CondPSE: A Polynomial-Filtered Structural Encoder with Conditional Modulation for Graphs

Anonymous Author(s)

Abstract

Message-passing graph neural networks are bounded by the 1-WL test and can miss topological structure that distinguishes non-isomorphic graphs. Positional and structural encodings (PSE) inject such topology-derived signals, and learned PSE encoders such as GPSE pretrain a single encoder to produce these signals from random node probes, which can then be frozen and reused as inputs across downstream graph models. We present CondPSE, a learned PSE encoder that applies a learnable polynomial graph filter bank to standard Gaussian node probes and refines the resulting structural-response branches through FiLM-style modulation conditioned on cross-filter, local message-passing, and graph-level signals. CondPSE is pretrained to reconstruct node-level positional/structural targets and graph-level invariants, and is then frozen for use as a downstream input encoding. On synthetic structural-discrimination benchmarks, CondPSE separates graph structures that 1-WL-bounded message passing cannot: it raises CSL accuracy from 42.9% to 97.3% and EXP accuracy from 68.3% to 99.9% relative to GPSE, and ablations show that the polynomial filter bank accounts for most of this gain. On real molecular property prediction, the picture is more limited. With a hybrid local-message-passing/global-attention backbone, CondPSE performs comparably to GPSE without surpassing it, and a ZINC backbone sweep shows no consistent ordering between the two encoders. We report these results and discuss why strong synthetic structural discrimination does not, on its own, yield a downstream advantage for frozen learned PSE encoders, including the role of downstream integration and possible mismatch between structural pretraining targets and molecular property labels.

CCS Concepts

• **Computing methodologies** → **Neural networks**; • **Mathematics of computing** → *Graph algorithms.*

Keywords

Graph Neural Networks, Positional Encoding, Graph Expressivity, Transfer Learning, Polynomial Graph Filtering

1 Introduction

Message-passing GNNs are bounded by the 1-WL test [16, 22]. Structural and positional encodings add topology-derived signals that such models may miss [3, 7, 13, 19, 23]. Hand-crafted encodings include random-walk statistics [19], Laplacian eigenvectors [7], shortest-path distances, and centrality encodings [23]; learned PSE encoders such as GPSE [3] are instead pretrained to produce positional/structural encodings by reconstructing multiple positional/structural targets. Because pretrained PSE encoders can be frozen and used to generate positional/structural inputs for downstream graph models, their value depends not only on what positional/structural targets they reconstruct, but also on how downstream architectures use the resulting encodings.

We study this gap through *CondPSE*, a learned structural encoder that applies polynomial graph filters to standard Gaussian node probes. CondPSE is pretrained to reconstruct node-level positional/structural targets and graph-level invariants. Relative to GPSE's [3] stacked message-passing extractor, CondPSE uses a polynomial graph filter bank to initialize multiple structural-response branches before conditional modulation. It modulates these branches using cross-filter, local, and global context via FiLM-style modulation [18].

Our main finding is that stronger benchmark-level structural discrimination obtained using frozen node encodings does not automatically translate into stronger downstream graph-level transfer. CondPSE achieves strong performance on CSL and near-perfect performance on EXP, suggesting that the encoder separates structural patterns useful for CSL and EXP, where labels are determined by graph topology rather than node or edge attributes. In our molecular evaluation, transfer with the hybrid local-message-passing/global-attention backbone remains close to GPSE, and the ZINC backbone sweep shows that the relative performance of GPSE and CondPSE varies by downstream architecture despite CondPSE's stronger CSL/EXP structural-discrimination performance.

We use CondPSE as a diagnostic encoder for studying when benchmark-level structural discrimination transfers to downstream graph-level prediction.

We make three contributions:

- We introduce CondPSE, a learned structural encoder that uses polynomial-filtered structural-response branches followed by conditional modulation.
- We find that CondPSE obtains strong CSL/EXP performance, and ablations indicate that the polynomial filter bank plays the largest role while cross-filter, local, and global conditioning provide additional contributions.
- We show that stronger CSL/EXP structural-discrimination performance does not automatically translate into stronger downstream molecular transfer, and use a ZINC backbone sweep to examine how downstream architecture affects the relative utility of GPSE and CondPSE.

2 Related Work

Structural and Positional Encodings. Many graph learning models augment node features with positional or structural information to overcome message-passing limits. Traditional message-passing graph neural networks (MPNNs) such as GCN [12], GIN [22], and GINE [10] are bounded by the 1-WL graph isomorphism test [16, 22]. Traditional positional/structural encodings include Laplacian eigenvectors (LapPE) and random-walk statistics (RWSE). To integrate these features, LSPE [7] learns decoupled structural and positional representations, while SAN [13] and Graphormer [23] incorporate

them into Transformer architectures. GraphGPS [19] combines local message passing, global attention, and positional encodings in a modular graph Transformer backbone.

Polynomial Graph Filtering and Spectral GNNs. Spectral graph neural networks define graph filters through the eigenstructure of graph operators, while practical models often implement such filters using polynomial functions of the adjacency or Laplacian matrix. ChebNet [5] approximates spectral graph filters with Chebyshev polynomials, replacing explicit eigendecomposition with polynomial functions of a graph operator. GPR-GNN [4] learns weights over multiple propagation powers. These methods motivate using powers of a graph operator to capture responses at different propagation depths. CondPSE uses polynomial filtering as an extraction stage: it forms initial structural-response branches from standard Gaussian node probes, then refines these branches through cross-filter, local, and global conditioning before producing the final structural encoding.

Random Features and Learned Encoders. Random node features have been shown to increase the distinguishing power of GNNs. Sato et al. [20] and Abboud et al. [1] show that random node initialization can increase the distinguishing power of GNNs beyond deterministic message passing in certain settings. GPSE [3] builds on this by learning a transferable positional and structural encoder from random features, using a GatedGCN [2] backbone to reconstruct multiple hand-crafted PSE targets. Unlike fixed encodings such as LapPE or RWSE, learned random-feature encoders do not directly prescribe a single structural statistic; instead, they learn to convert graph-induced random responses into positional/structural encodings. CondPSE shares GPSE's [3] goal of pretraining a transferable PSE encoder but constructs its intermediate responses differently: it first produces polynomial-filtered probe responses and then adapts them with conditional modulation modules, including local message passing.

3 Methodology

CondPSE starts from NormalSE probes sampled i.i.d. from a standard Gaussian distribution. These probes are graph-agnostic before adjacency propagation. The polynomial filter bank applies powers of the adjacency matrix to the probes and forms K learned mixtures of propagation-depth responses. These mixtures serve as initial structural-response branches that are later refined by conditioning modules. The conditioning modules use cross-filter comparisons, local message passing, and graph-level statistics, and FiLM modulation applies the resulting conditioning signals as adaptive channel-wise transformations. The encoder is pretrained once and then frozen; downstream experiments therefore evaluate how fixed structural encodings are incorporated by each backbone.

3.1 Extraction Phase: Polynomial Graph Filter Bank

Given a graph $G = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V}$, edge set $\mathcal{E}$, and unnormalized adjacency matrix A, we sample stochastic node probes $Z \in \mathbb{R}^{N \times d_{\text{in}}}$ as $Z_{v,c} \sim \mathcal{N}(0,1)$ and compute propagation responses from NormalSE probes. The probes serve two roles. First, stochastic node probes break node symmetry and can help distinguish structures that deterministic 1-WL-style message passing may fail to separate [1, 20]. Second, they act as random inputs for querying a matrix operator, analogous to randomized numerical linear algebra methods that use random vectors to estimate properties of matrix functions and operators [11, 21]. In CondPSE, $A^0 Z$ preserves the unpropagated probe channel, while $A^i Z$ for $i \geq 1$ records the same probes after i adjacency propagations. Thus, graph-structural information is introduced through adjacency propagation rather than through the initial probe values. Each polynomial filter learns a mixture of these propagation-depth responses, and the resulting filter branches serve as initial structural-response branches for the conditioning modules.

Following polynomial graph filtering ideas used in spectral GNNs such as ChebNet [5] and GPR-GNN [4], we filter Z through a bank of K learnable polynomial graph filters of degree d:

$$\tilde{H}_k = \sum_{i=0}^{d} \bar{c}_{k,i} A^i Z, \quad H_k^{(0)} = \mathrm{LN}(\tilde{H}_k W_{\text{proj}} + b_{\text{proj}}), \tag{1}$$

where $\bar{c}_{k,i}$ denotes the normalized polynomial coefficient:

$$\bar{c}_{k,i} = \frac{c_{k,i}}{\sqrt{\sum_{j=0}^{d} c_{k,j}^2} + \epsilon}, \tag{2}$$

and $W_{\text{proj}} \in \mathbb{R}^{d_{\text{in}} \times D}$ projects features into the model dimension D. Here and below, LN denotes LayerNorm. We use coefficient normalization and LayerNorm after projection to keep propagation responses numerically stable across graphs, filters, and propagation degrees. A single filter provides one learned mixture of propagation responses, while the polynomial graph filter bank computes multiple such responses across filters. Thus, each branch is a learned polynomial filter over A, implemented through repeated adjacency applications rather than eigendecomposition. The K branches retain different mixtures of propagation depths before conditioning. After B conditioning blocks, the K filter branches are fused into a single node encoding:

$$E_v = \psi\Big(H_{v,1}^{(B)} \,\|\, \cdots \,\|\, H_{v,K}^{(B)}\Big), \qquad \psi : \mathbb{R}^{KD} \to \mathbb{R}^{d_{\text{out}}}. \tag{3}$$

where $\|$ denotes feature concatenation, d_{out} is the output node-encoding dimension, and ψ is a two-layer MLP that produces the final node encoding E_v.

3.2 Conditioning Modulation and Injection

At each block $b \in \{1, \ldots, B\}$, the hidden tensor $H^{(b-1)} \in \mathbb{R}^{N \times K \times D}$ is modulated using three conditioning signals derived from the current filter-branch states:

Cross-filter conditioning: Cross-filter conditioning computes pairwise comparison features across polynomial-filter branches, allowing information exchange among current filter-branch states. For each node v and filter pair (k, j), we compute comparison features from the current hidden tensor:

$$F_{v,k,j} = \big[H_{v,k}^{(b-1)} \,\|\, H_{v,j}^{(b-1)} \,\|\, (H_{v,k}^{(b-1)} - H_{v,j}^{(b-1)}) \,\|\, (H_{v,k}^{(b-1)} \odot H_{v,j}^{(b-1)})\big].$$

Pairwise messages are summed over $j \neq k$ and mapped through shared MLPs to yield the cross-filter signal $I_{\text{xf},v,k} \in \mathbb{R}^D$.

Local conditioning: Local conditioning applies a shared two-layer ResGatedGCN [2] over each filter branch independently, aggregating edge-neighborhood information within each branch to produce the local spatial signal $I_{\mathrm{loc},v,k} \in \mathbb{R}^D$.

Global conditioning: Global conditioning computes graph-wide statistics, including mean, standard deviation, and maximum across nodes for each filter branch, projects them through an MLP, and broadcasts the resulting signal to nodes as $I_{\mathrm{glb},v,k} \in \mathbb{R}^D$.

These signals are concatenated into $U_{v,k} = [I_{\mathrm{xf},v,k} \parallel I_{\mathrm{loc},v,k} \parallel I_{\mathrm{glb},v,k}] \in \mathbb{R}^{3D}$. The block-specific FiLM MLP maps $U_{v,k}$ to a raw scale vector and shift vector, $(\gamma^r_{v,k}, \beta_{v,k}) = \phi_b(U_{v,k})$, where $\gamma^r_{v,k}, \beta_{v,k} \in \mathbb{R}^D$. The final channel-wise scale vector is constrained around one by $\gamma_{v,k} = \mathbf{1} + \tanh(\gamma^r_{v,k})$:

$$\begin{aligned}(\gamma^r_{v,k}, \beta_{v,k}) &= \phi_b(U_{v,k}), \quad \gamma_{v,k} = \mathbf{1} + \tanh(\gamma^r_{v,k}),\\ H^{(b)}_{v,k} &= \mathrm{LN}\Big(H^{(b-1)}_{v,k} + s_{\mathrm{res}}(\gamma_{v,k} \odot H^{(b-1)}_{v,k} + \beta_{v,k})\Big).\end{aligned} \tag{4}$$

Here, ϕ_b is the block-specific FiLM MLP, and $s_{\mathrm{res}} \in \mathbb{R}$ is a learnable residual scale initialized to 1. Thus, each filter branch receives a learned channel-wise scale and shift.

3.3 Downstream Integration

Once pretrained, the CondPSE encoder is frozen. The node-encoding matrix $E \in \mathbb{R}^{N \times d_{\mathrm{out}}}$, whose rows are E_v, is passed through a light-weight downstream adapter f_{PE} (a linear or MLP projection) to produce $E' \in \mathbb{R}^{N \times d_{\mathrm{pe}}}$. The final input to the downstream GNN is formed by concatenating the projected raw node features with the structural encoding:

$$X_{\mathrm{in}} = \big[X_{\mathrm{raw}} W_x \parallel E'\big], \tag{5}$$

where $W_x \in \mathbb{R}^{d_{\mathrm{in}} \times (d_{\mathrm{emb}} - d_{\mathrm{pe}})}$ projects the raw features to a complementary dimension before concatenation with E'. Freezing the encoder helps isolate the effect of the learned structural encoding under a fixed downstream integration setting.

4 Experiments

We pretrain CondPSE on OGBG-MolPCBA [9] using the same positional/structural pretraining targets as GPSE [3] and evaluate the frozen CondPSE encoder ($B = 6$ blocks, $K = 4$ filters, and degree $d = 15$) following GPSE's public evaluation setting, including dataset splits, seed protocols, and downstream backbone configurations. Pretraining uses graph-structure-derived targets and no downstream property labels. We report molecular transfer results using a GraphGPS (GPS) backbone and, for ZINC, an additional downstream backbone sweep across GCN [12], GIN [22], GatedGCN [2], GINE [10], and a Transformer-style backbone. Encoder, pretraining, downstream training details, baseline sources, and supervised objectives are given in Appendices A.1–A.3.

4.1 Benchmark-Level Structural Discrimination

We evaluate benchmark-level structural discrimination through GIN-based performance on the Circular Skip Link (CSL) [17] and EXP [1] graph isomorphism benchmarks. The CSL dataset consists of skip-link graphs where standard 1-WL GNNs struggle to distinguish different link lengths due to uniform degree distributions. The EXP dataset is a synthetic graph isomorphism benchmark designed to test GNN expressiveness against 1-WL limits.

Following Cantürk et al. [3], we compare CondPSE with GPSE and standard positional/structural encodings. LapPE uses the first four non-trivial Laplacian eigenvectors, while RWSE uses 20-dimensional random-walk return probabilities. Table 1 reports the results.

Table 1: Synthetic structural-discrimination benchmarks using ten-times stratified five-fold cross-validation (ACC % ↑). Bold indicates the best result for each benchmark.

Model	CSL	EXP
GIN (No PE)	10.0 ± 0.0	48.7 ± 2.2
GIN + GPSE	42.9 ± 7.9	68.3 ± 7.5
GIN + LapPE	92.5 ± 4.2	99.5 ± 0.8
GIN + RWSE	**100.0 ± 0.0**	99.7 ± 0.6
GIN + CondPSE w/o filter bank	12.0 ± 3.4	23.6 ± 2.2
GIN + CondPSE w/o global conditioning	87.3 ± 2.5	99.5 ± 0.3
GIN + CondPSE w/o local conditioning	92.0 ± 5.0	99.5 ± 0.3
GIN + CondPSE w/o cross-filter conditioning	96.4 ± 1.6	99.8 ± 0.5
GIN + CondPSE	97.3 ± 2.5	**99.9 ± 0.2**

As shown in Table 1, GIN without positional/structural encodings performs poorly on CSL and EXP, while explicit graph-operator encodings such as LapPE and RWSE provide strong positional/structural baselines. RWSE attains the highest CSL accuracy, which is consistent with the nature of CSL: the classes differ by circular skip-link structure under uniform degree patterns, and multi-step random-walk statistics can differentiate such structural differences. CondPSE reaches comparable CSL accuracy without concatenating RWSE directly as the downstream input encoding, although RWSE is included among the positional/structural pretraining targets. Thus, the comparison tests whether positional/structural supervision can be captured by a frozen learned encoder, rather than whether RWSE is explicitly supplied to the downstream GIN. Compared with GPSE, the closest learned random-feature encoder baseline, CondPSE increases CSL accuracy from 42.9 to 97.3 and EXP accuracy from 68.3 to 99.9. Compared with GPSE's stacked message-passing extractor, CondPSE explicitly forms multiple propagation-depth responses before conditional modulation. This design may make CSL/EXP-relevant structural differences easier for the downstream GIN to access. The ablations further separate the roles of the extraction and conditioning stages: removing the polynomial filter bank causes the largest degradation, suggesting that propagation-depth-specific graph-operator responses provide important initial inputs for the conditioning pipeline. Removing global, local, or cross-filter conditioning yields smaller but visible drops, especially on CSL, indicating that these contexts refine the filtered responses rather than replace the extraction stage. Together, these results suggest that CondPSE benefits from both polynomial-filtered initialization and subsequent conditional modulation.

4.2 GPS Molecular Downstream Transfer

We evaluate molecular downstream transfer with a GraphGPS (GPS) backbone, which combines local message passing with global attention [19]. We test on ZINC (subset) [6], MolHIV and MolPCBA [9], and PCQM4Mv2-subset [8], using the metrics shown in Table 2.

Table 2: GPS molecular downstream prediction results averaged over 10 seeds. Bold indicates the best result for each dataset.

Dataset	Metric	GPS + None	GPS + GPSE	GPS + CondPSE
ZINC (subset)	MAE ↓	0.1182 ± 0.0049	**0.0648 ± 0.0030**	0.0649 ± 0.0035
MolHIV	AUROC ↑	0.7798 ± 0.0077	**0.7815 ± 0.0133**	0.7797 ± 0.0070
MolPCBA	AP ↑	0.2869 ± 0.0012	0.2911 ± 0.0036	**0.2922 ± 0.0033**
PCQM4Mv2-subset	MAE ↓	0.1329 ± 0.0030	**0.1196 ± 0.0004**	0.1207 ± 0.0006

CondPSE is close to GPSE across the molecular tasks: it is slightly better on MolPCBA and slightly worse on ZINC, MolHIV, and PCQM4Mv2-subset. These small differences motivate a controlled ZINC sweep over downstream backbones.

4.3 ZINC Downstream Backbone Sweep

To test whether the relative utility of learned structural encodings varies across downstream architectures, we run a ZINC downstream backbone sweep [6]. For each architecture, we compare no structural encoding with encodings from GPSE and CondPSE under the same ZINC setup. Table 3 reports ZINC MAE for each backbone and encoding choice.

Table 3: ZINC downstream backbone sweep averaged over 4 seeds (MAE ↓). Bold indicates the best result for each architecture.

Backbone	None	GPSE	CondPSE
GCN	0.288 ± 0.004	0.129 ± 0.003	**0.126 ± 0.003**
GatedGCN	0.236 ± 0.008	0.113 ± 0.003	**0.110 ± 0.002**
GIN	0.285 ± 0.004	**0.124 ± 0.002**	0.131 ± 0.004
GINE	0.118 ± 0.005	**0.065 ± 0.003**	0.069 ± 0.002
Transformer	0.686 ± 0.017	0.189 ± 0.016	**0.186 ± 0.014**

Both learned encodings improve over the no-encoding baseline for every tested backbone, suggesting that these structural encodings provide useful information in this ZINC evaluation. However, the relative performance is mixed: CondPSE has lower MAE than GPSE with GCN [12], GatedGCN [2], and the Transformer-style backbone [19, 23], but higher MAE with GIN [22] and GINE [10]. Because several gaps are within reported variation, Table 3 is best read as a backbone-sensitivity check rather than a definitive superiority claim. This mixed pattern indicates that CSL/EXP performance alone does not determine downstream transfer.

5 Discussion

The results separate benchmark-level structural discrimination from downstream utility. We discuss this gap through backbone-dependent transfer, target alignment, and generalization.

Backbone-Dependent Transfer. Table 3 suggests that transfer behavior varies by architecture, although many GPSE–CondPSE differences are small. Both learned encodings improve over the no-encoding baseline across the ZINC backbones, but the mixed GPSE–CondPSE results suggest that their utility is not determined by CSL/EXP performance alone. We view the sweep as a diagnostic check against the hypothesis that stronger CSL/EXP structural-discrimination performance should consistently produce better downstream transfer. These results do not identify a single architectural factor that explains all trends, but they suggest that downstream integration is a relevant factor: different backbones may combine raw features, message passing, and positional/structural information in ways that favor different structural encoders.

Target Alignment Considerations. Beyond downstream integration, the choice of pretraining targets affects transfer, potentially leading to negative transfer [10]. Reconstructing graph-level invariants or hand-crafted positional/structural targets from node-level encodings may introduce a possible target-alignment mismatch. CondPSE reconstructs node-level positional/structural targets (ElectrostaticPE, LapPE, RWSE, HKdiagSE) and graph-level invariants (EigVals, CycleGE). These targets may not align perfectly with molecular labels, so optimizing these reconstruction objectives or achieving strong benchmark-level structural discrimination may not guarantee a task-aligned downstream inductive bias. We do not directly measure target-label alignment, so this remains a hypothesis. Future research should evaluate objectives that more directly anticipate downstream label structure.

Expressivity-Generalization Trade-off. The gap between strong CSL / near-perfect EXP performance and molecular transfer close to GPSE highlights a potential generalization trade-off. This empirical pattern is consistent with recent theory suggesting that greater expressivity can hurt generalization when structural distinctions reduce intra-class concentration or are misaligned with the downstream task metric [14, 15]. One possible explanation, not directly tested here, is that highly expressive structural encodings may induce overly fine structural partitions, making it harder for the downstream model to ignore irrelevant topological variation. Practical datasets like ZINC and MolHIV require models to generalize across varied graphs with similar properties. If structural encodings vary substantially for graphs with similar molecular properties, the downstream model may struggle to generalize.

Scope and Future Directions. The present experiments diagnose transfer behavior for learned structural encoders across the evaluated downstream architectures. Future work should isolate how cross-filter modulation, local message passing, and global context contribute under different graph families; extend the downstream backbone sweep beyond ZINC; and study positional/structural objectives that better anticipate task-specific semantics. More broadly, the pretraining objective, encoder design, and downstream adapter should be studied as interacting factors in transfer.

6 Conclusion

CondPSE serves as a diagnostic encoder for studying the alignment between benchmark-level structural discrimination and downstream transfer. It achieves strong CSL/EXP performance, while molecular transfer remains close to GPSE and the ZINC sweep shows architecture-dependent relative performance. For graph pipelines that use frozen learned PSE encoders as inputs to downstream backbones, these mixed transfer results suggest that benchmark-level structural discrimination alone does not by itself determine downstream transfer behavior.

References

[1] Ralph Abboud, Ismail Ilkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. 2021. The Surprising Power of Graph Neural Networks with Random Node Initialization. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. 2112–2118.

[2] Xavier Bresson and Thomas Laurent. 2017. Residual Gated Graph ConvNets. *arXiv preprint arXiv:1711.07553* (2017).

[3] Semih Cantürk, Renming Liu, Olivier Lapointe-Gagné, Vincent Létourneau, Guy Wolf, Dominique Beaini, and Ladislav Rampášek. 2024. Graph Positional and Structural Encoder. In *Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235)*. PMLR, 5533–5566.

[4] Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. 2021. Adaptive Universal Generalized PageRank Graph Neural Network. In *International Conference on Learning Representations*.

[5] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Advances in Neural Information Processing Systems*, Vol. 29. 3844–3852.

[6] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. 2023. Benchmarking Graph Neural Networks. *Journal of Machine Learning Research* 24, 43 (2023), 1–48.

[7] Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. 2022. Graph Neural Networks with Learnable Structural and Positional Representations. In *International Conference on Learning Representations*.

[8] Weihua Hu, Matthias Fey, Hongyu Ren, Maho Nakata, Yuxiao Dong, and Jure Leskovec. 2021. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

[9] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open Graph Benchmark: Datasets for Machine Learning on Graphs. In *Advances in Neural Information Processing Systems*, Vol. 33. 22118–22133.

[10] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. 2020. Strategies for Pre-training Graph Neural Networks. In *International Conference on Learning Representations*.

[11] Michael F. Hutchinson. 1990. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics: Simulation and Computation* 19, 2 (1990), 433–450.

[12] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

[13] Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. 2021. Rethinking Graph Transformers with Spectral Attention. In *Advances in Neural Information Processing Systems*, Vol. 34. 21618–21629.

[14] Shouheng Li, Floris Geerts, Dongwoo Kim, and Qing Wang. 2025. Towards Bridging Generalization and Expressivity of Graph Neural Networks. In *International Conference on Learning Representations*.

[15] Sohir Maskey, Raffaele Paolino, Fabian Jogl, Gitta Kutyniok, and Johannes F. Lutzeyer. 2026. Graph Representational Learning: When Does More Expressivity Hurt Generalization?. In *International Conference on Learning Representations*.

[16] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan E. Lenssen, Gaurav Rattan, and Martin Grohe. 2019. Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 4602–4609.

[17] Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. 2019. Relational Pooling for Graph Representations. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 4663–4673.

[18] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. 2018. FiLM: Visual Reasoning with a General Conditioning Layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. 3942–3951.

[19] Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. 2022. Recipe for a General, Powerful, Scalable Graph Transformer. In *Advances in Neural Information Processing Systems*, Vol. 35. 14501–14515.

[20] Ryoma Sato, Makoto Yamada, and Hisashi Kashima. 2021. Random Features Strengthen Graph Neural Networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining*. 333–341.

[21] Shashanka Ubaru, Jie Chen, and Yousef Saad. 2017. Fast estimation of tr(f(A)) via stochastic Lanczos quadrature. *SIAM J. Matrix Anal. Appl.* 38, 4 (2017), 1075–1099.

[22] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *International Conference on Learning Representations*.

[23] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do Transformers Really Perform Badly for Graph Representation?. In *Advances in Neural Information Processing Systems*, Vol. 34. 28877–28888.

A Implementation and Training Details

A.1 CondPSE Encoder Configuration

We use a standard CondPSE encoder configuration for all reported experiments. The architectural specifications are summarized in Table 4.

Table 4: CondPSE encoder configuration.

Component	Configuration
Input probe	NormalSE (256-dimensional)
Model dimension D	256
Output node-encoding dimension	256
Polynomial graph filter bank	$K = 4$ filters, polynomial degree $d = 15$
Conditioning blocks	$B = 6$ blocks
Local conditioning	2-layer ResGatedGCN
Cross-filter conditioning	2-layer pairwise-comparison MLP
Global conditioning	2-layer graph-statistics MLP
FiLM modulation	Channel-wise scale/shift modulation
FiLM residual scale s_{res}	Learnable scalar initialized to 1
Final fusion	Concatenate K filter branches and project
Adjacency operator A	Unnormalized adjacency matrix

A.2 Pretraining Setup

CondPSE is pretrained on OGBG-MolPCBA using the same positional/structural pretraining targets as GPSE [3]. The model reconstructs node-level positional/structural targets and graph-level invariants from NormalSE probes. The node-level targets are ElectrostaticPE, LapPE, RWSE, and HKdiagSE; the graph-level invariants are EigVals and CycleGE. We train the model for 120 epochs using the AdamW optimizer with a batch size of 512, a learning rate of 5×10^{-4}, weight decay of 1×10^{-4}, and a cosine learning rate scheduler with warmup.

A.3 Downstream Adapter and Training Details

For downstream tasks, the pretrained CondPSE encoder is frozen. The stored CondPSE node encoding is passed through a lightweight adapter and concatenated with projected raw node features before being passed to the downstream backbone.

Across Tables 1–3, baseline values for no encoding, GPSE, LapPE, and RWSE are taken from the GPSE paper or its public implementation reports, where applicable. CondPSE results are produced under the corresponding public GPSE evaluation settings, including dataset splits, seed protocols, and downstream backbone configurations. Backbone architectural hyperparameters, including the number of layers, hidden dimension, readout/pooling, and backbone-specific normalization or dropout settings, are inherited unchanged from the corresponding public GPSE configuration files. During CondPSE evaluation, the learned PSE encoder remains frozen; only the downstream adapter and downstream backbone are trained. In learned-PSE settings, the GPSE encoding is replaced by the stored frozen CondPSE encoding under the corresponding downstream configuration.

We use the dataset-specific supervised objectives from the corresponding GPSE downstream configurations: cross-entropy loss for CSL and EXP, L1 loss for ZINC and PCQM4Mv2-subset, binary cross-entropy with logits for MolHIV, and masked multi-label binary cross-entropy with logits for MolPCBA.

For downstream evaluation, CondPSE node encodings are generated once with a fixed NormalSE probe realization for each graph and stored. Downstream training and inference reuse these stored encodings.

For CSL and EXP, downstream models are trained for 200 epochs using the Adam optimizer with a Reduce-on-Plateau learning rate scheduler. For molecular datasets, downstream models are trained using AdamW for 2000, 100, 100, and 100 epochs on ZINC, MolHIV, MolPCBA, and PCQM4Mv2-subset, respectively, with a cosine decay learning rate scheduler and warmup. The adapter and training parameters are reported in Table 5.

Table 5: Downstream adapter and training hyperparameters. Dropout rates are reported before and after the adapter projection. The ZINC row reports the CondPSE adapter/training setting used with the GPS backbone in Table 2. CondPSE experiments in the ZINC downstream backbone sweep reuse this ZINC adapter setting while varying the downstream architecture.

Dataset	Backbone	Adapter	d_{pe}	Dropout	Batch Size	Learning Rate
CSL	GIN	Linear	128	0.0 / 0.0	128	2×10^{-3}
EXP	GIN	Linear	64	0.0 / 0.0	128	2×10^{-3}
ZINC	GPS	Linear	32	0.3 / 0.1	32	1×10^{-3}
MolHIV	GPS	Linear	32	0.3 / 0.1	32	1×10^{-4}
MolPCBA	GPS	2-layer MLP	48	0.3 / 0.1	512	5×10^{-4}
PCQM4Mv2-subset	GPS	Linear	64	0.2 / 0.1	256	2×10^{-4}

GenAI Usage Disclosure

AI tools were used to assist with light editing and grammar. The authors are responsible for all scientific content, experimental results, claims, and final manuscript decisions.