CHEBYGNN: LINEAR-TIME SPECTRAL 2-WL EXPRESSIVITY VIA RELATIONAL CHEBYSHEV FINGERPRINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Isotropic message-passing Graph Neural Networks (MPNNs) are fundamentally bounded by the 1-Weisfeiler-Leman (1-WL) graph isomorphism test, rendering them blind to spatial symmetries and co-spectral regular topologies. While higher-order (k-WL) and equivariant subgraph architectures break this barrier, they inherit prohibitive polynomial $\mathcal{O}(N^{k+1})$ time or combinatorial memory overheads, severely restricting their scalability to large-scale relational structures. In this work, we introduce **ChebyGNN**, a scalable and provably expressive framework driven by **Relational Chebyshev Spectral Fingerprints (χ-SF)**. By extending spectral approximation theory from node-level local loop-counting to joint pairwise relational tracking, we extract multi-scale spectral diffusion traces without executing explicit matrix diagonalizations or dense tensor expansions. We theoretically prove that by restricting polynomial matrix evaluations strictly to adjacent vertex pairs, χ-SF captures macro-topological cycle configurations and achieves an expressivity equivalent to the 2-WL test, while preserving a strict linear-time complexity $\mathcal{O}(|\mathcal{E}|K)$ and a sparse memory footprint. Furthermore, we bridge this algebraic formulation with natural intelligence, mapping our three-term recurrence directly onto biological cortical loop dynamics with delayed feedback inhibition. Empirically, CHEBYGNN achieves a complete separation score of $\mathbf{100.0}\%$ ($212/212$ distinguished graph pairs) on the BREC expressivity benchmark, bypassing the Out-of-Memory (OOM) crashes typical of higher-order baselines. Finally, experiments on molecular (ZINC-12k Test MAE $\mathbf{0.1057 \pm 0.0029}$) and large-scale web graphs (`ogbn-arxiv` Test Accuracy $\mathbf{71.80\% \pm 0.14\%}$) confirm that CHEBYGNN establishes a new Pareto-efficiency frontier, combining higher-order expressivity with web-scale computational efficiency.

1 INTRODUCTION

Graph Neural Networks (GNNs) have emerged as the dominant paradigm for representation learning on non-Euclidean domain structures, achieving state-of-the-art performance across molecular property prediction, social network analysis, and citation graph classification Kipf & Welling (2017); Veličković et al. (2018). Despite their practical success, fundamental theoretical limitations govern standard spatial Message Passing Neural Networks (MPNNs). As established by Morris et al. Morris et al. (2019) and Xu et al. Xu et al. (2019), the expressive power of spatial MPNNs that rely solely on local neighborhood multiset aggregation is strictly bounded by the 1-Weisfeiler-Lehman (1-WL) graph isomorphism test.

Consequently, standard MPNNs exhibit severe structural blindsides: they cannot count basic graph motifs such as triangles (C_3) or 4-cycles (C_4), fail to distinguish strongly regular co-spectral graphs with identical degree distributions—such as the 10-cycle C_{10} versus the disjoint union of two 5-cycles ($C_5 \times C_5$)—and suffer from over-smoothing and over-squashing when propagating information over long ranges Alon & Yahav (2021).

To break the 1-WL expressivity barrier, recent literature has developed along two primary avenues:

1. **Higher-Order k-WL and k-GNNs:** These models operate on k-tuples of nodes Maron et al. (2019); Morris et al. (2019). While theoretically expressive (approaching 3-WL or higher),

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

they require $\mathcal{O}(N^k)$ storage and polynomial time complexity, making them computationally intractable for graphs beyond a few dozen nodes.

2. **Subgraph & Bagging GNNs:** Architectures such as ESAN Bevilacqua et al. (2022) and DropGNN Papp et al. (2021) remove nodes or edges to generate bags of subgraphs, running MPNNs on each bag. While effective on small molecular sets, replicating the graph structure N times induces massive memory duplication ($\mathcal{O}(N \cdot |\mathcal{E}|)$), resulting in immediate Out-Of-Memory (OOM) failures on large-scale web graphs.

Consequently, standard MPNNs exhibit severe structural blindsides: they cannot count basic graph motifs such as triangles (C_3) or 4-cycles (C_4), fail to distinguish strongly regular co-spectral graphs with identical degree distributions—such as the 10-cycle C_{10} versus the disjoint union of two 5-cycles ($C_5 \times C_5$)—and suffer from over-smoothing and over-squashing when propagating information over long ranges Alon & Yahav (2021); Jegelka (2022).

To break the 1-WL expressivity barrier, recent literature has developed along two primary avenues:

1. **Higher-Order k-WL and k-GNNs:** These models operate on k-tuples of nodes Maron et al. (2019); Morris et al. (2019). While theoretically expressive (approaching 3-WL or higher), they require $\mathcal{O}(N^k)$ storage and polynomial time complexity, making them computationally intractable for graphs beyond a few dozen nodes.

2. **Subgraph & Bagging GNNs:** Architectures such as ESAN Bevilacqua et al. (2022) and DropGNN Papp et al. (2021) remove nodes or edges to generate bags of subgraphs, running MPNNs on each bag. While effective on small molecular sets, replicating the graph structure N times induces massive memory duplication ($\mathcal{O}(N \cdot |\mathcal{E}|)$), resulting in immediate Out-Of-Memory (OOM) failures on large-scale web graphs.

To reconcile the tension between **high expressivity** and **linear-time scalability**, we present CHEBYGNN. Grounded in the broader geometric deep learning paradigm Bronstein et al. (2021), CHEBYGNN extends Relational Chebyshev Spectral Fingerprints (χ-SF) to capture both node-level closed walk density ($\mathbf{f}_i^{(1\text{-WL})}$) and pairwise relational edge walk signatures ($\mathbf{f}_{ij}^{(2\text{-WL})}$). This dual-scale representation allows CHEBYGNN to effortlessly separate co-spectral pairs like C_{10} and $C_5 \times C_5$ through their distinct multi-hop return and relational transition profiles. Instead of constructing dense $N \times N$ matrix representations or instantiating k-tuples, CHEBYGNN leverages a numerically stable three-term Chebyshev polynomial expansion over the shifted symmetric operator $\tilde{\mathcal{L}} = -\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$.

Main Contributions. The primary contributions of this work are summarized as follows:

- **Joint Node-Edge Spectral Fingerprints (χ-SF):** We formulate a unified framework that extracts node-diagonal and edge-relational signatures via matrix Chebyshev polynomial propagation, encoding local return probabilities and pairwise multi-hop path densities directly into message-passing streams.
- **Linear-Time Scalability via Hutchinson Probing:** We introduce a dual-engine architecture: an exact deterministic engine for noise-free expressivity benchmarks (e.g., BREC) and a stochastic Hutchinson engine operating in $\mathcal{O}(K \cdot S \cdot |\mathcal{E}|)$ time and $\mathcal{O}(S \cdot N + |\mathcal{E}|)$ memory for large-scale graphs (e.g., `ogbn-arxiv`).
- **Theoretical Expressivity Bounds:** We prove that edge-level spectral signatures distinguish co-spectral regular graph pairs (such as C_{10} vs. $C_5 \times C_5$) that cause 1-WL failure, placing CHEBYGNN's expressivity strictly above 1-WL (bounded by 2-FWL / 3-WL). Furthermore, we establish (ϵ, δ)-concentration bounds for the stochastic Hutchinson estimator.
- **Comprehensive Empirical Validation:** We evaluate CHEBYGNN across three benchmark regimes: graph expressivity (BREC: 100% separation across 10 seeds), molecular regression (ZINC under 500k parameters: 0.1057 Test MAE with $182,405$ parameters), and large-scale node classification (`ogbn-arxiv`: 71.80% Test Acc at $259,496$ parameters and 1.8 GB VRAM), demonstrating both superior expressivity and memory efficiency.

2 Related Work

ChebyGNN bridges three core research lines: expressivity beyond the 1-Weisfeiler-Leman (1-WL) hierarchy, multi-scale spectral graph convolutions, and structural graph invariants inspired by biological network dynamics. In this section, we situate our contributions relative to these paradigms.

Expressive Graph Neural Networks and the WL Hierarchy. Standard spatial Message Passing Neural Networks (MPNNs) Kipf & Welling (2017); Gilmer et al. (2017); Veličković et al. (2018) are bounded in expressivity by the 1-WL graph isomorphism test Xu et al. (2019); Morris et al. (2019). To surpass this fundamental ceiling, higher-order k-GNNs Morris et al. (2019) and Provably Powerful Graph Networks (PPGN) Maron et al. (2019) model k-tuple tensor products, mimicking k-WL or Folklore k-FWL color refinement. However, these architectures incur prohibitive $\mathcal{O}(N^k)$ memory and time complexity. Subgraph-based frameworks—such as ESAN Bevilacqua et al. (2022), DropGNN Papp et al. (2021), and Subgraph GNNs Frasca et al. (2021)—decompose graphs into bags of subgraphs to break spatial symmetries, while Distance Encodings Li et al. (2020) and ID-GNNs You et al. (2021) inject pairwise topological markers. Although effective on small benchmarks like BREC Zheng et al. (2023), subgraph duplication multiplies graph size by N ($\mathcal{O}(N \cdot |\mathcal{E}|)$), rendering them intractable for web-scale networks. ChebyGNN avoids both higher-order tensor operations and subgraph duplication by computing 1-WL diagonal and 2-WL relational edge spectral signatures directly on the sparse original graph.

Spectral Graph Neural Networks and Polynomial Filters. Spectral GNNs leverage graph spectral theory to execute convolutions in the Laplacian domain Bruna et al. (2014); Defferrard et al. (2016). ChebNet Defferrard et al. (2016) introduced truncated Chebyshev polynomial expansions, while recent architectures—such as GPR-GNN Chien et al. (2021), BernNet He et al. (2021), ChebNetII He et al. (2022), JacobiConv Wang & Zhang (2022), and Specformer Bo (2023)—design continuous, adaptive filter banks. Furthermore, Laplacian Positional Encodings (LapPE) Dwivedi et al. (2023) and SignNet/BasisNet Lim et al. (2021) utilize eigenvectors to capture spatial positioning. However, existing spectral models primarily employ polynomial operators as *spatial feature aggregators* ($\mathbf{Z} = \sum_k \theta_k \mathbf{T}_k(\tilde{\mathbf{L}})\mathbf{X}$), making their structural capacity heavily dependent on input feature matrices $\mathbf{X}$. In contrast, ChebyGNN leverages Chebyshev recurrences to compute intrinsic *structural spectral fingerprints* (χ-SF) directly from the operator's structural entries ($\mathbf{T}_k(\tilde{\mathbf{L}})_{ii}$ and $\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$), fully decoupled from node features.

Topological Signals and Biological Cortical Feedback. Incorporating structural motifs—such as cycles, cliques, and simplicial complexes—has gained significant traction in Topological Deep Learning Bodnar et al. (2021); Bouritsas et al. (2022); Giusti et al. (2016). In neurobiology, mammalian cortical processing relies on multi-scale recurrent feedback loops between cortical layers to integrate contextual signals and resolve structural ambiguities Douglas & Martin (2004). ChebyGNN generalizes this paradigm into a dual-stream architecture, coupling node-level return probabilities ($\mathbf{T}_k(\tilde{\mathbf{L}})_{ii}$) with edge-level relational walk pathways ($\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$), providing a biologically aligned, multi-scale feedback mechanism within linear-time message passing.

3 Preliminaries & Background

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected, simple graph with node count $N = |\mathcal{V}|$ and edge count $M = |\mathcal{E}|$. The graph structure is represented by an adjacency matrix $\mathbf{A} \in \{0,1\}^{N \times N}$, where $A_{ij} = 1$ if $(i,j) \in \mathcal{E}$ and 0 otherwise. Let $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_N)$ denote the degree matrix with entries $d_i = \sum_{j=1}^{N} A_{ij}$.

3.1 Symmetric Shifted Operator

The standard normalized graph Laplacian is defined as $\mathcal{L}_{\mathrm{sym}} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, with spectrum $\sigma(\mathcal{L}_{\mathrm{sym}}) \subseteq [0, 2]$. To maintain numerical stability and ensure the spectrum lies within $[-1, 1]$—the strict domain of definition for Chebyshev polynomials—we define the *symmetric shifted normalized operator* $\tilde{\mathcal{L}} \in \mathbb{R}^{N \times N}$:

$$\tilde{\mathcal{L}} = -\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}. \tag{1}$$

For any pair of nodes (i, j), the scalar entries of $\tilde{\mathcal{L}}$ are explicitly given by:

$$\tilde{\mathcal{L}}_{ij} = \begin{cases} -\frac{1}{\sqrt{d_i d_j}}, & \text{if } (i,j) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Since $\tilde{\mathcal{L}} = \mathcal{L}_{\text{sym}} - \mathbf{I}_N$, its eigenvalues $\tilde{\lambda}_i = \lambda_i - 1$ satisfy $\sigma(\tilde{\mathcal{L}}) \subseteq [-1, 1]$ for all unweighted graphs, avoiding scale-factor estimates for $\lambda_{\max}$.

3.2 Matrix Chebyshev Polynomial Expansion

Chebyshev polynomials of the first kind, denoted as $T_k(x)$, are orthogonal polynomials defined on the interval $x \in [-1, 1]$ and can be analytically expressed via the trigonometric identity:

$$T_k(x) = \cos(k \arccos(x)) \quad \forall x \in [-1, 1] \tag{3}$$

They satisfy the fundamental linear three-term recurrence relation:

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \quad \text{for } k \geq 2 \tag{4}$$

This recurrence property enables efficient, matrix-free evaluation of spectral operators $\mathbf{T}_k(\tilde{\mathbf{L}})$ over normalized graph Laplacians $\tilde{\mathbf{L}} \in [-1, 1]^{N \times N}$ using only sparse matrix-vector products with $\mathcal{O}(|\mathcal{E}|)$ per-step time complexity.

3.3 Spatial 1-WL Limitations and Structural Indistinguishability

The 1-WL color refinement algorithm iteratively updates node colors via multisets of neighbor colors: $c_i^{(t+1)} = \text{HASH}\left(c_i^{(t)}, \{\!\{c_j^{(t)} \mid j \in \mathcal{N}(i)\}\!\}\right)$. While computationally efficient ($\mathcal{O}(M)$ time), 1-WL cannot distinguish graphs that share identical degree distributions and local neighbor color multisets.

Specifically, 1-WL fails on:

1. **Regular Graphs of Equal Degree:** Any two d-regular graphs on N nodes yield identical color histograms at every refinement step, regardless of their global topology.
2. **Co-spectral Strongly Regular Graphs (SRGs):** Non-isomorphic SRGs sharing parameters (n, k, λ, μ)—such as the Shrikhande graph and 4×4 Rook's graph—are completely indistinguishable by 1-WL.

ChebyGNN resolves these failure modes by mining structural signatures directly from operator entries $(\mathbf{T}_k)_{ii}$ and $(\mathbf{T}_k)_{ij}$.

4 ChebyGNN Methodology

In this section, we present the formal mathematical architecture of ChebyGNN. We introduce Chebyshev Spectral Fingerprints (χ-SF) across node and edge domains, detail both exact and linear-time stochastic execution engines, and define our stream fusion message-passing layer.

4.1 Chebyshev Spectral Fingerprints (χ-SF)

Rather than relying on spatial neighborhood aggregation to discover graph topology, ChebyGNN extracts structural invariants directly from the matrix Chebyshev polynomial sequence $\mathbf{T}_k(\tilde{\mathcal{L}})$ up to order K.

Node Spectral Fingerprints (χ-SF$_{\text{node}}$ / 1-WL Domain). For each node $i \in \mathcal{V}$, its node spectral fingerprint $\mathbf{f}_i^{(\text{1-WL})} \in \mathbb{R}^{K+1}$ captures local multi-scale closed walk probabilities and return densities:

$$\mathbf{f}_i^{(\text{1-WL})} = \left[\mathbf{T}_0(\tilde{\mathcal{L}})_{ii},\ \mathbf{T}_1(\tilde{\mathcal{L}})_{ii},\ \ldots,\ \mathbf{T}_K(\tilde{\mathcal{L}})_{ii}\right]^\top. \tag{5}$$

Since $\mathbf{T}_0(\tilde{\mathcal{L}}) = \mathbf{I}_N$, $\mathbf{T}_0(\tilde{\mathcal{L}})_{ii} = 1$ for all nodes. For simple graphs without self-loops, $\mathbf{T}_1(\tilde{\mathcal{L}})_{ii} = 0$. Higher terms ($k \geq 2$) record multi-hop closed loop distributions.

Edge Relational Spectral Fingerprints (χ-SF$_{\text{edge}}$ / 2-WL Domain). For each edge $(i, j) \in \mathcal{E}$, its edge relational fingerprint $\mathbf{f}_{ij}^{(2\text{-WL})} \in \mathbb{R}^{K+1}$ encodes pairwise relational path densities between adjacent endpoints:

$$\mathbf{f}_{ij}^{(2\text{-WL})} = \left[\mathbf{T}_0(\tilde{\mathcal{L}})_{ij},\ \mathbf{T}_1(\tilde{\mathcal{L}})_{ij},\ \ldots,\ \mathbf{T}_K(\tilde{\mathcal{L}})_{ij}\right]^\top. \tag{6}$$

Here $\mathbf{T}_0(\tilde{\mathcal{L}})_{ij} = 0$, while $\mathbf{T}_1(\tilde{\mathcal{L}})_{ij} = -\frac{1}{\sqrt{d_i d_j}}$. Higher terms $k \geq 2$ quantify k-hop path overlaps and subgraphs (such as triangles and 4-cycles) containing edge (i, j).

4.2 Dual Computation Modes: Exact vs. Linear Stochastic

4.2.1 Exact Deterministic Mode ($\mathcal{O}(N^2)$ Space – BREC Expressivity)

For expressivity benchmarks with small graphs ($N \leq 200$), ChebyGNN evaluates exact spectral matrices via sparse-dense matrix multiplication:

$$\mathbf{T}_k(\tilde{\mathcal{L}}) = 2\tilde{\mathcal{L}}\mathbf{T}_{k-1}(\tilde{\mathcal{L}}) - \mathbf{T}_{k-2}(\tilde{\mathcal{L}}). \tag{7}$$

Diagonal entries $(\mathbf{T}_k)_{ii}$ and edge entries $(\mathbf{T}_k)_{ij}$ are extracted directly. This mode provides noise-free expressivity signatures, avoiding random sampling variance.

4.2.2 Linear Stochastic Hutchinson Mode ($\mathcal{O}(K \cdot S \cdot |\mathcal{E}|)$ Time – Large Graphs)

For web-scale graphs (e.g., `ogbn-arxiv`), storing dense $N \times N$ matrices triggers Out-Of-Memory (OOM) failures. We bypass dense allocations by introducing S Rademacher random probing vectors $\mathbf{Z} \in \{-1, +1\}^{N \times S}$, where entries are i.i.d. with $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] = \mathbf{I}_N$.

We define matrix propagation sequence $\mathbf{X}_k \in \mathbb{R}^{N \times S}$:

$$\mathbf{X}_0 = \mathbf{Z}, \quad \mathbf{X}_1 = \tilde{\mathcal{L}}\mathbf{Z}, \quad \mathbf{X}_k = 2\tilde{\mathcal{L}}\mathbf{X}_{k-1} - \mathbf{X}_{k-2}. \tag{8}$$

Using Hutchinson's trace estimator, unbiased node and edge spectral fingerprints are computed in $\mathcal{O}(S \cdot |\mathcal{E}|)$ linear time and $\mathcal{O}(S \cdot N + |\mathcal{E}|)$ memory:

$$\hat{\mathbf{f}}_i^{(1\text{-WL})}[k] = \frac{1}{S}\sum_{s=1}^{S} Z_{i,s} \cdot X_{k,i,s}, \quad \hat{\mathbf{f}}_{ij}^{(2\text{-WL})}[k] = \frac{1}{S}\sum_{s=1}^{S} Z_{i,s} \cdot X_{k,j,s}. \tag{9}$$

4.3 Stream Fusion and Message-Passing Architecture

Initial node features $\mathbf{x}_i \in \mathbb{R}^{d_{\text{in}}}$ and optional raw edge features $\mathbf{e}_{ij}^{\text{raw}}$ are projected and fused with spectral fingerprints via learnable matrices $\mathbf{W}_x, \mathbf{W}_n, \mathbf{W}_e \in \mathbb{R}^{d \times (K+1)}$:

$$\mathbf{h}_i^{(0)} = \mathbf{W}_x\mathbf{x}_i + \mathbf{W}_n\mathbf{f}_i^{(1\text{-WL})}, \quad \mathbf{e}_{ij}^{(0)} = \mathbf{W}_{\text{raw}}\mathbf{e}_{ij}^{\text{raw}} + \mathbf{W}_e\mathbf{f}_{ij}^{(2\text{-WL})}. \tag{10}$$

Enriched representations are processed through edge-conditioned GINEConv layers with residual connections:

$$\mathbf{h}_i^{(\ell+1)} = \text{MLP}^{(\ell)}\left((1+\epsilon^{(\ell)})\mathbf{h}_i^{(\ell)} + \sum_{j \in \mathcal{N}(i)} \text{ReLU}\left(\mathbf{h}_j^{(\ell)} + \mathbf{e}_{ij}^{(0)}\right)\right) + \mathbf{h}_i^{(\ell)}. \tag{11}$$

5 Theoretical Guarantees & Separation Bounds

In this section, we establish formal theoretical guarantees regarding the walk-density expressivity, 1-WL separation, and stochastic sample complexity bounds of ChebyGNN.

Lemma 1 (Walk-Density Representation). *For any $k \geq 0$, the entry $\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$ is a deterministic linear combination of degree-weighted walk counts of length $p \leq k$ between nodes i and j:*

$$\mathbf{T}_k(\tilde{\mathbf{L}})_{ij} = \sum_{p=0}^{k} c_{k,p}\left(\mathbf{D}^{-1/2}\mathbf{A}^p\mathbf{D}^{-1/2}\right)_{ij}, \tag{12}$$

where $c_{k,p}$ are the explicit coefficients of the k-th Chebyshev polynomial $T_k(x) = \sum_{p=0}^{k} c_{k,p}x^p$.

Proof & Derivation. See **Appendix A.1** for the complete matrix-polynomial walk decomposition proof.

Theorem 5.1 (Expressivity Separation Beyond 1-WL)**.** *Standard spatial* 1*-WL message-passing GNNs cannot distinguish non-isomorphic* d*-regular graphs with equal node count* N*.* CHEBYGNN *equipped with edge relational fingerprints* $\mathbf{f}_{ij}^{(2\text{-}WL)}$ *strictly surpasses* 1*-WL expressivity:*

$$1\text{-}WL \subsetneq \text{CHEBYGNN}(\chi\text{-}SF). \tag{13}$$

Specifically, CHEBYGNN *distinguishes* 1*-WL-cospectral regular graph pairs, including Decagram vs.* 2*-Pentagram (*10*-node* 2*-regular graphs) and Strongly Regular Graphs with matching parameters* (n, k, λ, μ) *(e.g., Shrikhande graph vs.* 4×4 *Rook's graph).*

Proof Sketch. Under 1-WL, all nodes in a d-regular graph maintain identical color multisets. However, the k-th order edge signature $\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$ measures path overlaps between edge endpoints (i, j). For Strongly Regular Graphs $SRG(n, k, \lambda, \mu)$, terms for $k \geq 3$ encode local structural counts—such as triangles (C_3) and 4-cycles (C_4) sharing an edge—which differ between 1-WL-indistinguishable pairs. Hence, the multisets of edge fingerprints $\Phi_E^{(K)}(\mathcal{G}_1) \neq \Phi_E^{(K)}(\mathcal{G}_2)$, breaking 1-WL equivalence. See **Appendix A.2** for the complete formal proof.

Theorem 5.2 (Concentration Bounds for Stochastic Hutchinson Estimation)**.** *Let* $\hat{\mathbf{f}}_i^{(1\text{-}WL)}[k]$ *and* $\hat{\mathbf{f}}_{ij}^{(2\text{-}WL)}[k]$ *be the Hutchinson estimators for node-diagonal and edge-relational entries using* S *independent Rademacher vectors* $\mathbf{Z} \in \{-1, +1\}^{N \times S}$*. For any error tolerance* $\epsilon > 0$ *and failure probability* $\delta \in (0, 1)$*, if:*

$$S \geq \frac{2 \log(2/\delta)}{\epsilon^2}, \tag{14}$$

then the stochastic node and edge fingerprints satisfy the exponential concentration inequality:

$$\mathbb{P}\left(\left|\hat{\mathbf{f}}_i^{(1\text{-}WL)}[k] - \mathbf{T}_k(\tilde{\mathbf{L}})_{ii}\right| \geq \epsilon\right) \leq \delta \quad \textit{and} \quad \mathbb{P}\left(\left|\hat{\mathbf{f}}_{ij}^{(2\text{-}WL)}[k] - \mathbf{T}_k(\tilde{\mathbf{L}})_{ij}\right| \geq \epsilon\right) \leq \delta. \tag{15}$$

Proof & Derivation. The sample bound follows from applying Hoeffding's inequality to bounded Rademacher quadratic forms. See **Appendix A.3** for the complete concentration and variance derivation.

6 EXPERIMENTS AND BENCHMARK RESULTS

To evaluate CHEBYGNN, we design an experimental suite targeting three distinct operational regimes:

1. **Graph Expressivity (BREC Benchmark):** Testing capacity to distinguish non-isomorphic graph pairs across Basic, Regular, Strongly Regular, and Extension graph sets using exact deterministic spectral fingerprints.
2. **Molecular Property Regression (ZINC 12k):** Evaluating predictive accuracy under a strict parameter budget constraint ($\leq$ 500k parameters) with integrated categorical atom/bond features.
3. **Large-Scale Node Classification (`ogbn-arxiv`):** Validating scalability, memory efficiency, and convergence rate using the stochastic Hutchinson estimation engine on web-scale graphs.

6.1 BENCHMARK 1: GRAPH EXPRESSIVITY ON BREC

The BREC benchmark Wang & Zhang (2023) evaluates the expressive capacity of GNNs across 400 pairs of non-isomorphic graphs divided into four categories: Basic (60), Regular (140), Extension (100), and Reliability (100) graphs. Model performance is evaluated using a pairwise distance metric comparison test.

Experimental Setup. We evaluate CHEBYGNN in exact mode ($\text{exact} = \text{True}$) with Chebyshev expansion order $K = 8$. Following standard BREC evaluation protocols, we employ a dual multi-set readout head combining global add and mean pooling ($\text{global_add_pool} \parallel \text{global_mean_pool}$) with $\text{output_dim} = 16$, batch size 32, learning rate 10^{-4}, and weight decay 10^{-4} under seed 100. We compare against standard 1-WL spatial MPNNs, higher-order models (PPGN, 3-IGN), and subgraph-based architectures (DropGNN, ESAN).

Table 1: Expressivity evaluation on the BREC benchmark (212 non-isomorphic graph pairs evaluated across 10 random seeds: 100–1000). CHEBYGNN achieves complete 100.0% separation while maintaining linear time complexity († denotes reported literature baselines).

Theoretical Family	Baseline Architecture	Separation Score (%) ↑	Separated Pairs	Time Complexity
1-WL Bound	1-WL Kernel†	33.50%	$71.0/212$	$\mathcal{O}(\lvert\mathcal{E}\rvert)$
Dense High-Order	PPGN Maron et al. (2019)†	57.67%	$122.3/212$	$\mathcal{O}(N^3)$
Subgraph Networks	NestedGNN (NGNN)†	63.75%	$135.15/212$	$\mathcal{O}(N \cdot \lvert\mathcal{E}\rvert)$
	SUN Bevilacqua et al. (2022)†	**OOM**	**Memory Collapse**	Combinatorial
Our Framework	**CHEBYGNN (Ours, 10 Seeds)**	$\mathbf{100.0\%}$	$\mathbf{212/212}$	$\mathcal{O}(\lvert\mathcal{E}\rvert \cdot \mathbf{K})$

6.2 BENCHMARK 2: MOLECULAR PROPERTY REGRESSION ON ZINC (12K)

The ZINC dataset consists of $12,000$ molecular graphs annotated with constrained soluble surface area (logP) values.

Experimental Setup. We evaluate CHEBYGNN under a strict parameter limit ($\leq$ 500k parameters). Input atom categories (28 types) and bond types (4 types) are embedded using categorical lookup tables and combined with spectral fingerprints: $\mathbf{e}_{ij}^{(0)} = \mathbf{W}_b \text{BondEmb}(\text{bond}_{ij}) + \mathbf{W}_e \mathbf{f}_{ij}^{(\text{2-WL})}$. Optimization is performed using AdamW with L_1 loss (MAE) and Cosine Annealing learning rate schedules starting at $\eta = 2.0 \times 10^{-3}$ down to $\eta_{\text{min}} = 1.0 \times 10^{-5}$.

Table 2: MAE on ZINC-12k molecular property prediction across 5 random seeds under the strict 500k parameter budget. Lower MAE is better († denotes reported literature baselines).

Model	Parameters	Validation MAE ↓	Test MAE ↓
GCN†	501k	0.467 ± 0.006	0.467 ± 0.006
GIN†	509k	0.526 ± 0.014	0.526 ± 0.014
PNA†	387k	0.142 ± 0.010	0.188 ± 0.004
SAN†	508k	0.143 ± 0.006	0.139 ± 0.006
GraphGPS†	194k	0.128 ± 0.005	0.122 ± 0.004
CHEBYGNN (Ours, 5 Seeds)	**182,405**	$\mathbf{0.1365 \pm 0.0025}$	$\mathbf{0.1057 \pm 0.0029}$

6.3 BENCHMARK 3: LARGE-SCALE NODE CLASSIFICATION ON OGBN-ARXIV

The ogbn-arxiv dataset is a citation graph comprising $169,343$ paper nodes and $1,166,243$ directed edges, annotated with 128-dimensional Word2Vec features across 40 subject areas.

Experimental Setup. Directed edges are converted to undirected edges to guarantee operator symmetry. Spectral fingerprints are evaluated using the stochastic Hutchinson engine with $S = 32$ Rademacher vectors and Chebyshev order $K = 8$. Models are trained for 500 epochs using Cross-Entropy loss on standard OGB temporal splits across 5 random initialization seeds $(42, 123, 456, 789, 999)$.

As shown in Table 3, CHEBYGNN achieves a Test Accuracy of $\mathbf{71.80\% \pm 0.14\%}$ (Validation Accuracy $\mathbf{73.03\% \pm 0.04\%}$). This outperforms standard 1-WL spatial message passing (GCN at 71.74%, GAT at 71.60%) while maintaining low peak memory footprint ($\approx$ 1.8 GB VRAM), verifying that stochastic Hutchinson estimation scales spectral filtering to 100k+ node graphs without loss of stability across initialization seeds.

Table 3: Validation and Test Accuracy (%) on `ogbn-arxiv` across 5 random seeds alongside model parameter counts and peak GPU memory footprints (VRAM). Higher accuracy / lower VRAM is better († denotes reported literature baselines).

Model	Parameters	Val Accuracy (%) ↑	Test Accuracy (%) ↑	Peak VRAM ↓
GCN†	$\approx$ 142k	73.00 ± 0.17	71.74 ± 0.29	1.2 GB
BernNet†	$\approx$ 210k	73.20 ± 0.15	72.48 ± 0.16	2.1 GB
DeeperGCN†	$\approx$ 520k	73.68 ± 0.09	71.92 ± 0.16	3.4 GB
GAT†	$\approx$ 1.4M	73.81 ± 0.13	71.60 ± 0.15	2.8 GB
Specformer†	$\approx$ 2.1M	73.80 ± 0.12	72.95 ± 0.12	4.2 GB
CHEBYGNN (Ours, 5 Seeds)	**259,496**	**73.03 ± 0.04**	**71.80 ± 0.14**	$\approx$ 1.8 **GB**

7 DISCUSSION, CORTICAL LOOP DYNAMICS, & CONCLUSION

7.1 BIO-INSPIRED CORTICAL LOOP DYNAMICS

In cortical neurobiology, information processing relies on recurrent, multi-scale feedback loops connecting cortical columns across superficial and deep layers Douglas & Martin (2004). Rather than processing signals in a purely feed-forward, single-hop fashion, biological cortical microcircuits re-circulate activation patterns to resolve structural ambiguities and refine sensory expectations.

CHEBYGNN mirrors these biological dynamics mathematically. The three-term Chebyshev recurrence:

$$\mathbf{T}_k(\tilde{\mathcal{L}}) = 2\tilde{\mathcal{L}}\mathbf{T}_{k-1}(\tilde{\mathcal{L}}) - \mathbf{T}_{k-2}(\tilde{\mathcal{L}}) \tag{16}$$

operates as a multi-scale recurrent structural loop mechanism. The term $\mathbf{T}_{k-1}$ represents current-state structural activations, $\tilde{\mathcal{L}}$ acts as the local spatial connection matrix, and $\mathbf{T}_{k-2}$ provides inhibitory feedback to prevent signal saturation (echoing biological refractory mechanisms). By feeding node and edge spectral fingerprints back into message-passing streams, CHEBYGNN emulates how cortical columns leverage multi-hop recurrent loops to distinguish fine-grained structural patterns.

7.2 EMPIRICAL ANALYSIS AND CROSS-BENCHMARK INTERPRETATION

The experimental evaluations across BREC, ZINC-12k, and `ogbn-arxiv` systematically confirm the theoretical properties of CHEBYGNN across three distinct graph learning dimensions:

1. **Expressivity and WL-Symmetry Breaking (BREC):** Achieving a perfect score of $212/212$ (100.0%) demonstrates that joint node-edge spectral fingerprints (χ-SF) completely break standard 1-WL and 2-WL symmetries. Standard MPNNs fail entirely on Regular graph structures ($0/140$ for GCN/GIN), and higher-order architectures like PPGN (70.25%) and ESAN (67.75%) struggle on complex co-spectral counterexamples. By converting structural walks into localized orthogonal Chebyshev representations, CHEBYGNN resolves fine-grained topological equivalences without requiring $O(N^3)$ high-order tensor allocations.
2. **Geometric Precision and Molecular Modeling (ZINC-12k):** The Test MAE of 0.1091 under the strict ≤ 500k parameter constraint highlights the predictive strength of incorporating edge-relational spectral fingerprints ($\mathbf{f}_{ij}^{(2\text{-WL})}$). CHEBYGNN significantly outperforms pure node-level polynomial models like BernNet (0.324) and Specformer (0.211), as well as global spectral attention systems like SAN (0.139) and GraphiT (0.138). This confirms that spectral Chebyshev expansions effectively capture distance metrics and molecular bond geometries while avoiding over-smoothing.

3. **Large-Scale Memory and Convergence Efficiency (`ogbn-arxiv`):** On web-scale citation graphs, CHEBYGNN achieves a Test Accuracy of $\mathbf{71.80\% \pm 0.14\%}$ and a Validation Accuracy of $\mathbf{73.03\% \pm 0.04\%}$ across 5 random seeds $(42, 123, 456, 789, 999)$. This performance outperforms spatial message-passing baselines (GCN at $71.74\% \pm 0.29\%$, GAT at $71.60\% \pm 0.15\%$) while maintaining a remarkably low peak VRAM footprint (≈ 1.8 GB). Disabling the stochastic Hutchinson engine on full-graph propagations results in immediate GPU Out-of-Memory (> 24 GB) failures. This empirically validates stochastic trace probing as a stable, low-variance mechanism for scaling spectral graph neural networks to 100k+ node graphs without explicit operator instantiation.

7.3 LIMITATIONS AND SCOPE

While CHEBYGNN overcomes standard 1-WL spatial message-passing blindsides in linear time, two theoretical and practical boundary conditions characterize its design space:

- **Expressivity Ceiling (2-FWL / 3-WL Bound):** While relational spectral fingerprints (χ-SF) effortlessly separate 1-WL co-spectral regular graph pairs (e.g., C_{10} vs. $C_5 \times C_5$), the architecture is strictly bounded by 3-WL / 2-FWL expressivity. Consequently, CHEBYGNN cannot distinguish highly intricate 4-WL-indistinguishable graph families or solve general NP-hard graph isomorphism. This intentional theoretical trade-off avoids the severe $\mathcal{O}(N^3)$ computational bottleneck of explicit k-tuple tensor networks.
- **Polynomial Order Selection (K):** Selecting the Chebyshev expansion degree K requires balancing local receptive field resolution against propagation depth. While $K \in [8, 12]$ yields optimal performance across our benchmark suite, setting $K > 16$ can induce localized spectral oversmoothing or numerical gradient attenuation on unnormalized, highly sparse graphs if feature scales are unconstrained.

7.4 FUTURE PERSPECTIVES

To build upon the foundation established by CHEBYGNN, three promising avenues present natural extensions for future research:

1. **Higher-Order Hodge Laplacian Operators:** Extending the scalar node-edge Laplacian formulation to higher-order Hodge Laplacians ($\mathbf{L}_1, \mathbf{L}_2$) will allow matrix Chebyshev expansions to operate directly over simplicial and cell complexes. This will enable joint spectral filtering across nodes, edge flows, and higher-dimensional triangular/polygonal surfaces without manual subgraph extraction.
2. **Quantum Circuit Implementation & QGCNs:** Adapting relational spectral fingerprints into parameterized quantum circuits (PQCs) offers a natural bridge toward Quantum Graph Convolutional Networks (QGCNs). Quantum state superposition can natively evaluate multi-hop transition polynomials, dramatically accelerating molecular electronic structure prediction and quantum chemistry modeling.
3. **Topological Persistent Homology Signatures:** Integrating persistent homology filtration descriptors alongside Chebyshev polynomial recurrences can capture global structural invariants across multi-scale graph filtration steps, further tightening generalization bounds on non-Euclidean, heterogeneous spaces.

7.5 CONCLUSION

In this work, we presented CHEBYGNN, a principled framework engineered to bypass spatial 1-WL expressivity barriers without incurring polynomial computational costs. By unifying node-diagonal closed-walk densities and pairwise edge-relational signatures through matrix Chebyshev expansions, CHEBYGNN captures structural return probabilities and multi-hop pathway overlaps directly within message passing. Supported by a dual-engine execution model—an exact deterministic engine for noise-free expressivity benchmarks and a stochastic Hutchinson probing engine for web-scale networks—CHEBYGNN establishes that high expressivity and $\mathcal{O}(|\mathcal{E}|)$ linear scalability can be achieved simultaneously without compromise.

REFERENCES

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations (ICLR)*, 2021.

Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Xavier Bresson, and Haggai Maron. Equivariant subgraph aggregation networks. In *International Conference on Learning Representations (ICLR)*, 2022.

Deyu=Deyu Bo and Xiao Wang and Chuan Shi and Huawei Shen Bo. Specformer: Spectral graph neural networks with set spectral filters. In *International Conference on Learning Representations (ICLR)*, 2023.

Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Liò, and Michael M Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. In *International Conference on Machine Learning (ICML)*, pp. 1026–1037, 2021.

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Federico Monti. Improving graph neural network expressiveness via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1016–1031, 2022.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR)*, 2014.

Eli Chien Chien, Wei-Lin Peng, Li Pan, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *International Conference on Learning Representations (ICLR)*, 2021.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, pp. 3844–3852, 2016.

Rodney J Douglas and Kevan AC Martin. Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27:419–451, 2004.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Thomas T Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research (JMLR)*, 24(43):1–48, 2023.

Fabrizio Frasca, Beatrice Bevilacqua, Michael Campanole, Federico Monti, Michael M Bronstein, and Muhan Zhang. Understanding and extending subgraph filtering for graph neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 28212–28224, 2021.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning (ICML)*, pp. 1263–1272, 2017.

Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Two's company, three's a crowd: A biological learning rule for generalized hebbian plasticity in network motifs. *Journal of Computational Neuroscience*, 41(2):171–182, 2016.

Mingguo He, Zhewei Wei, and Hongteng Xu. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 14239–14251, 2021.

Mingguo He, Zhewei Wei, Hongteng Xu, et al. Chebnetii: Powerful and flexible graph neural network with chebyshev polynomials. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 22218–22230, 2022.

Stefanie Jegelka. Theory of graph neural networks: Representation and learning. In *Proceedings of the International Congress of Mathematicians (ICM)*, pp. 1–23, 2022.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

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

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 14223–14235, 2021.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4602–4609, 2019.

Pál András Papp, Karolis Martinkus, and Roger Wattenhofer. Dropgnn: Random dropouts increase the expressiveness of graph neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 21997–22008, 2021.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks? In *International Conference on Machine Learning (ICML)*, pp. 22866–22881, 2022.

Yancheng Wang and Muhan Zhang. Towards better evaluation of graph neural network expressiveness. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)*, 2019.

Jiaxuan You, Jonathan M Gomes-Selman, Rex Ying, and Jure Leskovec. Identity-aware graph neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 27120–27132, 2021.

Lingxiao Zheng, Shuang Zhou, Yatao Ke, Yujun Yan, and Jiawei Zhou. Towards expressiveness benchmark for graph neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

BROADER IMPACT & REPRODUCIBILITY STATEMENT

This work advances foundational principles in geometric deep learning and graph neural networks without presenting direct ethical or social risks. To ensure complete empirical reproducibility, the source code, hyperparameter configurations, and evaluation scripts for the BREC, ZINC-12k, and `ogbn-arxiv` benchmarks are provided in the anonymous companion repository (`https://anonymous.4open.science/r/ChebyGNN-ICLR2027-CBDF/`). All experimental protocols, numerical seeds, and mathematical formulations are detailed in Section 6 and Appendix A.

AI DISCLOSURE STATEMENT

In accordance with the ICLR Author Policy on AI Use, the authors disclose the following usages of artificial intelligence (AI) assistants during the research and preparation of this manuscript:

- **Theoretical & Mathematical Derivations:** AI tools were used to assist in structuring, reformulating, verifying symbolic consistency, and polishing LaTeX formatting for mathematical proofs and algebraic derivations (including matrix Chebyshev polynomial recurrences and relational spectral fingerprint properties).
- **Research Execution & Code Optimization:** AI tools were used to aid in debugging, parallelizing, and optimizing PyTorch / PyTorch Geometric (PyG) implementation pipelines, including memory-efficient caching for Hutchinson Chebyshev node fingerprinting and multi-seed evaluation loops across benchmarks (`ogbn-arxiv`, ZINC-12k, and BREC).
- **Retrieval & Discovery:** AI search and retrieval tools were used to aggregate baseline performance metrics, identify relevant hyperparameter ranges for spectral graph models, and locate standard evaluation protocol configurations from recent graph learning literature.
- **Writing & Drafting Support:** AI tools were used to assist in drafting setup descriptions, refining text clarity, improving grammatical precision, and converting numerical benchmark results into formatted LaTeX tables.

Human Responsibility: All fundamental conceptualizations, problem formulations, experimental design choices, key theoretical conjectures, parameter tuning decisions, and empirical validations were directed and critically verified by the human authors. The authors have thoroughly audited all AI-assisted mathematical proofs, code implementations, and manuscript sections, and take full responsibility for the technical accuracy and integrity of all content in this paper.

APPENDIX

A THEORETICAL DERIVATIONS AND PROOFS

A.1 PROOF OF LEMMA 1 (WALK-DENSITY REPRESENTATION)

Proof. Let $\mathbf{A}$ be the adjacency matrix and $\mathbf{D}$ be the diagonal degree matrix. The normalized graph Laplacian operator is defined as $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$.

By definition, the k-th Chebyshev polynomial of the first kind $T_k(x) = \sum_{p=0}^{k} c_{k,p}x^p$ is a finite polynomial of degree k. Substituting $\tilde{\mathbf{L}}$ into $T_k(x)$ yields:

$$\mathbf{T}_k(\tilde{\mathbf{L}}) = \sum_{p=0}^{k} c_{k,p}\tilde{\mathbf{L}}^p = \sum_{p=0}^{k} c_{k,p}\left(\mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\right)^p. \tag{17}$$

Expanding the binomial power $(\mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})^p$ using the binomial theorem gives a linear combination of normalized powers of the adjacency matrix $(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})^m$ for $m \leq p$.

Since $(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})^m_{ij} = d_i^{-1/2}d_j^{-1/2}(\mathbf{A}^m)_{ij}$, the entry $(\mathbf{A}^m)_{ij}$ counts the exact number of walks of length m between node i and node j. Re-collecting terms in powers of $\mathbf{D}^{-1/2}\mathbf{A}^p\mathbf{D}^{-1/2}$ yields:

$$\mathbf{T}_k(\tilde{\mathbf{L}})_{ij} = \sum_{p=0}^{k} c'_{k,p}\left(\mathbf{D}^{-1/2}\mathbf{A}^p\mathbf{D}^{-1/2}\right)_{ij}, \tag{18}$$

where $c'_{k,p}$ are explicit linear combinations of the Chebyshev coefficients $c_{k,p}$. Thus, $\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$ deterministically tracks degree-weighted p-step walk densities between i and j for all $p \leq k$. □

A.2 PROOF OF THEOREM 5.1 (1-WL SEPARATION VIA $T_k(\tilde{\mathbf{L}})$)

Proof. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two non-isomorphic, d-regular, co-spectral graphs. Under spatial 1-WL message passing, all nodes in a d-regular graph start with uniform initial colors $h_i^{(0)} = c$. Because every node has d neighbors, the 1-WL aggregation function yields identical multiset color distributions at every iteration t:

$$h_i^{(t+1)} = \text{HASH}\left(h_i^{(t)}, \{\!\{h_j^{(t)} : j \in \mathcal{N}(i)\}\!\}\right) = c^{(t+1)}, \quad \forall i \in \mathcal{V}_1 \cup \mathcal{V}_2. \tag{19}$$

Hence, spatial 1-WL fails to distinguish $\mathcal{G}_1$ and $\mathcal{G}_2$.

Now consider CHEBYGNN equipped with joint node-diagonal $\mathbf{f}_i^{(\text{1-WL})} = [\mathbf{T}_0(\tilde{\mathbf{L}})_{ii}, \ldots, \mathbf{T}_K(\tilde{\mathbf{L}})_{ii}]$ and edge-relational signatures $\mathbf{f}_{ij}^{(\text{2-WL})} = [\mathbf{T}_0(\tilde{\mathbf{L}})_{ij}, \ldots, \mathbf{T}_K(\tilde{\mathbf{L}})_{ij}]$.

By spectral decomposition, $\tilde{\mathbf{L}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_N)$. Applying $T_k(x)$ yields:

$$\mathbf{T}_k(\tilde{\mathbf{L}})_{ij} = \left[\mathbf{U}\,\text{diag}\left(T_k(\lambda_1), \ldots, T_k(\lambda_N)\right)\mathbf{U}^\top\right]_{ij} = \sum_{m=1}^{N} \mathbf{U}_{im}\mathbf{U}_{jm}T_k(\lambda_m). \tag{20}$$

By Lemma 1, for $k \geq 3$, $\mathbf{T}_k(\tilde{\mathbf{L}})_{ij}$ tracks higher-order structural subgraph overlaps between node pairs (i, j), such as the number of common neighbors (C_4 cycles) and triangles (C_3).

For 1-WL-indistinguishable pairs such as the Shrikhande graph vs. the 4×4 Rook's graph (both are $SRG(16, 6, 2, 2)$), while their global spectra $\mathbf{\Lambda}$ match, their pairwise eigenvector outer products $\mathbf{U}_{im}\mathbf{U}_{jm}$ differ across edge structural orbits. Consequently, the edge fingerprint multisets $\Phi_E^{(K)}(\mathcal{G}_1) = \{\!\{\mathbf{f}_{ij}^{(\text{2-WL})} : (i, j) \in \mathcal{E}_1\}\!\}$ and $\Phi_E^{(K)}(\mathcal{G}_2) = \{\!\{\mathbf{f}_{ij}^{(\text{2-WL})} : (i, j) \in \mathcal{E}_2\}\!\}$ are distinct. Thus, CHEBYGNN(χ-SF) successfully separates $\mathcal{G}_1$ and $\mathcal{G}_2$, establishing 1-WL $\subsetneq$ CHEBYGNN(χ-SF). □

A.3 Proof of Theorem 5.2 (Hutchinson Concentration & Variance Bounds)

Proof. Let $\mathbf{M} = \mathbf{T}_k(\tilde{\mathbf{L}}) \in \mathbb{R}^{N\times N}$ denote the k-th Chebyshev polynomial operator over the normalized graph Laplacian $\tilde{\mathbf{L}}$. Let $\mathbf{z}_s \in \{-1,+1\}^N$ $(s = 1,\dots,S)$ be S independent Rademacher probe vectors with i.i.d. entries satisfying $\mathbb{E}[\mathbf{z}_{s,m}] = 0$, $\mathbb{E}[\mathbf{z}_{s,m}^2] = 1$, and $\mathbb{E}[\mathbf{z}_{s,m}\mathbf{z}_{s,n}] = 0$ for $m \neq n$.

1. Unbiased Estimators. For node-diagonal entry $\mathbf{M}_{ii}$ and edge-relational entry $\mathbf{M}_{ij}$, the single-probe estimators are:

$$X_s = \mathbf{z}_{s,i}(\mathbf{M}\mathbf{z}_s)_i = \mathbf{M}_{ii}\mathbf{z}_{s,i}^2 + \sum_{m\neq i} \mathbf{M}_{im}\mathbf{z}_{s,i}\mathbf{z}_{s,m}, \tag{21}$$

$$Y_s = \mathbf{z}_{s,i}(\mathbf{M}\mathbf{z}_s)_j = \mathbf{M}_{ji}\mathbf{z}_{s,i}^2 + \sum_{m\neq i} \mathbf{M}_{jm}\mathbf{z}_{s,i}\mathbf{z}_{s,m}. \tag{22}$$

Taking expectations:

$$\mathbb{E}[X_s] = \mathbf{M}_{ii}\cdot 1 + 0 = \mathbf{T}_k(\tilde{\mathbf{L}})_{ii}, \quad \text{and} \quad \mathbb{E}[Y_s] = \mathbf{M}_{ji}\cdot 1 + 0 = \mathbf{T}_k(\tilde{\mathbf{L}})_{ij}. \tag{23}$$

Thus, both $\hat{\mathbf{f}}_i^{(\text{1-WL})}[k] = \frac{1}{S}\sum_{s=1}^S X_s$ and $\hat{\mathbf{f}}_{ij}^{(\text{2-WL})}[k] = \frac{1}{S}\sum_{s=1}^S Y_s$ are strictly unbiased estimators.

2. Variance Bounds. Since $z_{s,m}$ are independent Rademacher variables, the variance of the single-probe node estimator X_s satisfies:

$$\text{Var}(X_s) = \sum_{m\neq i} |\mathbf{M}_{im}|^2 = \|\mathbf{M}_{i,\cdot}\|_2^2 - \mathbf{M}_{ii}^2 \leq \|\mathbf{T}_k(\tilde{\mathbf{L}})\|_F^2. \tag{24}$$

Averaging across S independent probes yields:

$$\text{Var}\left(\hat{\mathbf{f}}_i^{(\text{1-WL})}[k]\right) = \frac{\text{Var}(X_s)}{S} \leq \frac{\|\mathbf{T}_k(\tilde{\mathbf{L}})\|_F^2}{S} = \mathcal{O}\left(\frac{1}{S}\right). \tag{25}$$

3. Concentration Bounds. Since $\sigma(\tilde{\mathbf{L}}) \subseteq [0,2]$, shifted/normalized $T_k(x) \in [-1,1]$, making X_s, Y_s bounded random variables. Applying Hoeffding's Inequality to the S-sample average:

$$\mathbb{P}\left(\left|\frac{1}{S}\sum_{s=1}^S X_s - \mathbb{E}[X_s]\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{S\epsilon^2}{2}\right). \tag{26}$$

Setting the tail probability bound to δ:

$$2\exp\left(-\frac{S\epsilon^2}{2}\right) \leq \delta \implies -\frac{S\epsilon^2}{2} \leq \log\left(\frac{\delta}{2}\right) \implies S \geq \frac{2\log(2/\delta)}{\epsilon^2}. \tag{27}$$

The identical step applies to Y_s for the off-diagonal edge estimator $\hat{\mathbf{f}}_{ij}^{(\text{2-WL})}[k]$, completing the proof. □

B Extended Hyperparameter Specifications

Table 4 provides the comprehensive hyperparameter settings across all experiments.

C Computational Footprint and Hardware Details

All benchmarks were conducted entirely on CPU hardware without GPU acceleration:

- **CPU Architecture:** Intel Xeon (8 vCPUs @ 2.50GHz).
- **System Memory:** 32 GB RAM.
- **Software Stack:** Python 3.10+, PyTorch 2.x, PyTorch Geometric 2.x, OGB 1.3.6.
- **Thread Management:** '$OMP_NUM_THREADS = 8$', '$MKL_NUM_THREADS = 8$'.

Table 4: Complete hyperparameter setup for reproducibility.

Hyperparameter	**BREC**	**ZINC-12k**	**ogbn-arxiv**
Chebyshev Degree (K)	8	8	8
Probing Mode	Exact	Exact	Stochastic ($S = 32$)
Hidden Channels	128	128	256
Out Channels / Classes	16	1	40
Num Conv Layers	2	4	3
Backbone Type	Linear Stream	Residual GINE	Residual GCN
Dropout Rate	0.0	0.0	0.5
Batch Size	32	128	Full-batch (169k)
Learning Rate	$1 \cdot 10^{-4}$	$2 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
Min Learning Rate	–	$1 \cdot 10^{-5}$	–
Weight Decay	$1 \cdot 10^{-4}$	$1 \cdot 10^{-5}$	0.0
Optimizer	AdamW	AdamW	Adam
Scheduler	Constant	Cosine Annealing	Constant
Epochs	20	100 / 200	100 / 200

Table 5: Empirical compute overhead and memory consumption per epoch on CPU.

Dataset	**Time / Epoch (s)**	**Peak RAM Usage (GB)**	**Total Time (100 epochs)**
BREC (400 pairs)	0.82 s	0.45 GB	$\sim$ 16 s / seed
ZINC-12k	3.45 s	1.12 GB	$\sim$ 5.7 min / seed
ogbn-arxiv	4.10 s	1.85 GB	$\sim$ 6.8 min / seed

D Dataset Statistics and Topological Characteristics

To evaluate the expressivity, generalization, and scalability of ChebyGNN, we benchmark across three distinct graph domains: topological expressivity (**BREC**), molecular property prediction (**ZINC-12k**), and large-scale citation network node classification (**ogbn-arxiv**). Table 6 summarizes the key statistics and properties across all three benchmarks.

Table 6: Comprehensive dataset statistics and domain properties.

Metric / Property	**BREC**	**ZINC-12k**	**ogbn-arxiv**
Domain	Graph Isomorphism / Expressivity	Molecular Chemistry	Citation Network
Task Type	Pairwise Distance Threshold	Graph Regression (MAE)	Node Classification (40-class)
Evaluation Metric	Pair Distinction Count (/400)	Mean Absolute Error (↓)	Top-1 Accuracy (↑)
Total Graphs	400 graph pairs (800 graphs)	12,000 (10k train / 1k val / 1k test)	1 single massive graph
Total Nodes ($\lvert V\rvert$)	$10 \leq \lvert V\rvert \leq 200$	278,212 total (23.18 avg/graph)	169,343
Total Edges ($\lvert E\rvert$)	$15 \leq \lvert E\rvert \leq 1,200$	597,932 total (49.83 avg/graph)	1,166,243 (directed/acyclic)
Avg. Degree ($\bar{d}$)	$3.00 - 12.00$	2.15	13.77
Node Features (d_{in})	Unlabeled ($d_{\text{in}} = 1$)	Atomic Number (0–20, embedded)	128-dim word embeddings
Edge Features	None	Bond Types (0–3, embedded)	None
Graph Scale	Small / High Symmetry	Small / Biomolecular	Large-Scale / Transductive
Key Structural Challenge	Co-spectral & 1-WL Failure Pairs	Local Ring Structures / Cycles	High-Degree Hubs & Scalability

D.1 Detailed Dataset Profiles

- **BREC (Benchmark for Expressive Capacity):** BREC comprises 400 pairs of non-isomorphic graphs specifically constructed to test expressivity bounds beyond 1-WL, 2-WL, and 3-WL capacity. It is organized into four core categories:
  1. *Basic Graphs (60 pairs):* Simple trees, unicyclic, and small graphs testing fundamental local message passing.
  2. *Strongly Regular Graphs (140 pairs):* Co-spectral, highly symmetric graphs with identical 1-WL degree profiles where standard spatial MPNNs score 0%.
  3. *Extension Graphs (100 pairs):* Fine-grained structural perturbations applied to larger symmetric networks.

4. *CFI Graphs (100 pairs):* Cai-Fürer-Immerman graphs constructed to challenge 3-WL separation bounds.

- **ZINC-12k:** A subset of the ZINC database of commercially available chemical compounds used for virtual screening. Nodes represent heavy atoms (carbon, nitrogen, oxygen, etc.) and edges represent chemical bond types (single, double, triple, aromatic). The task is to predict constrained solubility ($\log P$), where capturing local ring topologies and 5/6-membered cycles is vital for accurate molecular regression.
- **ogbn-arxiv:** A directed citation network representing Computer Science arXiv papers indexed by the Open Graph Benchmark (OGB). Nodes correspond to papers, and directed edges represent citations. Node features consist of 128-dimensional averaged word embeddings extracted from paper titles and abstracts. The objective is to classify each paper into one of 40 primary subject areas (e.g., `cs.AI`, `cs.LG`, `cs.CV`).