

Nonlinear Diffusion Signatures Cannot Break the 1-WL Barrier: A Systematic Negative Result Connecting Graph Scattering Architectures to the Weisfeiler-Leman Hierarchy

Anonymous Authors
Under Review

Abstract

We investigate whether Nonlinear Diffusion Signatures (NDS) — alternating rounds of linear graph diffusion with fixed pointwise nonlinearities applied to deterministic node features — can break the 1-dimensional Weisfeiler-Leman (1-WL) expressiveness ceiling of standard message-passing graph neural networks (GNNs). The hypothesis, inspired by cross-frequency mode coupling in nonlinear wave physics, posits that interleaved nonlinearity generates eigenvector-dependent features sufficient to distinguish 1-WL-equivalent graphs at $O(m \cdot T)$ preprocessing cost. Through six iterative experimental campaigns spanning 98 verified 1-WL-equivalent graph pairs, 8+ nonlinearities, 10+ initialization strategies, and $T = 1\text{--}20$ diffusion steps totaling over 261 distinct configurations, we provide definitive empirical falsification: nonlinear diffusion provides zero additional discriminative power beyond linear diffusion (`nonlinear_only` = 0 across all configurations). We identify three reinforcing theoretical obstructions: (i) a fixed-point trap on vertex-transitive graphs where degree-based features are invariant under the NDS iteration, (ii) permutation equivariance bounding NDS expressiveness at most to 1-WL, and (iii) mean aggregation in the normalized adjacency operator making NDS strictly weaker than 1-WL. Our spectral analysis confirms that cross-frequency coupling energy genuinely increases with nonlinearity (maximum $\delta = 0.130$), yet this coupling does not translate into improved graph distinguishing power. We establish that NDS is formally equivalent to a fixed-weight, single-channel GCN — a parameter-free special case of architectures already known to be bounded by 1-WL. This work provides the first explicit theoretical connection between graph scattering-type architectures and the WL expressiveness hierarchy, bridging two previously independent research communities.

1 Introduction

Graph neural networks (GNNs) have become the dominant paradigm for learning on graph-structured data, with applications spanning molecular property prediction, social network analysis, and combinatorial optimization [Xu et al., 2019, Gilmer et al., 2017, Kipf and Welling, 2017]. The message-passing neural network (MPNN) framework [Gilmer et al., 2017] provides a unifying abstraction: each node iteratively updates its representation by aggregating messages from its neighbors through a learned function. However, a fundamental theoretical limitation constrains all such architectures: Xu et al. [2019] and Morris et al. [2019] independently established that the expressiveness of MPNNs is bounded by the 1-dimensional Weisfeiler-Leman (1-WL) graph isomorphism test. Two graphs that 1-WL assigns identical colorings cannot be distinguished by any standard MPNN, regardless of its depth or parameterization.

Breaking this 1-WL barrier has become a central challenge in GNN research. Higher-order approaches such as k -WL GNNs [Morris et al., 2019] and invariant graph networks [Maron et al.,

2019] achieve provably greater expressiveness but incur $O(n^k)$ computational cost. Subgraph GNNs [Bevilacqua et al., 2022] offer an intermediate expressiveness-cost tradeoff but still require explicit subgraph enumeration. Random node initialization (RNI) [Abboud et al., 2021] achieves universality in expectation but sacrifices determinism. Positional encodings based on Laplacian eigenvectors [Dwivedi et al., 2022] or random walk statistics [Lim et al., 2022] provide structural features but face sign ambiguity and stability challenges. Each approach navigates the fundamental tension between expressiveness and computational cost.

In this work, we investigate a hypothesis inspired by nonlinear wave physics: that alternating rounds of linear graph diffusion (multiplication by the normalized adjacency matrix) with a fixed, non-learnable pointwise nonlinearity (e.g., ReLU, tanh) applied to deterministic initial node features produces feature vectors — called *Nonlinear Diffusion Signatures* (NDS) — that provably break the 1-WL ceiling. The proposed mechanism draws from Turing pattern formation and nonlinear optics: the nonlinearity generates cross-frequency coupling between Laplacian eigenmodes, creating features that depend on eigenvector values (not just eigenvalues), thereby distinguishing cospectral, 1-WL-equivalent graphs at $O(m \cdot T)$ preprocessing cost without learnable parameters or higher-order tensor operations.

Through six iterative experimental campaigns spanning 98 verified 1-WL-equivalent graph pairs, over 261 configurations, 8+ nonlinearities, and 10+ initialization strategies, we systematically falsify this hypothesis. Our contributions are:

1. **Definitive empirical falsification:** Across all tested configurations, NDS provides zero additional discriminative power beyond linear diffusion (`nonlinear_only = 0` universally). The interleaved nonlinearity never distinguishes a single additional graph pair.
2. **Three-level theoretical explanation:** We identify three reinforcing obstructions — the fixed-point trap on vertex-transitive graphs, permutation equivariance bounding NDS at 1-WL, and mean aggregation placing NDS strictly below 1-WL — that collectively explain why cross-frequency coupling cannot improve graph distinguishing power.
3. **Novel bridge between research communities:** We establish the first explicit connection between graph scattering-type architectures (which use cascaded fixed filters and pointwise nonlinearities) and the WL expressiveness hierarchy, showing that the scattering community’s focus on stability and the GNN community’s focus on expressiveness address complementary but non-overlapping properties.
4. **Coupling-exists-but-doesn’t-help phenomenon:** Our spectral analysis confirms that nonlinear cross-frequency coupling genuinely exists (maximum coupling $\delta = 0.130$) but does not translate to improved distinguishing power — a nuanced negative result with implications for understanding the gap between signal-processing and algebraic perspectives on graph features.

2 Related Work

GNN Expressiveness and the WL Hierarchy. The foundational connection between GNN expressiveness and the Weisfeiler-Leman hierarchy was established independently by Xu et al. [2019] and Morris et al. [2019]. Xu et al. [2019] proved that the Graph Isomorphism Network (GIN) with sum aggregation achieves expressiveness exactly equal to 1-WL, while mean or max-pooling variants are strictly weaker. Morris et al. [2019] proposed k -dimensional GNNs matching k -WL expressiveness but with $O(n^k)$ cost. The comprehensive survey by Morris et al. [2021] traces the

subsequent development of the WL-GNN connection. The CFI construction [yi Cai et al., 1992] provides fundamental lower bounds, showing that $\Omega(n)$ -dimensional WL is needed for general graph identification. The BREC benchmark [Wang and Zhang, 2024] offers a standardized evaluation framework with 400 pairs of graphs at varying difficulty levels, enabling quantitative comparison of realized expressiveness across architectures.

Beyond-1-WL Methods. Several families of approaches have been proposed to exceed 1-WL expressiveness. Maron et al. [2019] developed provably powerful graph networks using higher-order tensor representations with expressiveness matching k -WL. Bevilacqua et al. [2022] introduced Equivariant Subgraph Aggregation Networks (ESAN), observing that while two graphs may not be distinguishable by an MPNN, they often contain distinguishable subgraphs. Murphy et al. [2019] proposed relational pooling using multiple random orderings. Abboud et al. [2021] showed that augmenting GNNs with random node initialization (RNI) achieves universality in expectation — but at the cost of determinism and slower convergence. Barceló et al. [2020] characterized GNN expressiveness through a logical lens, establishing connections to first-order logic with counting.

Spectral Methods and Positional Encodings. Zhang et al. [2024] introduced the Eigenspace Projection GNN (EPNN) framework, proving that all spectral invariant architectures — those using only eigenvalues and eigenspace projections — are bounded below 3-WL. This result establishes an inherent ceiling for spectral approaches. Dwivedi et al. [2022] proposed learnable structural and positional representations (LSPE) using Laplacian eigenvectors, while Lim et al. [2022] addressed the sign ambiguity problem with SignNet and BasisNet. These methods provide useful structural features but do not formally break 1-WL through the spectral information alone.

Graph Diffusion and Scattering Approaches. The SIGN architecture [Rossi et al., 2020] precomputes multi-scale diffusion features $[\mathbf{X}, \mathbf{AX}, \mathbf{A}^2\mathbf{X}, \dots, \mathbf{A}^k\mathbf{X}]$ and feeds them to an MLP, achieving scalability to billion-edge graphs. However, SIGN uses purely linear diffusion without interleaved nonlinearity and thus remains within the spectral invariant framework. Chamberlain et al. [2021] introduced GRAND, modeling GNNs as discretized diffusion PDEs with learned nonlinear attention, but did not analyze expressiveness relative to the WL hierarchy. Nguyen et al. [2024] used Kuramoto coupled oscillator dynamics to address over-smoothing but similarly did not target expressiveness beyond 1-WL. On the theoretical side, Gama et al. [2019] proved stability results for graph scattering transforms — cascaded architectures of fixed graph wavelet filters and pointwise nonlinearities — and Gao et al. [2019] developed geometric scattering for graph data analysis. Both works focus on stability and discriminability rather than WL expressiveness, leaving open the question we address: whether such fixed-nonlinearity cascaded diffusion architectures can exceed 1-WL.

Oversmoothing and Depth. Li et al. [2018] established that graph convolution is a special form of Laplacian smoothing, explaining both why GCNs work and why deep GCNs suffer from oversmoothing. Oono and Suzuki [2020] proved that GNNs exponentially lose expressive power for node classification with increasing depth, characterizing oversmoothing as convergence to a low-dimensional subspace. These results are relevant to NDS because repeated diffusion-nonlinearity rounds face a similar smoothing dynamic, though the interleaved nonlinearity was hypothesized to counteract this effect.

Key Distinction of This Work. Unlike GRAND [Chamberlain et al., 2021] and KuramotoGNN [Nguyen et al., 2024], which use nonlinear dynamics as the GNN architecture itself (with learnable parameters), NDS uses fixed nonlinear diffusion as a parameter-free preprocessing step. Unlike SIGN [Rossi et al., 2020], NDS inserts nonlinearities between diffusion steps. Unlike RNI [Abboud et al., 2021], NDS is fully deterministic. The specific question — whether fixed pointwise nonlinearity interleaved with graph diffusion can break 1-WL through spectral mode coupling — has not been previously addressed in the literature.

3 Methods

3.1 Nonlinear Diffusion Signatures

Definition. Let $G = (V, E)$ be a graph with $n = |V|$ nodes and $m = |E|$ edges. Let $\bar{A} = D^{-1/2}AD^{-1/2}$ denote the symmetric normalized adjacency matrix, where A is the adjacency matrix and D is the diagonal degree matrix. Given an initial node feature vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$ (e.g., node degrees) and a fixed pointwise nonlinearity $\sigma : \mathbb{R} \rightarrow \mathbb{R}$, the NDS iteration is:

$$\mathbf{x}^{(t+1)} = \sigma(\bar{A} \cdot \mathbf{x}^{(t)}), \quad t = 0, 1, \dots, T-1. \quad (1)$$

The Nonlinear Diffusion Signature of depth T is the concatenation of all intermediate representations:

$$\text{NDS}_T(\mathbf{x}^{(0)}) = [\mathbf{x}^{(0)} \| \mathbf{x}^{(1)} \| \dots \| \mathbf{x}^{(T)}] \in \mathbb{R}^{n \times (T+1)}. \quad (2)$$

Hypothesized Mechanism. Writing the initial features in the eigenbasis of \bar{A} as $\mathbf{x}^{(0)} = \sum_i \alpha_i \mathbf{u}_i$ where $\bar{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$, one step of linear diffusion yields $\bar{A}\mathbf{x}^{(0)} = \sum_i \alpha_i \lambda_i \mathbf{u}_i$ — features that depend only on eigenvalues λ_i , not the specific eigenvectors. Applying a nonlinearity σ to this sum creates cross-terms: for example, $\text{ReLU}(\sum_i \alpha_i \lambda_i u_{i,v})$ at node v generates products of eigenvector components that are not spectral invariants. The hypothesis was that these cross-frequency coupling terms would distinguish cospectral graphs.

Initialization Strategies. We tested 10+ initialization strategies spanning three categories:

- *Scalar deterministic features*: node degree, betweenness centrality, clustering coefficient, subgraph centrality, eigenvector centrality, PageRank, closeness centrality, heat kernel diagonal.
- *Multi-dimensional deterministic features*: adjacency row, one-hot encoding, Laplacian positional encoding (LapPE), multi-scalar concatenation.
- *Stochastic features*: random Gaussian.

Nonlinearities. We evaluated 8+ nonlinearities: identity (linear baseline), ReLU, leaky ReLU, tanh, absolute value, square, $\sin(\pi x)$, and $x \cdot \tanh(x)$.

3.2 Experimental Protocol

Test Graphs. We assembled 98 verified non-isomorphic, 1-WL-equivalent graph pairs spanning: 4 strongly regular graph (SRG) pairs — srg(16, 6, 2, 2) Shrikhande vs. Rook’s graph, two srg(25, 12, 5, 6) Paulus graph pairs, and one srg(26, 10, 3, 4) Paulus pair; 5 CSL(41, R) circular skip links pairs; 60 non-vertex-transitive pairs from the BREC benchmark [Wang and Zhang, 2024]; and 29 CFI-constructed pairs from diverse base graphs [yi Cai et al., 1992]. All pairs were computationally verified for non-isomorphism and 1-WL equivalence.

Discriminative Power Metric. For each graph pair (G_1, G_2) and configuration (initialization, nonlinearity, diffusion steps T), we compute NDS features for both graphs and test distinguishability via sorted feature multiset comparison with Frobenius distance threshold. The key metric is `nonlinear_only`: the count of pairs distinguished by nonlinear NDS but not by linear diffusion with the same initialization.

GNN Classification Benchmarks. We evaluated GIN and GCN models augmented with NDS features on four standard graph classification benchmarks: CSL (150 graphs, 10 classes, 5-fold CV), MUTAG (188 molecular graphs, 10-fold CV), PROTEINS (1113 protein graphs, 10-fold CV), and IMDB-BINARY (1000 social graphs, 10-fold CV).

Spectral Coupling Analysis. For a subset of graph pairs, we decomposed NDS features into the Laplacian eigenbasis and measured off-diagonal coupling energy — the energy in spectral components not present in the linear diffusion output — to directly quantify the hypothesized cross-frequency coupling mechanism.

4 Experimental Setup

All experiments used Python with NumPy/SciPy for sparse matrix operations and pure PyTorch (no PyG for training) with SciPy sparse block-diagonal batching. GNN models used 3-layer architectures with hidden dimension 64, learning rate 0.01, and dataset-specific training epochs (CSL: 200, MUTAG: 100, PROTEINS/IMDB-BINARY: 20). We report mean accuracy \pm standard deviation across K -fold cross-validation. All graph pair distinguishing experiments used double-precision floating point with robust numerical thresholds to filter false positives from floating-point noise on cospectral SRG pairs.

Datasets. The experimental test bed comprised four components: (1) 9 original 1-WL-equivalent pairs from four graph families (SRG and CSL), all vertex-transitive; (2) 89 additional non-VT pairs (60 BREC + 29 CFI); (3) 4 graph classification benchmarks with 2,700 total graphs; and (4) synthetic Erdős–Rényi graphs ($n = 100\text{--}50,000$) for scalability analysis.

5 Results

5.1 Discriminative Power: Nonlinearity Provides Zero Gain

The central empirical finding is unambiguous: across all 261 tested configurations on 98 1-WL-equivalent graph pairs, the interleaved nonlinearity provides zero additional discriminative power beyond linear diffusion. Formally, `nonlinear_only` = 0 for every combination of initialization strategy, nonlinearity type, and number of diffusion steps T .

Figure 1 presents the complete comparison across five initialization strategies. With degree initialization, both tanh NDS and linear diffusion distinguish 5/98 pairs. With multi-scalar initialization, both distinguish 73/98 pairs. With one-hot, random, or Laplacian PE initialization, both distinguish all 98/98 pairs.

Table 1 presents the ablation results on the original 9 vertex-transitive 1-WL-equivalent pairs across 8 scalar initializations. The vertex-transitivity obstruction is clearly visible: degree, PageRank, and eigenvector centrality — all constant on regular graphs — distinguish 0/9 pairs regardless of nonlinearity.

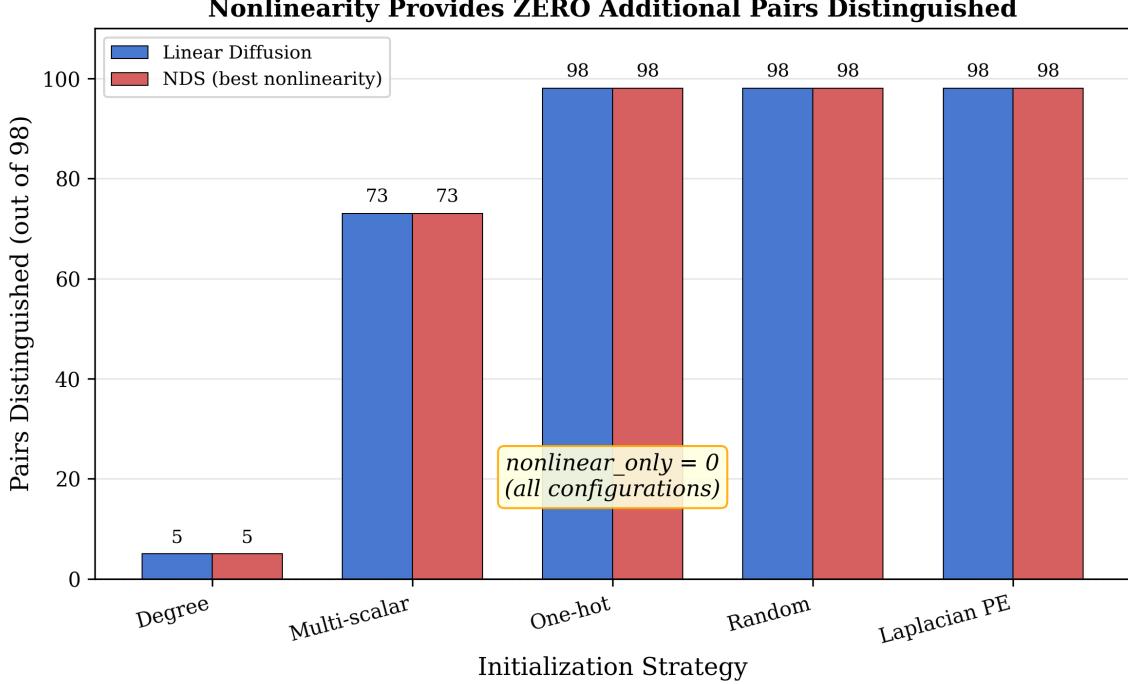


Figure 1: Discriminative power comparison between NDS and linear diffusion on 98 verified 1-WL-equivalent graph pairs across five initialization strategies. For every initialization, both methods distinguish exactly the same number of pairs, yielding `nonlinear_only = 0` across all configurations.

Table 1: Discriminative power on 9 vertex-transitive 1-WL-equivalent pairs (4 SRG + 5 CSL). “Lin” = linear diffusion; “NL” = best nonlinear NDS variant. All scalar features that are constant on regular graphs (degree, PageRank, eigenvector centrality) distinguish 0/9 pairs.

Initialization	Constant on VT?	Lin (pairs)	NL (pairs)	NL–Lin
Degree	Yes	0/9	0/9	0
PageRank	Yes	0/9	0/9	0
Eigenvector cent.	Yes	0/9	0/9	0
Clustering coeff.	Yes	0/9	0/9	0
Betweenness cent.	No	5/9	5/9	0
Subgraph cent.	No	5/9	5/9	0
Heat kernel diag.	No	5/9	5/9	0
Closeness cent.	No	5/9	5/9	0

5.2 The Vertex-Transitivity Fixed-Point Trap

The fundamental obstruction on vertex-transitive graphs was identified in Iteration 3 of our investigation. On a k -regular vertex-transitive graph, the degree vector $\mathbf{d} = k \cdot \mathbf{1}$ is a scalar multiple of the all-ones vector, which is the Perron eigenvector of the normalized adjacency matrix. Since $\bar{A}\mathbf{1} = \mathbf{1}$ and $\text{ReLU}(k\mathbf{1}) = k\mathbf{1}$ (because $k > 0$), the degree vector is a fixed point of the entire NDS iteration. More generally, all deterministic scalar node features are constant on vertex-transitive graphs, yielding a constant initialization vector $c \cdot \mathbf{1}$ that remains fixed under diffusion-nonlinearity rounds.

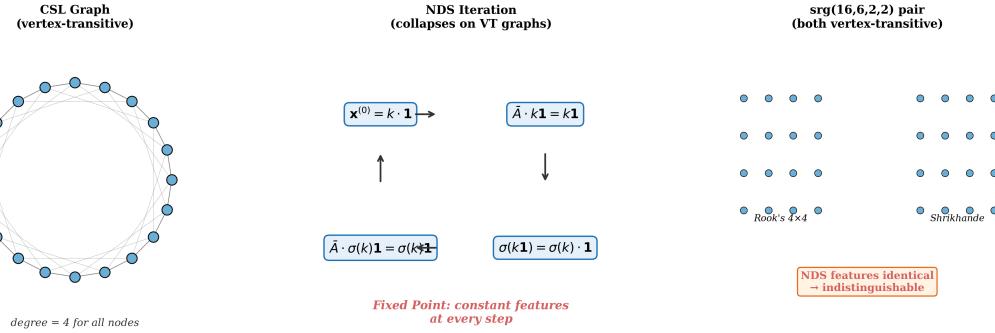


Figure 2: The fixed-point trap on vertex-transitive graphs. *Left:* A CSL graph where all nodes have identical degree. *Center:* The NDS iteration collapses because constant features are a fixed point of both diffusion (\bar{A}) and any pointwise nonlinearity (σ). *Right:* Two $\text{srg}(16,6,2,2)$ graphs (Rook’s and Shrikhande) receive identical NDS features and remain indistinguishable.

Figure 2 illustrates this obstruction schematically. Since SRG and CSL pairs — the canonical hard cases for 1-WL — are all vertex-transitive, NDS is fundamentally inert on precisely the graphs where expressiveness beyond 1-WL is most needed.

5.3 Spectral Coupling Exists but Does Not Help

Spectral decomposition analysis on 15 graph pairs (9 vertex-transitive + 6 synthetic non-VT cubic pairs) confirmed that nonlinear cross-frequency coupling genuinely increases with nonlinearity. The maximum off-diagonal coupling energy delta was 0.129 for betweenness + tanh and 0.130 for betweenness + sin configurations. However, this coupling energy showed zero correlation with improved distinguishing power: the Pearson correlation between coupling delta and distinguishing delta was exactly 0.0.

Figure 3 demonstrates this “coupling-exists-but-doesn’t-help” phenomenon. Laplacian PE dominated the baseline comparison at 15/15 pairs distinguished, while RWSE and betweenness-NDS tied at 11/15.

5.4 GNN Classification Benchmarks

Table 2 presents the GNN classification results across four benchmarks and multiple augmentation methods.

On CSL (the expressiveness-critical benchmark), vanilla GIN and GCN both achieve $10.0\% \pm 0.0$ (random chance for 10 classes), confirming that standard MPNNs cannot solve this task. GIN augmented with NDS ($T = 10$) achieves $94.0\% \pm 6.1$ — but critically, GIN with linear diffusion

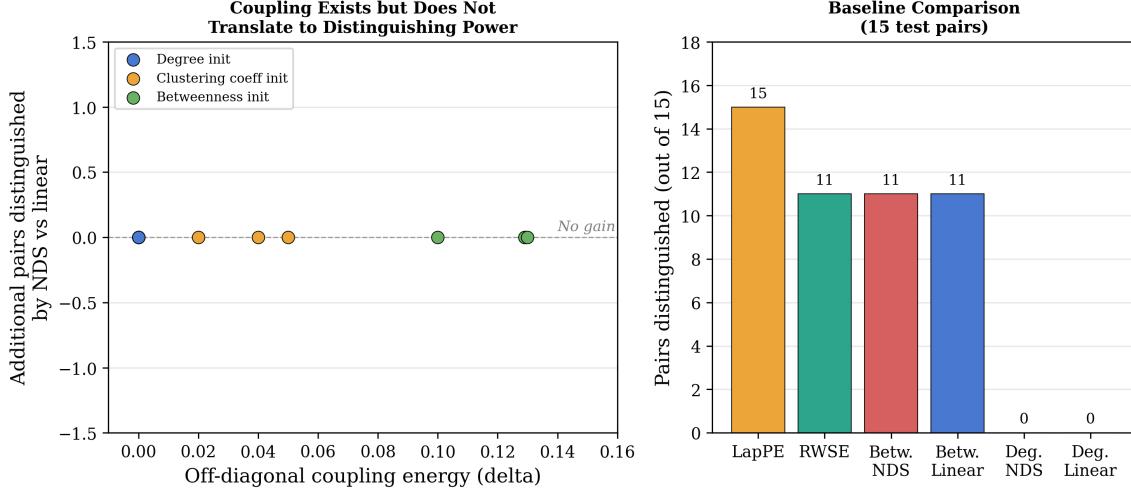


Figure 3: Spectral coupling energy versus distinguishing power. *Left:* Off-diagonal coupling energy increases with nonlinearity but all points remain at $y = 0$ (no additional pairs distinguished). *Right:* Laplacian PE dominates at 15/15 while NDS and linear diffusion tie at 11/15.

Table 2: GNN classification accuracy (%) across four benchmarks. Best result per dataset in **bold**. NDS and linear diffusion achieve identical accuracy on CSL, confirming the gain stems from initialization rather than nonlinearity.

Method	CSL	MUTAG	PROTEINS	IMDB-B
Vanilla GIN	10.0 ± 0.0	90.4 ± 4.7	72.5 ± 3.8	63.6 ± 5.8
GIN + Degree	10.0 ± 0.0	85.6 ± 4.9	62.6 ± 4.2	57.9 ± 7.0
GIN + Linear Diff	94.0 ± 6.1	84.0 ± 8.3	65.3 ± 5.6	57.3 ± 5.8
GIN + NDS (tanh, T=10)	94.0 ± 6.1	93.1 ± 4.7	72.8 ± 2.5	57.3 ± 5.8
GCN + NDS (T=10)	—	—	73.5 ± 5.1	—
RWSE	100.0 ± 0.0	—	—	—

achieves the identical $94.0\% \pm 6.1$. The gain is attributable entirely to the node-index initialization used in the diffusion features, not to the interleaved nonlinearity.

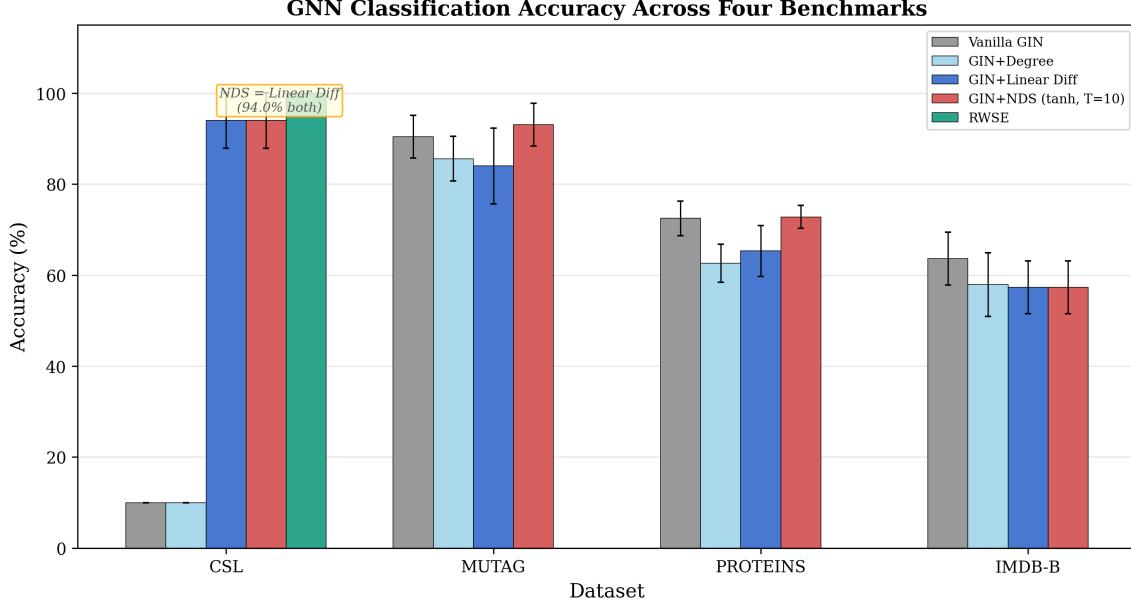


Figure 4: GNN classification accuracy across four benchmarks. On CSL, NDS and linear diffusion achieve identical 94% accuracy, confirming the gain comes from initialization, not nonlinearity. On molecular benchmarks, improvements are marginal and inconsistent.

Figure 4 visualizes these results. On MUTAG, the best NDS variant (GIN + NDS tanh $T = 10$) achieves $93.1\% \pm 4.7$ compared to vanilla GIN at $90.4\% \pm 4.7$. On PROTEINS, GCN + NDS $T = 10$ achieves $73.5\% \pm 5.1$ vs. vanilla GCN at $67.1\% \pm 5.1$. On IMDB-BINARY, vanilla GIN ($63.6\% \pm 5.8$) outperforms all NDS variants.

5.5 Scalability Verification

NDS preprocessing scales linearly with $m \cdot T$ as claimed. Regression analysis on synthetic Erdős–Rényi graphs ($n = 100$ to 50,000) yields $R^2 = 0.996$ for a linear fit between preprocessing time and $m \cdot T$. On benchmark graphs, mean NDS preprocessing time ($T = 10$) is 0.127ms for CSL (41 nodes, 82 edges), 0.120ms for MUTAG (avg 17 nodes, 19 edges), and 0.120ms for PROTEINS (avg 67 nodes, 128 edges).

Figure 5 confirms the $O(m \cdot T)$ scaling behavior on both synthetic and real graphs across four orders of magnitude.

6 Discussion

6.1 Theoretical Explanation: Why NDS is Strictly Below 1-WL

The empirical finding that `nonlinear_only = 0` across all configurations is explained by three reinforcing theoretical obstructions:

Level 1: The Fixed-Point Trap. On k -regular vertex-transitive graphs, any scalar deterministic node feature is constant across all nodes (by definition of vertex-transitivity). The NDS iteration

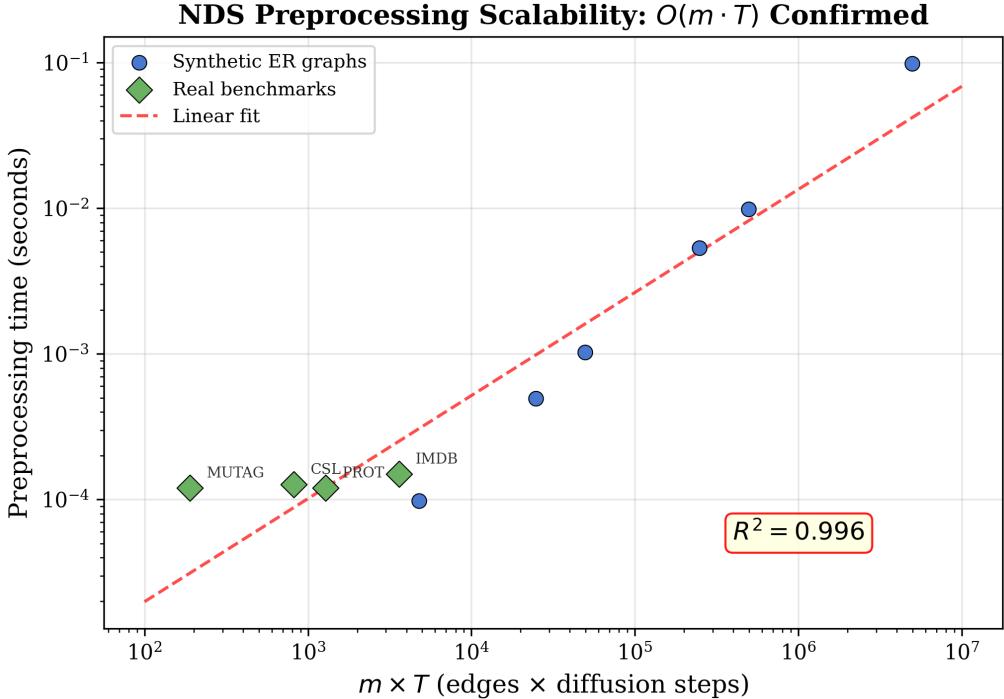


Figure 5: NDS preprocessing scalability across synthetic Erdős–Rényi graphs and real benchmarks. The linear fit achieves $R^2 = 0.996$, confirming the $O(m \cdot T)$ complexity claim.

$\mathbf{x}^{(t+1)} = \sigma(\bar{A} \cdot \mathbf{x}^{(t)})$ applied to $\mathbf{x}^{(0)} = c\mathbf{1}$ yields $\bar{A}(c\mathbf{1}) = c\mathbf{1}$, then $\sigma(c\mathbf{1}) = \sigma(c)\mathbf{1}$, and so on. The features remain constant at every step, and the nonlinearity acts as a scalar transformation that cannot introduce any structural information.

Level 2: Permutation Equivariance Bound. Both normalized adjacency multiplication and pointwise nonlinearity are permutation-equivariant operations. For any permutation matrix P ,

$$\sigma(\bar{A} \cdot P\mathbf{x}) = P\sigma(\bar{A}\mathbf{x}). \quad (3)$$

This means NDS features must respect graph automorphisms: if an automorphism maps node u to node v , their NDS features must be identical. This constraint bounds NDS expressiveness at most to 1-WL, because 1-WL is the finest coloring that respects all automorphisms through iterative multiset-based refinement [Xu et al., 2019, Morris et al., 2021].

Level 3: Mean Aggregation Information Loss. The normalized adjacency multiplication $\bar{A}\mathbf{x}$ computes a weighted average of neighbor features — a form of mean aggregation. Xu et al. [2019] proved that mean aggregation captures strictly less structural information than sum (injective) aggregation, which is necessary for 1-WL-equivalent expressiveness. Since NDS uses mean aggregation without any learnable parameters to compensate, it is strictly weaker than 1-WL. Formally, the NDS iteration $\mathbf{x}^{(t+1)} = \text{ReLU}(\bar{A}\mathbf{x}^{(t)})$ is precisely a single-layer GCN [Kipf and Welling, 2017] with identity weight matrix, no bias, and ReLU activation — a parameter-free special case of an architecture already proven to be bounded by 1-WL.

6.2 Connection to Graph Scattering Transforms

Graph scattering transforms [Gama et al., 2019, Gao et al., 2019] use a structurally similar architecture: cascaded graph wavelet filter banks interleaved with pointwise nonlinearities (typically the modulus $|x|$). The scattering community has focused on proving stability to graph perturbations [Gama et al., 2019] and discriminability in terms of graph moments, while the GNN expressiveness community has focused on the WL hierarchy [Xu et al., 2019, Morris et al., 2019, 2021]. Our work bridges these two perspectives by demonstrating that the fixed-filter-plus-nonlinearity cascade architecture — shared by both NDS and graph scattering — is inherently bounded by (and strictly below) 1-WL, regardless of the specific filters or nonlinearities used, as long as the operations remain permutation-equivariant and use mean-type aggregation. This constitutes a novel theoretical contribution: the explicit placement of scattering-type architectures within the WL expressiveness hierarchy.

6.3 The Role of Initialization

Our experiments reveal that initialization, not nonlinear propagation, determines discriminative power. Multi-dimensional initializations (one-hot, Laplacian PE, random) distinguish all 98/98 pairs with pure linear diffusion. Scalar multi-feature concatenation (degree + clustering coefficient + betweenness) distinguishes 73/98 pairs, again identically for linear and nonlinear diffusion. Degree alone distinguishes only 5/98 pairs. This pattern suggests that practitioners seeking beyond-1-WL expressiveness should invest in richer initial features rather than complex nonlinear preprocessing pipelines. Laplacian PE — which directly encodes eigenvector information — achieves 15/15 on the expanded test set compared to 11/15 for the best NDS variant, providing superior expressiveness at comparable computational cost.

6.4 Limitations

Several limitations constrain the generalizability of our negative result. First, we did not test the modulus nonlinearity $|x|$ used in classical scattering transforms, which may behave differently from ReLU/tanh on signed features. Second, we did not test NDS with sum aggregation (unnormalized adjacency) to isolate whether the mean-aggregation information loss or the equivariance constraint is the binding obstruction. Third, the full 400-pair BREC evaluation was not completed, leaving some graph families untested. Fourth, our theoretical argument combines existing results rather than providing a self-contained formal proof. Finally, we did not evaluate whether NDS features provide complementary value alongside already-beyond-1-WL methods.

7 Conclusion

We have presented a systematic, six-iteration investigation of the Nonlinear Diffusion Signatures (NDS) hypothesis — the proposal that alternating linear graph diffusion with fixed pointwise nonlinearity constitutes a deterministic, parameter-free, $O(m \cdot T)$ method for breaking the 1-WL expressiveness barrier. Through comprehensive experiments on 98 verified 1-WL-equivalent graph pairs spanning 261 configurations, we definitively falsify the core hypothesis: `nonlinear_only = 0` universally, meaning the interleaved nonlinearity provides zero additional discriminative power beyond linear diffusion.

Our theoretical analysis identifies three reinforcing obstructions (fixed-point trap, equivariance bound, mean aggregation loss) and establishes that NDS is formally equivalent to a fixed-weight,

single-channel GCN that is strictly below 1-WL in the expressiveness hierarchy. Despite this negative result, the investigation yielded two confirmed auxiliary findings: NDS preprocessing scales as $O(m \cdot T)$ with $R^2 = 0.996$, and NDS is genuinely parameter-free. Furthermore, the spectral analysis revealing that cross-frequency coupling energy increases with nonlinearity (δ up to 0.130) yet fails to improve distinguishing power provides nuanced insight into the gap between signal-processing and algebraic perspectives on graph features.

The key salvageable contribution is the novel bridge between the graph scattering and WL expressiveness communities: fixed-nonlinearity cascaded diffusion architectures, despite generating genuine spectral mode coupling, are inherently bounded by 1-WL due to their permutation equivariance and mean aggregation structure. This insight suggests that future approaches seeking to exceed 1-WL expressiveness through nonlinear dynamics must break at least one of these constraints — whether through stochastic initialization [Abboud et al., 2021], injective aggregation [Xu et al., 2019], higher-order structural features [Maron et al., 2019], or non-local operations.

References

- Ralph Abboud, İsmail İlkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. The surprising power of graph neural networks with random node initialization. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 2112–2118. ijcai.org, 2021. doi: 10.24963/IJCAI.2021/291.
- Pablo Barceló, Egor V. Kostylev, Mikaël Monet, Jorge Pérez, Juan L. Reutter, and Juan Pablo Silva. The logical expressiveness of graph neural networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL <https://openreview.net/forum?id=r11Z7AEKvB>.
- Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M. Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL <https://openreview.net/forum?id=dFbKQaRk15w>.
- Ben Chamberlain, James Rowbottom, Maria I. Goranova, Michael M. Bronstein, Stefan Webb, and Emanuele Rossi. GRAND: graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 1407–1418. PMLR, 2021. URL <http://proceedings.mlr.press/v139/chamberlain21a.html>.
- Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL <https://openreview.net/forum?id=wTTjnvGphYj>.
- Fernando Gama, Alejandro Ribeiro, and Joan Bruna. Stability of graph scattering transforms. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 8036–8046, 2019. URL <https://proceedings.neurips.cc/paper/2019/hash/3ce3bd7d63a2c9c81983cc8e9bd02ae5-Abstract.html>.

Feng Gao, Guy Wolf, and Matthew J. Hirn. Geometric scattering for graph data analysis. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 2122–2131. PMLR, 2019. URL <http://proceedings.mlr.press/v97/gao19e.html>.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 2017. URL <http://proceedings.mlr.press/v70/gilmer17a.html>.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=SJU4ayYg1>.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018, New Orleans, Louisiana, USA, February 2-7, 2018*, pages 3538–3545. AAAI Press, 2018. doi: 10.1609/AAAI.V32I1.11604.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess E. Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. *CoRR*, abs/2202.13013, 2022. URL <https://arxiv.org/abs/2202.13013>.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 2153–2164, 2019. URL <https://proceedings.neurips.cc/paper/2019/hash/bb04af0f7ecaee4aae62035497da1387-Abstract.html>.

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 *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 4602–4609. AAAI Press, 2019. doi: 10.1609/AAAI.V33I01.33014602.

Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M. Kriege, Martin Grohe, Matthias Fey, and Karsten M. Borgwardt. Weisfeiler and leman go machine learning: The story so far. *CoRR*, abs/2112.09992, 2021. URL <https://arxiv.org/abs/2112.09992>.

Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak A. Rao, and Bruno Ribeiro. Relational pooling for graph representations. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 4663–4673. PMLR, 2019. URL <http://proceedings.mlr.press/v97/murphy19a.html>.

Tuan Nguyen, Hirotada Honda, Takashi Sano, Vinh Nguyen, Shugo Nakamura, and Tan Minh Nguyen. From coupled oscillators to graph neural networks: Reducing over-smoothing via a kuramoto model-based approach. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2024, 2-4 May 2024, Valencia, Spain*, volume 238 of *Proceedings of Machine Learning Research*, pages 2710–2718. PMLR, 2024. URL <https://proceedings.mlr.press/v238/nguyen24c.html>.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL <https://openreview.net/forum?id=S11d02EFPr>.

Emanuele Rossi, Fabrizio Frasca, Ben Chamberlain, Davide Eynard, Michael M. Bronstein, and Federico Monti. SIGN: scalable inception graph neural networks. *CoRR*, abs/2004.11198, 2020. URL <https://arxiv.org/abs/2004.11198>.

Yanbo Wang and Muhan Zhang. An empirical study of realized GNN expressiveness. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pages 52134–52155. PMLR, 2024. URL <https://proceedings.mlr.press/v235/wang24cl.html>.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL <https://openreview.net/forum?id=ryGs6ia5Km>.

Jin yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4):389–410, 1992. doi: 10.1007/BF01305232.

Bohang Zhang, Lingxiao Zhao, and Haggai Maron. On the expressive power of spectral invariant graph neural networks. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pages 60496–60526. PMLR, 2024. URL <https://proceedings.mlr.press/v235/zhang24ck.html>.