B	IMDB-M	REDB
Max #Nodes	28	64	620	5748	136	89	3782
Mean #Nodes	17.93	14.29	39.06	284.3	19.77	13	429.63
#Graphs	188	344	1113	1178	1000	1500	2000
#Classes	2	2	2	2	2	3	2
Domain	Bio	Bio	Bio	Bio	SN	SN	SN

Table 2: Statistics of benchmark graph classification datasets.

Method	MUTAG	PTC_MR	PROTEINS	DD	IMDB-B	IMDB-M	REDB
QS-GCN (Ours)	**90.91±0.61**	**63.42±1.42**	**75.15±0.21**	77.33±0.29	**73.46±1.40**	48.78±0.64	88.76±0.20
RWGK (2003)	80.77±0.72	55.91±0.37	74.20±0.40	71.70±0.47	67.94±0.77	46.72±0.30	72.73±0.39
SPGK (2005)	83.38±0.31	56.55±0.53	75.10±0.50	78.45±0.26	71.26±1.04	**51.33±0.57**	84.20±0.70
WLSK (2011)	82.88±0.57	56.05±0.51	73.52±0.43	**79.78±0.36**	71.88±0.77	49.50±0.49	76.56±0.30
JTQK (2014)	85.50±0.55	57.39±0.46	72.86±0.41	79.49±0.32	72.45±0.81	50.33±0.49	77.60±0.35
EDBMK (2021)	86.35	56.75	–	78.19	–	–	–
QBMK (2024)	88.55±0.43	59.38±0.36	–	77.60±0.47	–	–	–
AEGK (2025)	89.11±0.40	59.38±0.36	75.11±0.28	76.32±0.46	–	–	**88.86±0.12**

Table 3: Classification accuracy (% ± standard error) compared with graph kernel methods. The best results are in bold.

Dataset	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Cora	61.68	81.19	**82.62**	81.67	41.69	41.29
OGB-MolHIV	77.87	79.08	**79.76**	78.44	77.32	77.74

Table 4: The Performance under different polynomial orders k on additional benchmarks.

on representative graph classification datasets from the TUDataset. For each dataset, the model hyperparameters are kept fixed across different values of k. The detailed hyperparameter settings are provided in Section 6.2. We report the average results over 10-fold cross-validation.

As shown in Table 1, the polynomial order k has a substantial impact on classification performance. Across all datasets, $k = 2$ generally achieves the best results, in agreement with the trade-off analysis in Section 4.3. When k is too small, the graph filter has limited frequency selectivity and tends to bias toward low-frequency components, resulting in overly smooth node representations and degraded discriminability. In contrast, larger k increases spectral flexibility but simultaneously enlarges the spatial receptive field, compressing weak long-range signals into fixed-dimensional representations and aggravating OSQ. Overall, the quadratic setting achieves an effective balance between spectral discrimination and spatial locality, thereby mitigating both OSM and OSQ.

To further validate the generalizability of our method, we additionally evaluate it on the node classification dataset Cora and the large-scale molecular graph classification benchmark OGB-MolHIV (∼41k graphs), following the standard evaluation protocol used in GCN-based studies. We report classification accuracy (%) on Cora and ROC-AUC on OGB-MolHIV. The results in Table 4 exhibit the same trend as our theoretical analysis and the observations on the TUDataset.

6.2 Performance Evaluation of QS-GCN

We further evaluate the overall performance of the proposed QS-GCN model, considering classification accuracy and the ability to mitigate the OSM and OSQ problems. Specifically, experiments are conducted on multiple benchmark graph classification datasets covering two major application domains: bioinformatics (Bio) and social networks (SN). Detailed statistics of the datasets are summarized in Table 2. All experiments are implemented using PyTorch and PyTorch Geometric and are executed on a single NVIDIA GeForce RTX 3090 GPU with 24GB of VRAM.

We initialize the polynomial spectral filters as band-pass filters whose peak frequency response is located in the mid-frequency region ($\lambda \approx 1$). Specifically, for $k = 2$, the coefficients are designed to produce a unimodal spectral response, given by $g(\lambda) = -\lambda^2+2\lambda$. For higher-order polynomials, the coefficients are extended into a smoothly varying oscillatory pattern. The number of convolutional layers is set between 2 and 4, the node embedding dimension is fixed as 64, and the number of clusters in the readout stage is set to 8. We adopt the ReLU activation function and optimize the models using Adam. The learning rate is selected from $\{0.001, 0.005\}$, the weight decay from $\{10^{-5}, 10^{-4}, 10^{-3}\}$, and the batch size is set to 64. Depending on the dataset, the number of training epochs ranges from 100 to 1000. To ensure statistical robustness, we perform 10 independent repetitions of 10-fold cross-validation and report the mean classification accuracy and the standard error. Code is available on GitHub.[1]

Comparison on Graph Classification Performance. We compare QS-GCN with a range of representative graph kernel methods and GNNs. The graph kernels include RWGK [Gartner *et al.*, 2003], SPGK [Borgwardt and Kriegel, 2005], WLSK [Shervashidze *et al.*, 2011], JTQK [Bai *et al.*, 2014] (with $q = 2$), EDBMK [Xu *et al.*, 2021], QBMK [Bai *et al.*, 2024], and AEGK [Bai *et al.*, 2025]. The GNNs include DGCNN [Zhang *et al.*, 2019], DiffPool [Ying *et al.*, 2018], ECC [Simonovsky and Komodakis, 2017], GIN [Xu *et al.*, 2019], GraphSAGE [Hamilton *et al.*, 2017], S^2GCN [Zhu and Koniusz, 2021], JacobiConv [Wang and Zhang, 2022], MidGCN [Huang *et al.*, 2023], GWRNN [Nikolentzos and Vazirgiannis, 2023], Co-GNN [Finkelshtein *et al.*, 2024],

[1]https://github.com/Xiaoqin0421/QS-GCN

Method	MUTAG	PTC_MR	PROTEINS	DD	IMDB-B	IMDB-M	REDB
QS-GCN (Ours)	**90.91±0.61**	**63.42±1.42**	75.15±0.21	**77.33±0.29**	**73.46±1.40**	48.78±0.64	88.76±0.20
DGCNN (2019)	84.0±6.7	58.3±7.0	72.9±3.5	76.6±4.3	69.2±3.0	45.6±3.4	87.8±2.5
DiffPool (2018)	79.8±7.1	60.8±7.0	73.7±3.5	75.0±3.5	68.4±3.3	45.6±3.4	89.4±1.6
ECC (2017)	75.4±6.2	55.7±3.3	72.3±3.4	72.69±4.1	67.7±2.8	43.5±3.1	–
GIN (2019)	84.7±6.7	58.8±5.5	73.3±4.0	75.3±2.9	71.2±3.9	48.5±3.3	89.9±1.9
GraphSAGE (2017)	83.6±9.6	60.1±4.7	73.0±4.5	72.9±2.0	68.8±4.5	47.6±3.5	84.3±1.9
S^2GCN (2021)	75.29	59.76	69.97	64.43	67.78	46.81	70.80
JacobiConv (2022)	85.52	58.61	63.30	68.46	67.10	39.39	67.45
MidGCN (2023)	81.92	59.95	70.67	69.06	66.60	46.57	74.10
GWRNN (2023)	83.4±5.6	–	74.9±3.5	75.6±4.6	72.8±4.2	**49.0±2.9**	**90.0±1.8**
Co-GNN (2024)	–	–	73.1±2.3	–	70.8±3.3	48.5±4.0	88.6±2.2
LASER (2024)	82.20	–	74.39	–	64.33	–	85.46
GNN-VPA (2024)	76.10	61.30	73.90	–	71.70	46.70	85.50
CGP (2025)	83.75	–	73.04	–	56.20	–	67.05
GKNN-WL (2025)	85.73±2.70	58.29±2.54	74.94±1.10	–	69.70±2.20	47.87±1.78	–
GKNN-GL (2025)	85.24±2.28	60.13±1.94	**75.36±1.12**	–	69.90±1.44	45.67±1.22	–

Table 5: Classification accuracy (% ± standard error) compared with GNN-based methods. The best results are in bold.

LASER [Barbero *et al.*, 2024], GNN-VPA [Schneckenreiter *et al.*, 2024], CGP [Wilson *et al.*, 2025], GKNN-WL, and GKNN-GL [Cosmo *et al.*, 2025]. For DGCNN, DiffPool, ECC, GIN, and GraphSAGE, we adopt the results from [Errica *et al.*, 2020] or evaluate them under the same experimental settings for datasets not originally reported. For S^2GCN [Zhu and Koniusz, 2021], JacobiConv [Wang and Zhang, 2022], and MidGCN [Huang *et al.*, 2023], we re-implement the models under the same experimental settings for fair comparison. For the remaining methods, we report results as presented in their original publications.

Table 3 and 5 report the classification accuracy and standard error on each dataset. Overall, QS-GCN achieves the best or highly competitive performance on most benchmarks, demonstrating its effectiveness for graph classification. This advantage mainly stems from two aspects. First, the QS-Conv employs a learnable quadratic spectral polynomial filter that captures information across different frequency bands, enhancing spectral expressiveness, while the limited 2-hop propagation helps alleviate OSQ. Second, the alignment-based readout module performs structured cluster alignment and aggregation, effectively integrating local node features with global graph structure to improve discriminative capability. To provide spatial interpretability, Figure 2 illustrates the attention weights assigned by a 3-layer QS-GCN to multi-hop neighbors of a central node on a sample graph from the MUTAG dataset. The results show that the proposed QS-GCN can dynamically adjust the importance of multi-hop information across layers.

Analysis of Over-Smoothing (OSM). To quantify the diversity of graph representations, we adopt the *Average Cosine Distance (AD)* as the evaluation metric. Given a graph-level representation matrix $H \in \mathbb{R}^{N\times d}$, the AD value is defined as

$$\mu(H) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - \frac{H_i^\top H_j}{\|H_i\|\,\|H_j\|}\right), \qquad (27)$$

where H_i and H_j denote the embeddings of the i-th and j-th graphs. A higher AD value indicates greater representation diversity and thus reduced OSM. To examine robustness to network depth, we compare QS-GCN with GCN, GIN, and GraphSAGE on four representative datasets, varying the number of layers over $\{2, 4, 8, 16, 32, 64\}$. Both classification accuracy and AD are reported in Figures 3. As depth increases, QS-GCN exhibits significantly less performance degradation and consistently maintains higher AD values, whereas the baselines suffer sharp AD drops, indicating severe OSM. These results show that QS-GCN effectively alleviates OSM and remains robust under increased depth.

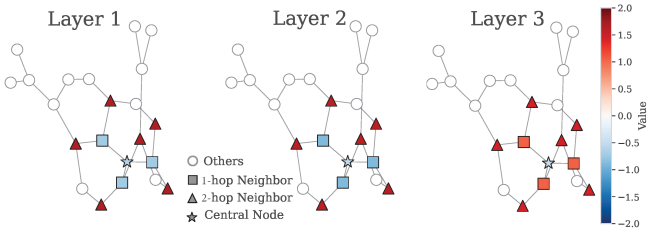
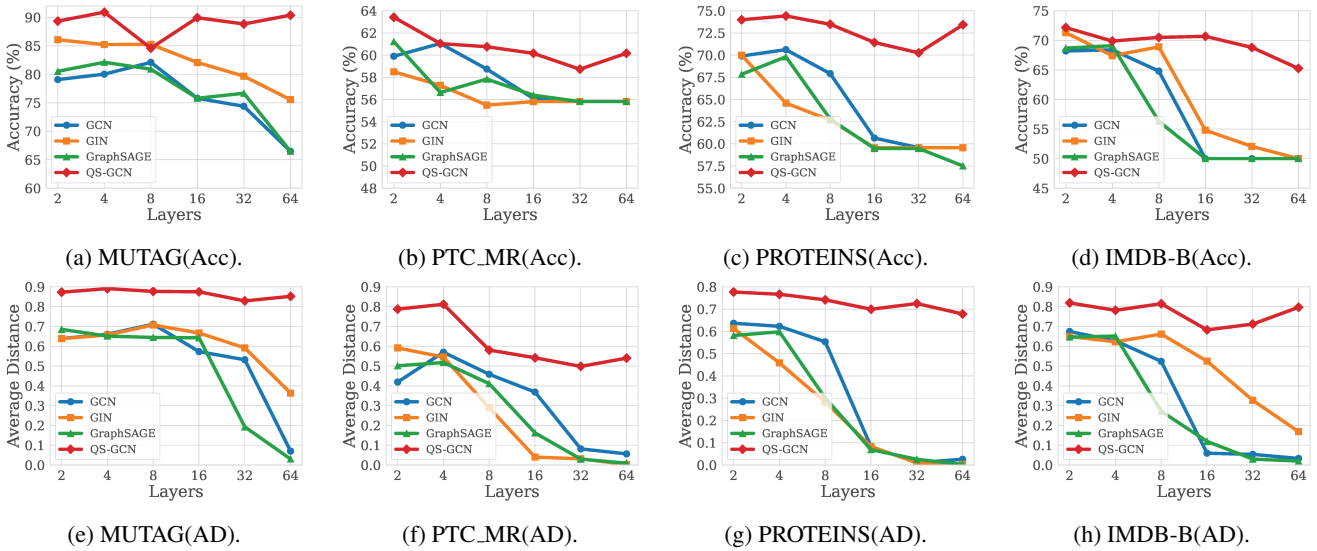


Figure 2: An example of adaptive weight distribution.

Method	MUTAG	PROTEINS	IMDB_B	REDB
QS-GCN	**90.91±0.61**	75.15±0.21	**73.46±1.40**	88.76±0.20
FA	83.45±1.74	72.30±0.67	71.48±0.88	90.22±0.48
SDRF	82.70±1.78	70.92±0.79	70.21±0.81	86.83±0.52
FoSR	86.15±1.49	75.25±0.86	71.96±0.69	**90.94±0.47**
GTR	86.10±1.76	**75.78±0.76**	71.49±0.93	90.41±0.41

Table 6: Classification accuracy (% ± standard error) for comparison with GNNs addressing the OSQ issue.

Analysis of Over-Squashing (OSQ). We evaluate the effectiveness of QS-GCN in mitigating the over-squashing problem by comparing our method with several GNNs specifically designed to address OSQ, including FA [Alon and Yahav, 2021], SDRF [Topping *et al.*, 2022], FoSR [Karhadkar *et al.*, 2023], and GTR [Black *et al.*, 2023], and report the results as presented in their original papers. As summarized in



(a) MUTAG(Acc). (b) PTC_MR(Acc). (c) PROTEINS(Acc). (d) IMDB-B(Acc).

(e) MUTAG(AD). (f) PTC_MR(AD). (g) PROTEINS(AD). (h) IMDB-B(AD).

Figure 3: Classification accuracies (%) and the AD values on the four datasets.

k	Readout	MUTAG	PTC_MR	PROTEINS	IMDB-B
1	Align	85.58	60.55	74.57	71.30
1	Avg	83.15	59.17	73.13	69.40
2	Align	**89.33**	**63.42**	**75.15**	**73.46**
2	Avg	86.63	55.81	71.78	71.80
3	Align	82.89	61.92	74.21	73.26
3	Avg	82.15	60.94	70.25	70.60

Table 7: Ablation study of alignment-based readout (Align) and average readout (Avg) under different polynomial orders k.

Table 6, QS-GCN achieves competitive performance among these methods, suggesting its effectiveness in alleviating OSQ while preserving strong predictive capability.

Ablation Study on Alignment-based Readout. To further isolate the contribution of the alignment-based readout module, we replace it with a standard average readout and evaluate both variants on MUTAG, PTC_MR, PROTEINS, and IMDB-B under different polynomial orders. The results are reported in Table 7. A consistent trend is observed across all datasets: regardless of the readout strategy, the performance first improves and then declines as k increases, with the best results consistently achieved at $k = 2$. This suggests that the spectral design governed by the quadratic polynomial plays the dominant role in balancing OSM and OSQ. Moreover, under the same polynomial order, the alignment-based readout consistently outperforms average readout, indicating its effectiveness in further enhancing graph-level representation learning through more discriminative structural preservation.

Computational Complexity. Finally, we analyze the computational complexity of QS-GCN. Let n denote the number of nodes, m the number of edges, d the feature dimension, s the number of clusters, l the number of layers, and k the polynomial order. Under a recursive propagation-based implementation, the QS-Conv requires $O(kmdl)$ complexity. The readout module aligns node representations into s clusters followed by MLP aggregation, with complexity $O(nds)$. Therefore, the overall complexity of QS-GCN is $O(kmdl + nds)$, which scales approximately linearly with k when other hyperparameters are fixed. To validate this analysis, we further measure runtime per epoch and GPU memory usage under different polynomial orders on the MUTAG dataset. As summarized in Table 8, the empirical results confirm the predicted scaling behavior, showing that increasing k introduces only moderate computational overhead.

k	0	1	2	3	4	5
Memory (MB)	362	368	370	372	373	374
Runtime (ms)	5.2	9.8	11.4	14.3	16.9	18.2

Table 8: Empirical runtime and memory consumption under different polynomial orders k on MUTAG.

7 Conclusion

This work investigates the phenomena of OSM and OSQ in GNNs from a unified spectral-spatial perspective. We first develop a unified spectral-spatial framework for analyzing polynomial graph convolution, providing a principled understanding of how the polynomial order k jointly affects spectral filtering and spatial information propagation. Based on this framework, we reveal an intrinsic trade-off induced by k, where higher-order filters enhance spectral expressiveness and help alleviate OSM, while simultaneously expanding the receptive field and exacerbating information compression, thereby intensifying OSQ. Motivated by this insight, we propose the QS-GCN, which achieves an effective balance between these two phenomena. Extensive experiments validate both our theoretical analysis and the effectiveness of the proposed model. Future work will explore more principled and generalizable metrics for quantifying OSQ, as well as adaptive, data-dependent selection of the polynomial order k to further improve the balance between OSM and OSQ.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No. 62576371, 62576198, and 62536006), and the Open Project Foundation of Key Laboratory of Computation Intelligence and Chinese Information Processing of Ministry of Education and Key Laboratory of Data Intelligence and Cognitive Computing of Shanxi Province.

References

[Alon and Yahav, 2021] Uri Alon and Eran Yahav. On the Bottleneck of Graph Neural Networks and its Practical Implications. In *Proceedings of ICLR*, 2021.

[Bai *et al.*, 2014] Lu Bai, Luca Rossi, Horst Bunke, and Edwin R. Hancock. Attributed Graph Kernels Using the Jensen-Tsallis q-Differences. In *Proceedings of ECML-PKDD*, pages 99–114, 2014.

[Bai *et al.*, 2024] Lu Bai, Lixin Cui, Ming Li, Yue Wang, and Edwin Hancock. QBMK: Quantum-based Matching Kernels for Un-attributed Graphs. In *Proceedings of ICML*, pages 2364–2374, 2024.

[Bai *et al.*, 2025] Lu Bai, Lixin Cui, Ming Li, Peng Ren, Yue Wang, Lichi Zhang, Philip S. Yu, and Edwin R. Hancock. AEGK: Aligned Entropic Graph Kernels Through Continuous-Time Quantum Walks. *IEEE Trans. Knowl. Data Eng.*, pages 1064–1078, 2025.

[Balcilar *et al.*, 2021] Muhammet Balcilar, Guillaume Renton, Pierre Héroux, Benoit Gaüzère, Sébastien Adam, and Paul Honeine. Analyzing the Expressive Power of Graph Neural Networks in a Spectral Perspective. In *Proceedings of ICLR*, 2021.

[Barbero *et al.*, 2024] Federico Barbero, Ameya Velingker, Amin Saberi, Michael M. Bronstein, and Francesco Di Giovanni. Locality-Aware Graph Rewiring in GNNs. In *Proceedings of ICLR*, 2024.

[Black *et al.*, 2023] Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. Understanding Oversquashing in GNNs through the Lens of Effective Resistance. In *Proceedings of ICML*, pages 2528–2547, 2023.

[Borgwardt and Kriegel, 2005] K.M. Borgwardt and H.P. Kriegel. Shortest-Path Kernels on Graphs. In *Proceedings of ICDM*, pages 74–81, 2005.

[Bruna *et al.*, 2013] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral Networks and Locally Connected Networks on Graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[Chen *et al.*, 2020a] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and Relieving the Over-Smoothing Problem for Graph Neural Networks from the Topological View. In *Proceedings of AAAI*, pages 3438–3445, 2020.

[Chen *et al.*, 2020b] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and Deep Graph Convolutional Networks. In *Proceedings of ICML*, pages 1725–1735, 2020.

[Chung, 1997] Fan R. K. Chung. *Spectral Graph Theory*, volume 92. Amer. Math. Soc., 1997.

[Cosmo *et al.*, 2025] Luca Cosmo, Giorgia Minello, Alessandro Bicciato, Michael M. Bronstein, Emanuele Rodolà, Luca Rossi, and Andrea Torsello. Graph Kernel Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.*, pages 6257–6270, 2025.

[Dai *et al.*, 2023] Shi Dai, Andi Han, Lequan Lin, Yi Guo, and Junbin Gao. Exposition on Over-Squashing Problem on GNNs: Current Methods, Benchmarks and Challenges. *arXiv preprint arXiv:2311.07073*, 2023.

[Defferrard *et al.*, 2016] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Proceedings of NeurIPS*, pages 3844–3852, 2016.

[Errica *et al.*, 2020] Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. A Fair Comparison of Graph Neural Networks for Graph Classification. In *Proceedings of ICLR*, 2020.

[Finkelshtein *et al.*, 2024] Ben Finkelshtein, Xingyue Huang, Michael Bronstein, and İsmail İlkan Ceylan. Cooperative Graph Neural Networks. In *Proceedings of ICML*, 2024.

[Gartner *et al.*, 2003] Thomas Gartner, Peter A. Flach, and Stefan Wrobel. On Graph Kernels: Hardness Results and Efficient Alternatives. In *Proceedings of COLT*, pages 129–143, 2003.

[Hamilton *et al.*, 2017] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive Representation Learning on Large Graphs. In *Proceedings of NeurIPS*, pages 1025–1035, 2017.

[Huang *et al.*, 2023] Jincheng Huang, Lun Du, Xu Chen, Qiang Fu, Shi Han, and Dongmei Zhang. Robust Mid-Pass Filtering Graph Convolutional networks. In *Proceedings of WWW*, pages 328–338, 2023.

[Jamadandi *et al.*, 2024] Adarsh Jamadandi, Celia Rubio-Madrigal, and Rebekka Burkholz. Spectral Graph Pruning Against Over-Squashing and Over-Smoothing. In *Proceedings of NeurIPS*, 2024.

[Karhadkar *et al.*, 2023] Kedar Karhadkar, Pradeep Kr. Banerjee, and Guido Montúfar. FoSR: First-Order Spectral Rewiring for Addressing Oversquashing in GNNs. In *Proceedings of ICLR*, 2023.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of ICLR*, 2017.

[Li *et al.*, 2018] Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. In *Proceedings of AAAI*, 2018.

[Li *et al.*, 2023] Guohao Li, Matthias Müller, Guocheng Qian, Itzel C. Delgadillo, Abdulellah Abualshour, Ali Thabet, and Bernard Ghanem. DeepGCNs: Making GCNs Go as Deep as CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 6923–6939, 2023.

[Min *et al.*, 2021] Shengjie Min, Zhan Gao, Jing Peng, Liang Wang, Ke Qin, and Bo Fang. STGSN - a Spatial-Temporal Graph Neural Network framework for Time-Evolving Social Networks. *Knowl.-Based Syst.*, 2021.

[Nguyen *et al.*, 2023] Kha Dang Nguyen, Tan Minh Nguyen, Nhat Ho, Khuong N. Nguyen, Hieu Nong, and Vinh Phu Nguyen. Revisiting Over-smoothing and Over-squashing using Ollivier-Ricci Curvature. In *Proceedings of ICML*, 2023.

[Nikolentzos and Vazirgiannis, 2023] Giannis Nikolentzos and Michalis Vazirgiannis. Geometric Random Walk Graph Neural Networks via Implicit Layers. In *Proceedings of AISTATS*, 2023.

[NT and Maehara, 2019] Hoang NT and Takanori Maehara. Revisiting Graph Neural Networks: All We Have is Low-Pass Filters. *arXiv preprint arXiv:1905.09550*, 2019.

[Rong *et al.*, 2020] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. DropEdge: Towards Deep Graph Convolutional Networks on Node Classification. In *Proceedings of ICLR*, 2020.

[Rusch *et al.*, 2022a] T. Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-Coupled Oscillator Networks. In *Proceedings of ICML*, pages 18888–18909, 2022.

[Rusch *et al.*, 2022b] T. Konstantin Rusch, Benjamin Paul Chamberlain, Michael W. Mahoney, Michael M. Bronstein, and Siddhartha Mishra. Gradient Gating for Deep Multi-Rate Learning on Graphs. In *Proceedings of ICLR*, 2022.

[Rusch *et al.*, 2023] T. Konstantin Rusch, Michael M. Bronstein, and Siddhartha Mishra. A Survey on Oversmoothing in Graph Neural Networks. *arXiv preprint arXiv:2303.10993*, 2023.

[Schneckenreiter *et al.*, 2024] Lisa Schneckenreiter, Richard Freinschlag, Florian Sestak, Johannes Brandstetter, Günter Klambauer, and Andreas Mayr. GNN-VPA: A Variance-Preserving Aggregation Strategy for Graph Neural Networks. In *Proceedings of ICLR Tiny Papers Track*, 2024.

[Shervashidze *et al.*, 2011] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler-Lehman Graph Kernels. *J. Mach. Learn. Res.*, pages 2539–2561, 2011.

[Simonovsky and Komodakis, 2017] Martin Simonovsky and Nikos Komodakis. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. In *Proceedings of CVPR*, pages 29–38, 2017.

[Singh, 2025] Akansha Singh. Over-Squashing in Graph Neural Networks: A Comprehensive Survey. *Neurocomputing*, 2025.

[Topping *et al.*, 2022] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding Over-Squashing and Bottlenecks on Graphs via Curvature. In *Proceedings of ICLR*, 2022.

[Wang and Zhang, 2022] Xiyuan Wang and Muhan Zhang. How Powerful are Spectral Graph Neural Networks. In *Proceedings of ICML*, pages 23341–23362, 2022.

[Wang *et al.*, 2023] Yu Wang, Yuying Zhao, Yi Zhang, and Tyler Derr. Collaboration-Aware Graph Convolutional Networks for Recommendation Systems. In *Proceedings of WWW*, pages 91–101, 2023.

[Wilson *et al.*, 2025] J. J. Wilson, Maya Bechler-Speicher, and Petar Veličković. Cayley Graph Propagation. In *Proceedings of LoG*, 2025.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks? In *Proceedings of ICLR*, 2019.

[Xu *et al.*, 2021] Lixiang Xu, Lu Bai, Xiaoyi Jiang, Ming Tan, Daoqiang Zhang, and Bin Luo. Deep Rényi Entropy Graph Kernel. *Pattern Recognit.*, 2021.

[Yan *et al.*, 2023] Yujun Yan, Gao Li, and Danai Koutra. Size Generalizability of Graph Neural Networks on Biological Data: Insights and Practices from the Spectral Perspective. *arXiv preprint arXiv:2305.15611*, 2023.

[Ying *et al.*, 2018] Rex Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. Hierarchical Graph Representation Learning with Differentiable Pooling. In *Proceedings of NeurIPS*, page 4805–4815, 2018.

[Zhang *et al.*, 2019] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An End-to-end Deep Learning Architecture for Graph Classification. In *Proceedings of AAAI*, pages 4438–4445, 2019.

[Zhao and Akoglu, 2019] Lingxiao Zhao and Leman Akoglu. PairNorm: Tackling Oversmoothing in GNNs. *arXiv preprint arXiv:1909.12223*, 2019.

[Zhu and Koniusz, 2021] Hao Zhu and Piotr Koniusz. Simple Spectral Graph Convolution. In *Proceedings of ICLR*, 2021.