Graph Mediator Networks: Bridging Local and Global Semantics via Serial Message Passing

Jiangfeng Sun sun2017@bupt.edu.cn
SiHao He sihaohe@bupt.edu.cn
Yanlong Lin starim@bupt.edu.cn
Zhonghong Ou zhonghong.ou@bupt.edu.cn
Meina Song mnsong@bupt.edu.cn
School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China

Editors: Hung-yi Lee and Tongliang Liu

Abstract

Graph Neural Networks (GNNs) have achieved remarkable success in modeling structured data through local message passing. However, their effectiveness diminishes on graphs with low homophily or irregular structures, where long-range dependencies are hard to capture and features tend to suffer from over-smoothing and noise amplification. To address these limitations, we propose GMN, a novel dual-path Graph Mediator Network that explicitly enhances both global information propagation and spectral stability. In the spatial path, GMN introduces a lightweight Mediator node connected to all graph nodes, allowing long-range communication to occur in a single hop without increasing network depth. In parallel, the spectral path leverages multi-scale Chebyshev filtering along with a spectral energy regularization term that suppresses high-frequency noise, leading to smoother and more stable node embeddings. These two complementary pathways are adaptively integrated via a gated fusion mechanism, which dynamically balances their contributions based on structural context. Final graph-level representations are obtained through task-specific pooling strategies, enabling GMN to generalize effectively across different tasks. Extensive experiments on benchmark datasets with varying homophily levels and structural perturbations demonstrate that GMN consistently achieves state-of-the-art performance in terms of accuracy, robustness, and generalization. Code is available at: https://github.com/sun2017bupt/GMN.

Keywords: Graph Neural Networks, Long-Range Dependency Modeling.

1. Introduction

Graph-structured data is ubiquitous in domains ranging from social networks and molecular chemistry to biological interaction systems. To extract rich, task-relevant representations from such data, Graph Neural Networks (GNNs) have emerged as the leading framework, achieving state-of-the-art results in a variety of prediction tasks. Prominent GNN architectures—Graph Convolutional Networks (GCNs) Kipf and Welling (2016), Graph Attention Networks (GATs) Veličković et al. (2017)—employ a local message-passing paradigm in which each node iteratively aggregates and transforms features from its immediate neighbors to build expressive embeddings.

© 2025 J. Sun, S. He, Y. Lin, Z. Ou & M. Song.

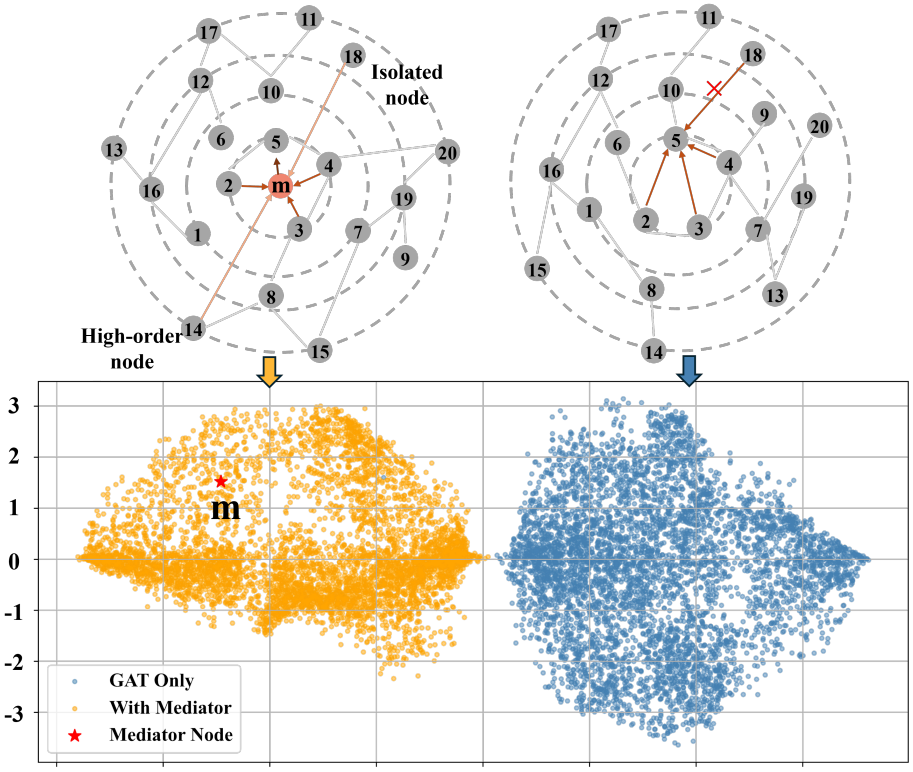

Figure 1: **Motivation of GMN. Top**: Message propagation comparison. **Left**: GMN introduces a Mediator node (red star) for efficient one-hop global communication. **Right**: Standard GNNs rely on multi-hop local propagation, which may cause information bottlenecks. **Bottom**: t-SNE visualization of node embeddings from GAT (blue) and GMN (orange), where GMN exhibits clearer class separation. In addition to the Mediator, GMN incorporates spectral energy regularization to enhance representation stability.

However, the inductive bias toward local neighborhood aggregation in most GNNs inherently constrains their ability to capture high-order and long-range dependencies. In heterophilic graphs—where adjacent nodes often belong to different classes or exhibit dissimilar features—this bias propagates noisy or conflicting information, diluting the discriminative power of node embeddings. Deepening the network only aggravates the effect, leading to over-smoothing in which embeddings converge to nearly identical values Oono and Suzuki (2019). These shortcomings undermine both the expressive capacity and generalization of GNNs, highlighting the need for mechanisms that explicitly enable global communication while preserving feature integrity.

To address these limitations, we propose GMN, a unified framework that combines explicit global routing with frequency-aware regularization to learn robust graph representations. Figure 1 provides an overview of our design, which stems from two complementary insights. The first insight addresses the inherent locality bias in standard GNNs, which

confines message passing to each node's immediate neighbors and thus struggles to capture high-order dependencies in sparse or heterophilic graphs. To overcome this, GMN introduces Mediator node, which connects to every original vertex, forming a star-like topology that approximates global connectivity. In each propagation step, any two vertices can exchange information via a Mediator node in at most two hops distance (≤ 2 hops), thereby expanding the receptive field without increasing network depth. This global relay mitigates over-squashing of long-range signals and preserves discriminative features, while the additional $\mathcal{O}(|V|)$ edges incur only linear computational and memory overhead. The second insight adopts a spectral perspective on graph signals, where the graph Laplacian's eigenmodes decompose node features into low- and high-frequency components. Low-frequency modes capture smooth, global structures, whereas high-frequency modes reflect local variations and are prone to noise amplification Shuman et al. (2013); Chien et al. (2020). To enforce spectral stability, GMN introduces a spectral energy regularization that penalizes the magnitude of high-frequency activations in the node embeddings. By suppressing these unstable components, the model yields smoother representations, reduces sensitivity to structural perturbations, and improves generalization across graphs with diverse topologies. These complementary mechanisms are integrated within a dual-path architecture. In the spatial path, standard message passing is performed on the mediator-augmented graph, enabling both local neighborhood aggregation and direct global routing through the mediator node. Simultaneously, the spectral path applies multi-scale filtering—implemented via a Chebyshev polynomial expansion—to extract structural patterns across a range of frequency bands. Outputs from the two paths are then merged by a learnable gated fusion module that adaptively balances spatial and spectral information. Finally, a global pooling layer aggregates the fused features into a compact graph-level embedding for downstream prediction. In the next section, we describe each component in detail and outline the end-to-end training procedure.

The main contributions are summarized as follows:

- Mediator Node for Global Routing: We introduce a Mediator node connected to all graph vertices, enabling direct long-range message propagation without increasing network depth. This design effectively mitigates locality bottlenecks and over-squashing, enhancing the model's ability to capture high-order dependencies.

- Spectral Energy Regularization: We propose a regularization term that penalizes high-frequency components in the spectral domain, promoting smoother node embeddings and improving robustness to structural noise and perturbations.

- Dual-Path Architecture with Gated Fusion: GMN employs a dual-path architecture that combines spatial aggregation on the Mediator-augmented graph with multi-scale spectral filtering. A learnable gated mechanism adaptively fuses these two streams, yielding expressive and balanced representations.

- Comprehensive Empirical Evaluation: We conduct extensive experiments on benchmarks with varying levels of homophily and graph scales. The results demonstrate that GMN achieves superior accuracy, robustness, and generalization compared to state-of-the-art baselines.

2. Related Work

Graph Neural Networks (GNNs) are a leading approach for representation learning on graphs, with notable examples including GCNs Kipf and Welling (2016) and GATs Veličković et al. (2017). Most rely on local message passing, where nodes aggregate features from immediate neighbors. While effective for local structure, this bias limits modeling of long-range dependencies and often leads to over-smoothing in deeper networks Oono and Suzuki (2019); Li et al. (2020). Recent attempts, such as label propagation Wang and Leskovec (2020) and distance-aware encodings Li et al. (2020), provide partial solutions but still rely on heuristic designs, lacking fully learnable global interaction modeling.

Spectral-based methods provide an alternative approach to overcome the inherent locality constraints of traditional message-passing GNNs by explicitly operating in the frequency domain of graph signals. Initial approaches such as ChebNet Defferrard et al. (2016) utilize polynomial spectral filters derived from the eigenbasis of the graph Laplacian to capture multi-scale structural patterns. More recent advances, including adaptive spectral propagation methods Chien et al. (2020); Dwivedi et al. (2021) and graph wavelet frameworks Shen et al. (2021), further illustrate the benefit of decomposing signals into frequency components: low-frequency signals correspond to smooth global structures, while high-frequency signals typically capture local details and abrupt structural variations. Despite their conceptual appeal, spectral methods remain susceptible to noise and instability due to high-frequency components, particularly in small-scale or irregularly structured graphs. Moreover, explicit regularization of these spectral responses is rarely considered in prior works, leaving these methods prone to spectral instability and potential overfitting.

In parallel, a separate line of research has explored hierarchical and pooling-based GNNs designed to extract coarse-grained, global graph representations Ying et al. (2018); Lee et al. (2019); Gao and Ji (2019); Sun et al. (2024a). Notably, DiffPool Ying et al. (2018) employs differentiable hierarchical clustering, SAGPool Lee et al. (2019) leverages graph self-attention to identify salient structures, and Graph Filter Networks (GFN) Chen et al. (2019) utilize graph filtering operations for structural summarization. More recent approaches explicitly target structural bottlenecks and over-squashing issues by strategies such as expanding the width of central nodes to facilitate broader information flow Choi et al. (2024), or employing graphon-based fine-tuning techniques to align graph structures at different scales Sun et al. (2024b). Nevertheless, these approaches typically overlook the explicit use of dedicated mediator structures that can directly bridge distant nodes and systematically alleviate long-range communication bottlenecks. In contrast to these prior works, our proposed GMN framework integrates a Mediator node explicitly designed to enhance global message propagation that circumvents the locality constraints of traditional pooling operations, together with a spectral energy regularization mechanism, thereby jointly improving the robustness, stability, and expressive power of learned representations.

3. Proposed Method

To effectively capture long-range dependencies and stabilize spectral representations in graph-structured data, we propose **GMN**, a novel framework explicitly designed to enhance global communication and enforce frequency-domain robustness for graph classification tasks. As illustrated in Figure 2, GMN consists of three tightly integrated modules.

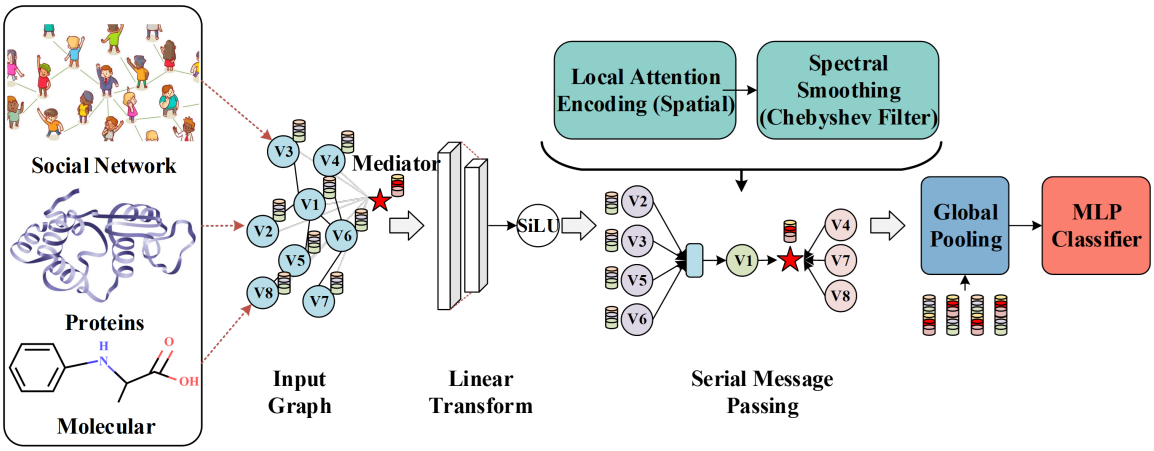

Figure 2: **Overall architecture of GMN.** GMN learns graph-level representations by first adding a Mediator node (red star) connected to all nodes, enabling global message exchange. Node features are processed through two parallel paths: a spatial path with local attention and a spectral path with Chebyshev filtering and spectral energy regularization. The outputs are fused via a gating mechanism and pooled to form a graph-level representation for downstream classification.

First, the Mediator node is introduced, which directly connects to all vertices, significantly extending the receptive field to efficiently propagate global context without increasing model depth. Second, the Spectral Graph Filter module utilizes Chebyshev polynomial approximations to extract structural patterns across multiple frequency bands, effectively capturing both global smoothness and local variations. Third, the Spectral Energy Regularization component explicitly penalizes high-frequency components, promoting smoother, more stable node embeddings and mitigating sensitivity to structural perturbations. These components are jointly incorporated into a dual-path architecture, where spatial and spectral signals are processed in parallel and integrated through a learnable gated fusion mechanism. The fused representations are then aggregated via global pooling to produce compact, discriminative graph-level embeddings. In the following subsections, we formally define the learning problem and detail each module of the GMN framework.

3.1. Problem Formulation

Given a dataset of labeled graphs $\mathcal{D} = \{(G_i, y_i)\}_{i=1}^{N}$, each graph is denoted by $G = (V, E, \mathbf{X})$, where V is a set of vertices, $E \subseteq V \times V$ represents edges, and $\mathbf{X} \in \mathbb{R}^{|V| \times d}$ contains the associated node features. The primary task addressed in this paper is graph classification, aiming to learn a predictive function $f : \mathcal{G} \to \mathcal{Y}$ that maps each graph to a discrete label $y \in \mathcal{Y}$.

Specifically, we formulate the prediction process as:

$$\hat{y} = f(G) = \phi(\mathcal{F}(G)), \tag{1}$$

where $\mathcal{F}$ is a representation learning module responsible for capturing comprehensive structural patterns at both local and global scales from input graph G. Subsequently, the classification head ϕ maps the learned representation into the label space. Our ultimate objective is to optimize the parameters of f by minimizing the classification error over the training set, thus ensuring accurate predictions on unseen graphs.

3.2. Mediator as Semantic Relay for Global Interaction

Conventional message-passing schemes aggregate information only from local neighborhoods, inherently restricting their capability to capture distant semantic relationships, especially in sparse or heterophilic graphs. To overcome this limitation, we introduce Mediator node, a dedicated semantic relay designed to explicitly facilitate global message propagation.

Specifically, we augment $G = (V, E, \mathbf{X})$ with a virtual Mediator node v_M to form $\tilde{G} = (\tilde{V}, \tilde{E}, \tilde{\mathbf{X}})$ where $\tilde{V} = V \cup \{v_M\}$ and $\tilde{E} = E \cup \{(v_i, v_M) \mid v_i \in V\}$. For any two nodes $u, v \in V$, there exists a path $u \to v_M \to v$ of length exactly 2 (or 1 if $(u, v) \in E$), which guarantees global communication without increasing layer depth. This augmentation adds $\mathcal{O}(|V|)$ edges and makes $\deg(v_M) = |V|$, preserving linear-time message passing on sparse graphs.

At layer l, we update node embeddings using a normalized aggregator that includes the mediator:

$$\mathbf{h}_i^{(l+1)} = \mathrm{AGG}\big(\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\} \cup \{\mathbf{h}_M^{(l)}\}\big), \quad \mathbf{h}_M^{(l+1)} = \mathrm{AGG}\big(\{\mathbf{h}_j^{(l)} : j \in V\}\big), \tag{2}$$

where AGG can be implemented as attention or degree-normalized convolution to prevent v_M from dominating due to its high degree.

By leveraging Mediator node, GMN significantly expands the receptive field of message passing without increasing the depth of the network, effectively alleviating over-squashing issues that commonly occur in deep GNN architectures. The resulting embeddings inherently capture richer global context, facilitating more discriminative and robust representations.

3.3. Spectral Graph Filter and Energy Regularization

While the spatial pathway equipped with Mediator node effectively facilitates global message propagation, graph signals inherently possess rich structural information that can be more naturally interpreted in the spectral domain. Specifically, from the viewpoint of graph spectral theory, node representations can be decomposed into eigenmodes of the normalized graph Laplacian, where low-frequency components capture smooth, global structural patterns, and high-frequency components characterize local variations and rapid structural changes. However, it has been empirically observed that high-frequency components are more susceptible to noise and structural perturbations, leading to unstable and inconsistent node embeddings Shuman et al. (2013); Chien et al. (2020).

To explicitly leverage multi-scale spectral information while mitigating instability from high-frequency signals, we incorporate a Spectral Graph Filter module based on Chebyshev polynomial approximation Defferrard et al. (2016). Given a graph Laplacian L, the spectral filtering operation is defined as:

$$\mathbf{Z} = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})\mathbf{X}, \tag{3}$$

where $\tilde{L} = 2L/\lambda_{\max} - I_n$ denotes the scaled Laplacian, $T_k(\cdot)$ represents the k-th Chebyshev polynomial, θ_k are learnable filter coefficients, and $\mathbf{Z}$ are the filtered node features. This approach efficiently captures structural patterns across multiple frequency bands without explicit eigen-decomposition, thus significantly reducing computational complexity and enabling scalability to large graphs.

We optimize a composite loss that combines the primary supervised objective with a pre-fusion consistency regularization, and the term $\mathcal{L}_{\text{NLL}}$ denotes the supervised negative log-likelihood on the training labels. The overall loss is:

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + \mathcal{L}_{\text{spec}}. \tag{4}$$

Let $\mathbf{H}^{\text{spatial}}, \mathbf{H}^{\text{spectral}} \in \mathbb{R}^{|V| \times d}$ denote the node embeddings produced by the spatial and spectral paths (before fusion). We define:

$$\mathcal{L}_{\text{spec}} = \log\left(1 + \frac{1}{|V|d} \left\|\mathbf{H}^{\text{spectral}} - \mathbf{H}^{\text{spatial}}\right\|_F^2\right), \tag{5}$$

which encourages the two paths to remain consistent prior to gating, thereby suppressing unstable high-frequency deviations while preserving complementary information at fusion time.

This regularization penalizes large discrepancies between spatial and spectral branches, encouraging their consistency and promoting more stable, robust embeddings across varying graph structures. By jointly employing spectral filtering and explicit energy regularization, GMN effectively leverages the complementary strengths of spatial and spectral domains, providing graph representations that are simultaneously expressive, robust, and stable across diverse graph topologies.

3.4. Gated Fusion and Graph-level Representation

Given the distinct yet complementary characteristics captured by the spatial and spectral pathways, we propose a gated fusion mechanism to adaptively integrate these features into unified node representations. Specifically, for node i, let $\mathbf{h}_i^{\text{spatial}} \in \mathbb{R}^d$ and $\mathbf{h}_i^{\text{spectral}} \in \mathbb{R}^d$ denote embeddings derived from the spatial and spectral paths respectively. We compute gating weights $\mathbf{g}_i \in \mathbb{R}^d$ as:

$$\mathbf{g}_i = \sigma\left(\mathbf{W}_g \left[\mathbf{h}_i^{\text{spatial}} \| \mathbf{h}_i^{\text{spectral}}\right] + \mathbf{b}_g\right), \tag{6}$$

where $\mathbf{W}_g \in \mathbb{R}^{d \times 2d}$ and $\mathbf{b}_g \in \mathbb{R}^d$ are learnable parameters, $\sigma(\cdot)$ is the sigmoid activation, and $\|$ represents concatenation.

The fused node embedding is then obtained as a weighted combination:

$$\mathbf{h}_i^{\text{fused}} = \mathbf{g}_i \odot \mathbf{h}_i^{\text{spatial}} + (1 - \mathbf{g}_i) \odot \mathbf{h}_i^{\text{spectral}}, \tag{7}$$

where $\odot$ denotes element-wise multiplication. This adaptive fusion strategy enables the model to dynamically balance local spatial information and global spectral patterns, resulting in more expressive and discriminative node-level representations.

Table 1: Statistics of graph classification datasets used in our experiments.

Category	Dataset	#Graphs	Avg. #Nodes	Avg. #Edges	#Classes
Social	COLLAB	5000	74.49	2457.78	3
	IMDb-B	1000	19.77	96.53	2
	IMDb-M	1500	13.00	65.94	3
Bioinfo	PROTEINS	1113	39.06	72.82	2
	COX2	467	41.22	86.89	2
Molecule	MUTAG	188	17.93	19.79	2
	PTC-MR	344	14.29	14.69	2
	BZR	405	35.75	76.72	2

Finally, to obtain a graph-level representation, we employ global pooling (e.g., sum or mean pooling) over all node embeddings:

$$\mathbf{h}_G = \text{Pooling}\left(\{\mathbf{h}_i^{\text{fused}}\}_{i \in V}\right). \tag{8}$$

The resulting embedding $\mathbf{h}_G$ serves as the graph-level feature, subsequently used by a classifier head for prediction.

4. Experiments

4.1. Datasets and Metrics

We evaluate GMN on eight widely-used benchmark datasets spanning three domains: **social networks**, **bioinformatics**, and **molecular graphs**. These datasets encompass a wide spectrum of graph topologies and label semantic complexities, establishing a comprehensive evaluation framework for assessing model robustness and generalization capabilities.

Social network graphs, including *COLLAB*, *IMDB-BINARY* (IMDb-B), and *IMDB-MULTI* (IMDb-M), where nodes represent people (e.g., actors or researchers), and edges denote collaborations or co-occurrence. These graphs often exhibit heterophilic patterns, making it challenging for standard GNNs to propagate meaningful information across distant regions.

Bioinformatics datasets, such as *PROTEINS* and *COX2*, describe macromolecular structures where nodes are amino acids or atoms, and edges reflect spatial or chemical interactions. These graphs are medium-sized with varying degrees of structural complexity.

Molecular datasets, including *MUTAG*, *PTC-MR*, and *BZR*, represent chemical compounds as graphs, where nodes are atoms and edges are chemical bonds. These graphs tend to be small and highly sensitive to structural noise, which makes them ideal for evaluating spectral stability.

All datasets are framed as graph-level classification tasks. Following prior work, we randomly split the data into 80% training, 10% validation, and 10% testing. We apply classification accuracy (%) as the main evaluation metric and report the average performance across 10 random seeds to ensure stability.

Table 2: Graph Classification Test Accuracy (%) on Benchmark Datasets.

Method	SOCIAL NETWORKS			BIOINFO		MOLECULES		
	COLLAB	IMDb-B	IMDb-M	PROTEINS	COX2	MUTAG	PTC-MR	BZR
GAT-based Methods								
CapsGNN	79.62	73.10	50.27	76.28	-	86.67	-	-
AutoGCL	70.12	73.30	53.54	75.80	77.98	88.64	68.21	83.34
RGCL	70.92	71.85	53.88	75.03	78.15	85.67	59.87	81.48
ESA	81.76	76.52	54.38	78.26	80.74	88.12	66.09	84.87
GCN-based Methods								
GIN	80.20	75.10	52.30	76.20	77.87	89.40	64.60	85.73
DGCNN	68.34	70.00	47.80	75.10	76.48	85.80	65.43	81.34
GFN	81.50	73.00	51.80	76.46	-	90.84	66.83	-
GFN-light	81.34	73.00	51.20	77.44	-	89.89	61.32	-
GIUNet	82.02	74.92	53.26	79.38	81.40	90.32	67.10	86.00
TFGW	84.33	78.34	56.81	82.93	88.94	96.42	72.45	92.28
Other Methods								
PPGN	81.38	72.20	44.73	76.39	75.75	88.88	64.70	86.54
WEGL	79.80	75.40	52.00	76.50	-	88.30	58.60	-
PANDA	82.44	76.85	54.99	78.90	81.28	89.21	66.80	85.76
G-TUNING	80.45	75.60	52.90	77.12	79.20	89.63	65.42	85.31
GMN (Ours)	**91.44**	**81.00**	**61.00**	**85.71**	**91.66**	**96.57**	**83.33**	**97.59**

4.2. Comparison with SOTA Methods

To evaluate the effectiveness of GMN, we compare it with a wide range of state-of-the-art methods, categorized into GAT-based models, GCN-based models, and other advanced graph learning frameworks.

GAT-based methods. CapsGNN Xinyi and Chen (2019) integrates capsule networks with attention mechanisms and dynamic routing to improve graph-level representation. AutoGCL Yin et al. (2022) adopts graph-level contrastive learning with GAT encoders and automatically learns augmentation distributions. RGCL Li et al. (2022) further enhances contrastive learning by extracting rationale-aware subgraphs and leveraging GATs to perform instance discrimination. ESA Buterez et al. (2024) discards message passing and treats graphs as edge sets, using interleaved masked and vanilla self-attention layers to learn edge-level representations that outperform classical GATs in long-range tasks.

GCN-based models. GIN Xu et al. (2018) designs an injective aggregation scheme that enhances the expressive power of standard GCNs. DGCNN Wu et al. (2018) uses disordered graph convolution followed by a sort pooling layer to preserve critical structure. GFN and GFN-light Chen et al. (2019) decouple graph filtering and permutation-invariant set functions to build fast and accurate classifiers. TFGW Vincent-Cuaz et al. (2022) introduces a template-based approach using Fused Gromov-Wasserstein (FGW) distances, learning graph representations by measuring dissimilarities to learnable template graphs, combining both structural and feature information through optimal transport. GIUNet Amouzad et al. (2024) applies a GIN-based U-Net architecture with spectral-aware pooling, utilizing node centrality and low-frequency components for hierarchical graph learning.

Other competitive models. PPGN Maron et al. (2019) achieves powerful representation by combining MLPs with tensor-based aggregation. PANDA Choi et al. (2024)

Table 3: Ablation Study of GMN Components (Test Accuracy %).

Variant	SOCIAL NETWORKS			BIOINFO		MOLECULES		
	COLLAB	IMDb-B	IMDb-M	PROTEINS	COX2	MUTAG	PTC-MR	BZR
GMN (Full)	**91.44**	**81.00**	**61.00**	**85.71**	**91.66**	**96.57**	**83.33**	**97.59**
w/o Spectral Energy Regularization	87.20	78.10	57.24	81.39	86.00	90.31	76.20	91.42
Sum Fusion (No Gate)	80.01	77.84	53.10	79.41	80.55	89.48	66.73	91.20
Concat Only (No Fusion)	77.42	74.93	51.28	76.13	77.91	85.79	63.21	83.26
w/o Mediator & Spectral Transformation	65.56	52.53	46.12	65.00	64.12	63.55	52.42	72.86
w/ Mediator only	73.20	63.80	50.00	70.20	72.00	78.20	61.85	82.13
w/ Spectral Transformation only	75.40	69.33	52.44	74.80	76.30	83.70	66.10	84.52

avoids graph rewiring and instead expands the hidden dimensions of high-centrality nodes to mitigate over-squashing. G-TUNING Sun et al. (2024b) addresses structural divergence in transfer learning by approximating downstream graph patterns. Finally, WEGL Kolouri et al. (2020) serves as a non-neural baseline that computes global graph embeddings via Wasserstein distances between node distributions.

Table 2 shows that GMN achieves the highest accuracy on all datasets, outperforming GAT-based, GCN-based, and long-range-aware models. It yields large gains on social graphs (e.g., +5.2% on COLLAB) due to effective long-range modeling via the Mediator, and strong improvements on small or irregular datasets (e.g., +4.3% on PTC-MR) from spectral regularization. These results validate the robustness of GMN's unified spatial–spectral design across diverse graph types.

4.3. Ablation Study

We conduct ablation studies on all datasets (Table 3) to evaluate each component.

Fusion and regularization. Removing spectral regularization reduces accuracy, especially on small or noisy graphs, confirming its role in controlling high-frequency noise. Replacing gated fusion with summation or removing it entirely causes clear drops, showing the need for adaptive integration.

Backbone modules. Removing both the Mediator and spectral module leads to severe degradation, indicating local aggregation alone is insufficient. The Mediator benefits social graphs by enabling long-range communication, while the spectral module performs better on bio/molecular data, highlighting its structural modeling strength.

Overall, the spectral path, Mediator, fusion, and regularization are all critical, and their combination consistently achieves the best results.

4.4. Efficiency and Interpretability

We evaluate GMN in terms of computational efficiency and interpretability.

Parameter and Runtime Comparison. Figure 3 shows model parameter counts (log scale), training time per epoch, and inference time per graph. GMN is lightweight and close to GCN in runtime, while being much smaller and faster than GAT, indicating that the dual-path design and Mediator node add minimal overhead.

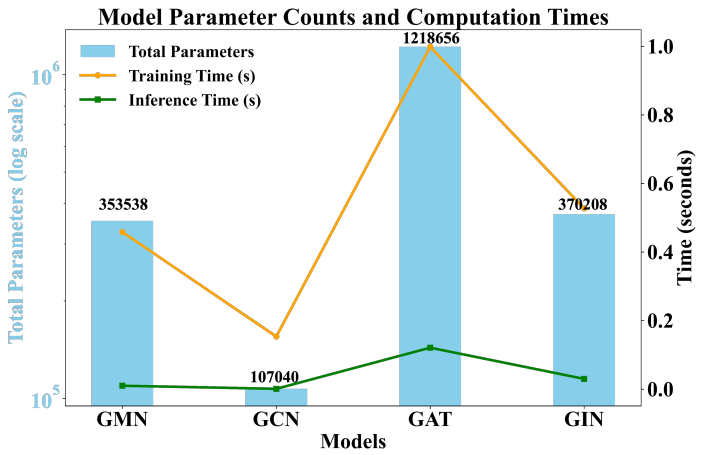

Figure 3: **Model size and efficiency.** Bars represent parameter counts; lines show per-epoch training time and per-graph inference time. GMN achieves high efficiency with modest complexity.

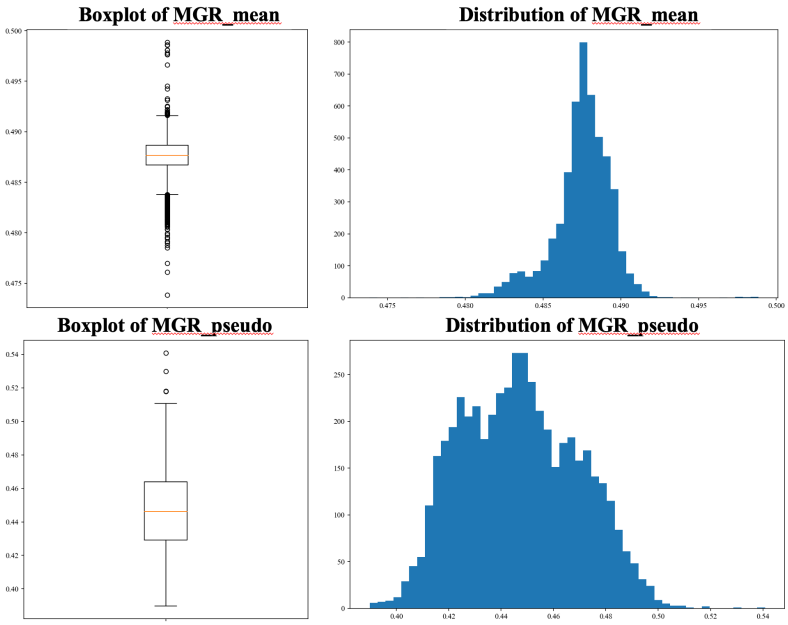

Figure 4: **Mediator Gate Ratio (MGR).** Top: $\mathrm{MGR}_{\mathrm{mean}}$ distribution, showing high stability across nodes. Bottom: $\mathrm{MGR}_{\mathrm{mediator}}$ distribution, reflecting Mediator adaptivity between spectral and spatial paths.

Mediator Gate Ratio (MGR) Analysis. To study the fusion mechanism, we compute gating metrics for node v with gate $\mathbf{g}_v \in [0,1]^d$:

$$\mathrm{MGR}_{\mathrm{mean}} = \frac{1}{|V|d} \sum_{v \in V} \sum_{c=1}^{d} g_{v,c}, \quad \mathrm{MGR}_{\mathrm{mediator}} = \frac{1}{d} \sum_{c=1}^{d} g_{v_M,c},$$

Table 4: **Robustness Analysis. Top:** Generalization across different node proportions (small / medium / large). **Bottom:** Test accuracy under increasing edge deletion. Bold denotes best results.

Robustness to Node Percentiles (30%, 60%, 90%)										
Model	**COLLAB (Social)**			**PROTEINS (Bioinfo)**			**BZR (Molecule)**			**Avg**
	30%	60%	90%	30%	60%	90%	30%	60%	90%	Acc.
GIN	74.12	78.35	80.03	70.83	75.22	**77.80**	78.84	81.20	83.14	77.95
DGCNN	64.12	66.37	68.20	70.14	71.59	74.07	78.91	79.85	81.03	72.70
ESA	79.05	80.90	81.32	75.38	76.88	78.40	82.20	83.21	84.55	80.21
TFGW	80.88	82.24	84.10	78.45	80.19	81.32	89.55	91.76	92.23	84.52
GMN (Ours)	**82.30**	**85.25**	**88.41**	**78.22**	**80.25**	**83.50**	**90.74**	**93.31**	**95.78**	**86.42**
Robustness to Edge Deletion (10%, 15%, 20%)										
Model	**COLLAB**			**PROTEINS**			**BZR**			**Avg**
	10%	15%	20%	10%	15%	20%	10%	15%	20%	Acc.
GIN	77.85	75.34	73.00	72.15	70.06	68.30	78.26	76.31	74.00	73.47
DGCNN	70.12	67.48	65.10	69.21	67.93	65.89	76.34	74.21	71.92	69.91
ESA	80.16	78.05	75.80	75.70	74.21	72.60	83.12	80.84	78.90	77.60
TFGW	83.10	81.54	78.46	78.85	76.93	74.18	90.10	88.47	85.93	81.95
GMN (Ours)	**85.62**	**82.34**	**78.50**	**81.45**	**79.58**	**75.40**	**92.45**	**91.83**	**89.01**	**84.02**

where v_M is the Mediator node. MGR_{mean} measures overall fusion stability; $\text{MGR}_{\text{mediator}}$ reflects Mediator adaptivity.

As shown in Figure 4, MGR_{mean} is concentrated near 0.485, showing consistent gating across nodes. $\text{MGR}_{\text{mediator}}$ varies more (mean 0.44–0.46, occasional peaks > 0.5), indicating adaptive adjustment between spectral and spatial paths based on graph structure.

4.5. Robustness Analysis

As shown in Table 4, we evaluate the robustness of all models under two perturbation settings: node percentiles and edge deletion. For the node percentile experiment, we evaluate model performance on graphs with 30%, 60%, and 90% of nodes, using proportional splits to account for the large variation in graph sizes and to avoid bias from graphs with few nodes. GMN consistently achieves the highest accuracy across all datasets and node proportion levels, demonstrating superior generalization across graphs of varying scales and an enhanced ability to capture long-range dependencies. In the edge deletion experiment, GMN maintains the highest test accuracy under increasing levels of edge removal, showing substantially smaller performance degradation compared to baseline models (e.g., 89.01% on BZR at 20% edge deletion).

Figure 5 shows robustness curves under increasing edge deletion. GMN consistently suffers smaller or comparable performance drops than GCN and GAT. On datasets such as COLLAB, COX2, and MUTAG, its degradation curve is noticeably flatter, demonstrating strong resilience to connectivity loss. Even on datasets with similar decline trends (e.g., BZR, IMDb-M, PROTEINS), GMN maintains higher accuracy, confirming its ability to preserve key structural information under severe perturbations. These results highlight GMN's robustness in noisy or incomplete graphs.

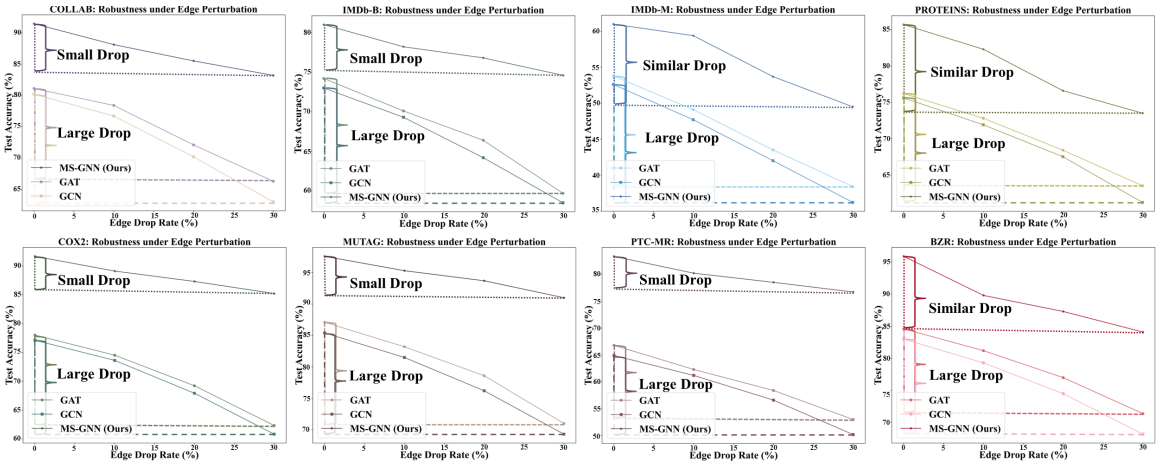

Figure 5: Robustness of GAT, GCN, and GMN under increasing edge drop rates on BZR, COLLAB, and PROTEINS datasets. GMN demonstrates significantly smaller accuracy degradation under both small and large drop rates.

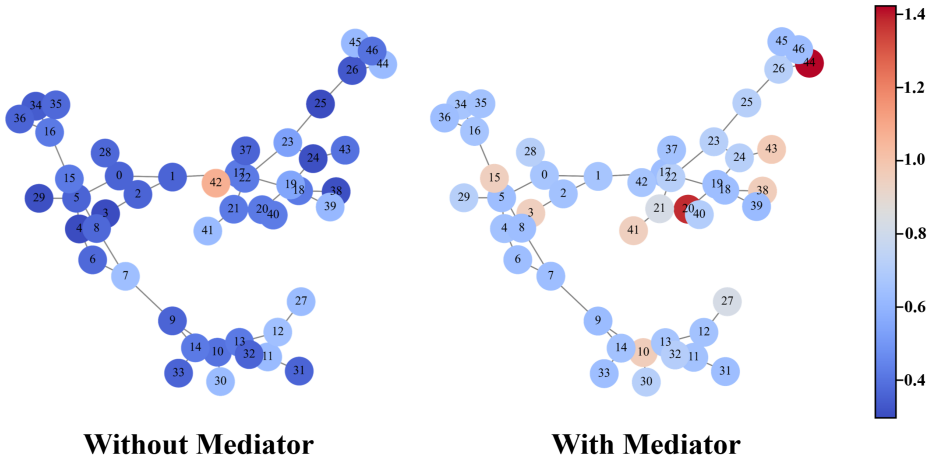

Figure 6: Node-level feature heatmaps of a BZR molecule without and with the mediator node. The mediator amplifies key node activations (e.g., 15, 40, 44), enhancing global context and highlighting important molecular substructures.

4.6. Case Study

We conduct a case study on a BZR molecule to examine the effect of the mediator node. Figure 6 presents node-level feature heatmaps without (left) and with (right) the mediator. Without the mediator, feature activations remain uniformly low, suggesting a reliance on local neighborhoods. With the mediator, key nodes (e.g., 15, 40, 44) show heightened activations, corresponding to chemically meaningful substructures such as branch points and

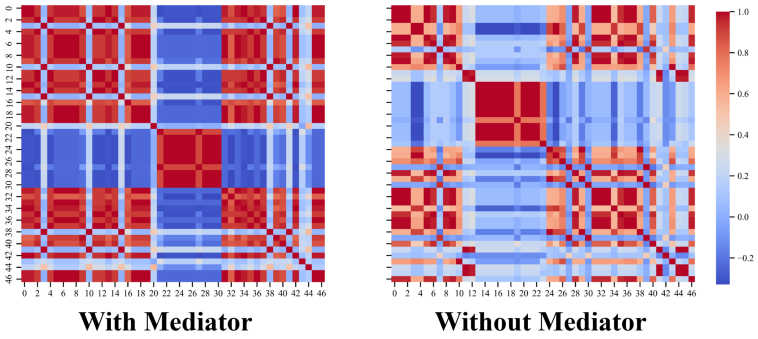

Figure 7: Pairwise node similarity heatmaps of a BZR molecule with and without the mediator node, showing improved global connectivity and cross-region interactions with the mediator.

functional groups. This indicates that the mediator enhances global context propagation and helps the model focus on structurally important regions.

Figure 7 further shows pairwise similarity heatmaps. Without the mediator, the matrix exhibits isolated local blocks, reflecting poor long-range connectivity. In contrast, the mediator induces a more globally coherent similarity pattern, bridging distant regions and enabling better integration of local and global structure. These results highlight the mediator's role in enhancing representation expressiveness and improving molecular property modeling.

5. Conclusion

In this paper, we propose GMN, a graph neural network that captures multi-scale dependencies by integrating spatial and spectral features, enhanced by a mediator node for efficient long-range communication. Through gated fusion and spectral energy regularization, GMN effectively mitigates feature degradation. Extensive experiments on social, biological, and molecular graphs show that GMN consistently outperforms state-of-the-art baselines in accuracy and robustness. Our analyses further demonstrate its interpretability and resilience, making it well-suited for real-world noisy graph scenarios. The mediator-based, multi-scale design of GMN offers a promising direction for future robust graph learning.

Acknowledgments

This work is supported by the National Key Research and Development Program of China under Grant 2024YFC3308500, National Natural Science Foundation of China under Grant 62406036, Beijing Municipal Natural Science Foundation under Grant L251042, the State Key Laboratory of Networking and Switching Technology under Grant NST20250110, and also sponsored by SMP-Zhipu. AI Large Model Cross-Disciplinary Fund under Grant ZPCG20241029322.

References

Alireza Amouzad, Zahra Dehghanian, Saeed Saravani, Maryam Amirmazlaghani, and Behnam Roshanfekr. Graph isomorphism u-net. *Expert Systems with Applications*, 236: 121280, 2024.

David Buterez, Jon Paul Janet, Dino Oglic, and Pietro Lió. An end-to-end attention-based approach for learning on graphs. 2024. URL https://api.semanticscholar.org/CorpusID:267740349.

Ting Chen, Song Bian, and Yizhou Sun. Are powerful graph neural nets necessary? a dissection on graph classification. *arXiv preprint arXiv:1905.04579*, 2019.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. *arXiv preprint arXiv:2006.07988*, 2020.

Jeongwhan Choi, Sumin Park, Hyowon Wi, Sung-Bae Cho, and Noseong Park. Panda: Expanded width-aware message passing beyond rewiring. *arXiv preprint arXiv:2406.03671*, 2024.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. *arXiv preprint arXiv:2110.07875*, 2021.

Hongyang Gao and Shuiwang Ji. Graph u-nets. In *international conference on machine learning*, pages 2083–2092. PMLR, 2019.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Soheil Kolouri, Navid Naderializadeh, Gustavo K Rohde, and Heiko Hoffmann. Wasserstein embedding for graph learning. *arXiv preprint arXiv:2006.09430*, 2020.

Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In *International conference on machine learning*, pages 3734–3743. PMLR, 2019.

Pan Li, Yanbang Wang, Hongwei Wang, and Jure Leskovec. Distance encoding: Design provably more powerful neural networks for graph representation learning. *Advances in Neural Information Processing Systems*, 33:4465–4478, 2020.

Sihang Li, Xiang Wang, An Zhang, Yingxin Wu, Xiangnan He, and Tat-Seng Chua. Let invariant rationale discovery inspire graph contrastive learning. In *International conference on machine learning*, pages 13052–13065. PMLR, 2022.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. *Advances in neural information processing systems*, 32, 2019.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. *arXiv preprint arXiv:1905.10947*, 2019.

Yangmei Shen, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Multi-scale graph convolutional network with spectral graph wavelet frame. *IEEE Transactions on Signal and Information Processing over Networks*, 7:595–610, 2021.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

Jiangfeng Sun, Xinyue Lin, Fangyu Hao, and Meina Song. Hybrid convolution method for graph classification using hierarchical topology feature. In *Asian Conference on Machine Learning*, pages 1308–1320. PMLR, 2024a.

Yifei Sun, Qi Zhu, Yang Yang, Chunping Wang, Tianyu Fan, Jiajun Zhu, and Lei Chen. Fine-tuning graph neural networks by preserving graph generative patterns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9053–9061, 2024b.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Cédric Vincent-Cuaz, Rémi Flamary, Marco Corneli, Titouan Vayer, and Nicolas Courty. Template based graph neural network with optimal transport distances. *Advances in Neural Information Processing Systems*, 35:11800–11814, 2022.

Hongwei Wang and Jure Leskovec. Unifying graph convolutional neural networks and label propagation. *arXiv preprint arXiv:2002.06755*, 2020.

Bo Wu, Yang Liu, Bo Lang, and Lei Huang. Dgcnn: Disordered graph convolutional neural network based on the gaussian mixture model. *Neurocomputing*, 321:346–356, 2018.

Zhang Xinyi and Lihui Chen. Capsule graph neural network. In *International conference on learning representations*, 2019.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Yihang Yin, Qingzhong Wang, Siyu Huang, Haoyi Xiong, and Xiang Zhang. Autogcl: Automated graph contrastive learning via learnable view generators. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 8892–8900, 2022.

Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *Advances in neural information processing systems*, 31, 2018.