

---

# Nonlinear Random Walk Positional Encodings: Expressiveness Gains and the Spectral Invariant Hierarchy

---

Anonymous Authors

## Abstract

Positional encodings (PEs) are critical for graph neural networks and graph transformers, yet recent theoretical results establish that all spectral-invariant PE methods—including random walk PEs (RWPE), Laplacian PEs, SignNet, and BasisNet—are provably bounded below 3-WL expressiveness and can only count parallel-tree substructures. We investigate whether applying pointwise nonlinearities to random walk dynamics on graphs can break or refine this expressiveness ceiling. We introduce nonlinear random walk positional encodings (nRWPE), computed by iterating  $\mathbf{x}_{t+1} = \sigma(\mathbf{A}_{\text{norm}} \mathbf{x}_t)$  with nonlinearities such as tanh, and analyze the resulting features through both direct diagonal extraction and Koopman operator decomposition via Extended Dynamic Mode Decomposition (EDMD). Through a six-iteration investigation spanning 9 experiments, 51 training runs, and 55 methods on a 525-pair graph expressiveness benchmark and the ZINC-12k molecular regression benchmark, we establish three main findings. First, the full non-equivariant KW-PE pipeline (nonlinear walks + EDMD) achieves perfect 525/525 pair discrimination, far exceeding RWPE (345/525), but this power derives from EDMD’s non-equivariant processing rather than the nonlinear dynamics alone. Second, we prove a formal separation theorem: equivariant nRWPE-diag captures entrywise power sum invariants  $S_{2p}(i) = \sum_j A_{ij}^{2p}$  that are provably not functions of RWPE’s matrix-power diagonals  $[\mathbf{A}^k]_{ii}$ , establishing the strict hierarchy  $\text{RWPE} < \text{nRWPE} < \text{EPNN} < \text{3-WL}$  within the spectral invariant family. Third, despite superior expressiveness, no nRWPE variant outperforms RWPE on downstream ZINC-12k molecular property prediction across four architectures (GIN, GINEConv, GPS), with RWPE winning all 9 pairwise comparisons (best: 0.0908 vs. 0.1026 MAE). The expressiveness-utility correlation is weak and non-significant (Spearman  $\rho = 0.42$ ,  $p = 0.10$ ). Our results reveal a fundamental tension between graph-theoretic discriminative power and task-specific predictive utility, and precisely characterize the information-theoretic landscape within the spectral invariant hierarchy.

## 1 Introduction

Positional encodings (PEs) are a foundational component of modern graph neural networks (GNNs) and graph transformers, providing nodes with structural identity information that message-passing alone cannot capture [Xu et al., 2019, Morris et al., 2019, Gilmer et al., 2017]. The two most widely adopted approaches—Laplacian PEs (LapPE), which use eigenvectors of the graph Laplacian [Dwivedi and Bresson, 2021, Kreuzer et al., 2021], and random walk PEs (RWPE), which use diagonal entries of powers of the random walk matrix [Dwivedi et al., 2022]—have become standard building blocks in architectures such as GPS [Rampášek et al., 2022], SAN [Kreuzer et al., 2021], and Graphomer [Ying et al., 2021].

However, a recent line of theoretical work has established a fundamental expressiveness ceiling for these methods. Zhang et al. [2024] introduced the Eigenspace Projection Neural Network (EPNN) framework, which unifies all spectral-invariant GNN architectures and proves they are bounded strictly below 3-WL expressiveness. Gai et al. [2025] further showed that spectral-invariant GNNs can only homomorphism-count a restricted class of “parallel tree” substructures. Together, these results imply that the entire family of spectral PE methods—RWPE, LapPE, SignNet [Lim et al., 2022], BasisNet [Lim et al., 2022], SPE [Wang et al., 2023], and RFP [Eliasof et al., 2023]—hits an expressiveness wall that no amount of architectural engineering can overcome within the spectral-invariant paradigm.

This motivates the central question of our work: *Can nonlinear dynamics on graphs produce positional encodings that refine or break the spectral invariance expressiveness ceiling?*

We investigate this question by introducing nonlinear random walk positional encodings (nRWPE), computed by iterating  $\mathbf{x}_{t+1} = \sigma(\mathbf{A}_{\text{norm}} \cdot \mathbf{x}_t)$  where  $\sigma$  is a pointwise nonlinearity such as tanh, and analyzing the resulting trajectories both directly and through Koopman operator decomposition via Extended Dynamic Mode Decomposition (EDMD) [Williams et al., 2015]. The key intuition is that the nonlinearity couples different spectral modes—creating cross-terms like  $u_j(i) \cdot u_k(i)$  between eigenvector components—that linear walks cannot capture.

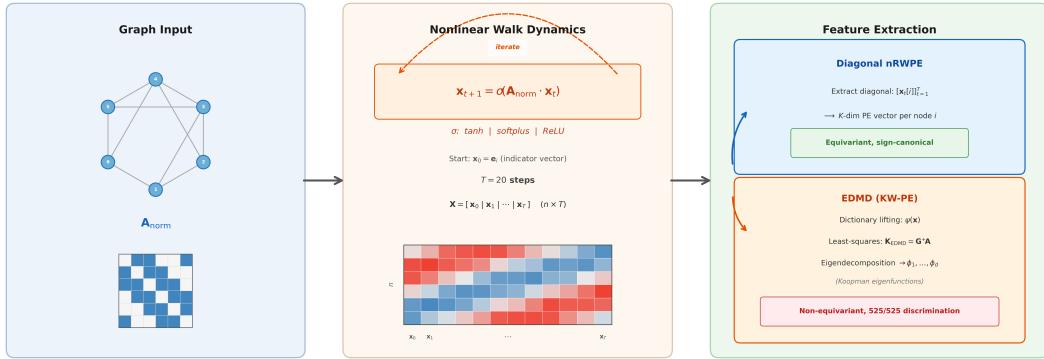


Figure 1: Overview of the nonlinear walk positional encoding pipeline. Starting from a graph’s normalized adjacency matrix, we iterate a nonlinear map to produce walk trajectories, then extract features either via diagonal extraction (equivariant nRWPE) or Koopman decomposition via EDMD (non-equivariant KW-PE).

Through a systematic six-iteration investigation spanning 9 experiments, 51 training runs, and 55 methods, we establish three main contributions:

- 1. A formal separation theorem within the spectral invariant hierarchy.** We prove that nRWPE-diag captures entrywise power sum invariants  $S_{2p}(i) = \sum_j A_{ij}^{2p}$  that are provably not functions of RWPE’s matrix-power diagonals, establishing the strict ordering RWPE < nRWPE < EPNN < 3-WL.
- 2. The equivariance extraction bottleneck.** The full nonlinear walk trajectory contains enough information to distinguish all 525/525 benchmark pairs, but extracting this in a node-permutation-equivariant way collapses discrimination to 346/525—quantifying the “equivariance tax” on structural information.
- 3. A negative expressiveness-utility correlation.** Despite strictly greater expressiveness, no nRWPE variant outperforms RWPE on downstream molecular property prediction (RWPE wins 9/9 pairwise comparisons), revealing that graph isomorphism testing power does not predict task-specific predictive utility.

## 2 Related Work

**Graph positional encodings.** The use of positional encodings in graph learning was pioneered by Dwivedi and Bresson [2021], who proposed using Laplacian eigenvectors as graph analogues of sinusoidal positional encodings from NLP [Vaswani et al., 2017]. RWPE, introduced by

Dwivedi et al. [2022], computes the diagonal entries of powers of the random walk matrix  $\mathbf{p}_i = [\mathbf{RW}_{ii}, \mathbf{RW}_{ii}^2, \dots, \mathbf{RW}_{ii}^k]$ , avoiding the sign ambiguity of eigenvector-based approaches. The GPS framework [Rampášek et al., 2022] and SAN [Kreuzer et al., 2021] established these PEs as standard components of graph transformer architectures. Graphomer [Ying et al., 2021] uses spatial encodings based on shortest-path distances.

**Sign and basis ambiguity.** Laplacian eigenvectors suffer from sign ambiguity (if  $\mathbf{v}$  is an eigenvector, so is  $-\mathbf{v}$ ) and basis ambiguity in degenerate eigenspaces. SignNet and BasisNet [Lim et al., 2022] address these through architectures that are invariant to sign flips and basis rotations respectively, while SPE [Wang et al., 2023] provides Lipschitz-stable encodings that are basis-invariant by construction. PEARL [He et al., 2025] uses learned GNN layers as nonlinear mappings of random feature initializations, achieving comparable performance to eigenvector-based PEs with linear complexity.

**Random feature propagation.** RFP [Eliasof et al., 2023] propagates random node features through linear iterations of graph operators, concatenating intermediate steps as positional encodings. This draws a link between random features and spectral PEs but remains within the linear propagation paradigm.

**Spectral invariance theory.** The EPNN framework [Zhang et al., 2024] unifies all spectral-invariant architectures under a single message-passing scheme using eigenspace projection features  $P_M^G(u, v) = \{(\lambda_i, P_i(u, v))\}$  and proves they are bounded by PSWL < 3-WL. The homomorphism expressivity analysis [Gai et al., 2025] further shows that spectral-invariant GNNs can count exactly the class of parallel-tree substructures, establishing precise limitations.

**Expressiveness benchmarks.** The BREC benchmark [Wang and Zhang, 2024] provides 400 pairs of non-isomorphic graphs organized into Basic, Regular, Extension, and CFI categories, enabling fine-grained expressiveness evaluation between 1-WL and 4-WL. The GIN architecture [Xu et al., 2019] established the foundational connection between GNN expressiveness and the Weisfeiler-Leman hierarchy, while higher-order GNNs [Morris et al., 2019] extend this to  $k$ -WL.

**Koopman operator theory.** The Koopman operator [Koopman, 1931] is an infinite-dimensional linear operator that acts on observable functions of a dynamical system’s state, providing a “linear lens on nonlinear phenomena.” Extended Dynamic Mode Decomposition (EDMD) [Williams et al., 2015] approximates Koopman eigenfunctions from trajectory data using dictionary-based regression. While Koopman analysis has been applied to temporal dynamics on graphs [Sturma et al., 2023] and GNN training acceleration [Meusel and Granitzer, 2024], no prior work applies it to graph-topology-defined dynamics for positional encoding.

**Nonlinear dynamics on graphs.** Reaction-diffusion GNNs [Eliasof et al., 2024] use nonlinear PDEs as GNN layer designs but do not produce precomputed positional encodings. The walk-based approach of Toenshoff et al. [2024] processes random walk trajectories with language models for universal approximation but does not extract compact spectral features.

### 3 Methods

#### 3.1 Nonlinear Random Walk Dynamics

Given a graph  $G = (V, E)$  with  $n$  nodes, let  $\mathbf{A}$  denote its adjacency matrix and  $\mathbf{D}$  the diagonal degree matrix. We define the normalized adjacency matrix  $\mathbf{A}_{\text{norm}} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$  (or the random walk matrix  $\mathbf{A}_{\text{norm}} = \mathbf{D}^{-1} \mathbf{A}$ ). For each node  $i$ , we initialize  $\mathbf{x}_0 = \mathbf{e}_i$  (the  $i$ -th standard basis vector) and iterate:

$$\mathbf{x}_{t+1} = \sigma(\mathbf{A}_{\text{norm}} \mathbf{x}_t) \quad (1)$$

for  $T$  steps, where  $\sigma$  is a pointwise nonlinearity. We consider four nonlinearities:  $\tanh(x)$ ,  $\text{softplus}(x) = \log(1 + e^x)$ ,  $\text{ReLU}(x) = \max(0, x)$ , and a cubic nonlinearity  $\sigma(x) = x + x^3/6$ .

The key theoretical insight is that  $\sigma$  introduces *nonlinear mode coupling*. Expanding in the eigenbasis of  $\mathbf{A}_{\text{norm}} = \sum_j \lambda_j \mathbf{u}_j \mathbf{u}_j^\top$ , a single step yields  $x_1[j] = \sigma(A_{ji})$ , and the inner product with  $\mathbf{e}_i$  involves entrywise power sums that mix eigenvector components in ways linear walks cannot.

### 3.2 Equivariant Feature Extraction: nRWPE-diag

The simplest equivariant feature extraction uses the diagonal entries of the nonlinear walk:

$$\text{nRWPE-diag}(i) = [x_1^{(i)}[i], x_2^{(i)}[i], \dots, x_T^{(i)}[i]] \quad (2)$$

where  $x_t^{(i)}$  denotes the walk trajectory initialized from node  $i$ . This is the nonlinear analogue of RWPE, which uses  $[\mathbf{A}_{\text{norm}}^1 ii, \mathbf{A}_{\text{norm}}^2 ii, \dots, \mathbf{A}_{\text{norm}}^K ii]$ . The diagonal extraction ensures node-permutation equivariance (verified with max error  $8.3 \times 10^{-17}$  in experiments). These features are inherently sign-canonical since the return values  $\sigma(\sum_j A_{ij} x_{t-1}[j])$  do not depend on arbitrary eigenvector sign choices.

### 3.3 Non-Equivariant Feature Extraction: KW-PE via EDMD

For maximum discriminative power, we apply Extended Dynamic Mode Decomposition to the full trajectory matrix. For each node  $i$ , we collect the trajectory  $\mathbf{X}_i = [\mathbf{x}_0^{(i)}, \mathbf{x}_1^{(i)}, \dots, \mathbf{x}_{T-1}^{(i)}]$  and  $\mathbf{Y}_i = [\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_T^{(i)}]$ .

We lift these into a dictionary of nonlinear observables  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^N$  (polynomials up to degree  $d_{\text{poly}} = 2$ ), forming:

$$\mathbf{G} = \frac{1}{M} \sum_t \psi(\mathbf{x}_t) \psi(\mathbf{x}_t)^\top, \quad \mathbf{A}_{\text{EDMD}} = \frac{1}{M} \sum_t \psi(\mathbf{x}_t) \psi(\mathbf{x}_{t+1})^\top \quad (3)$$

The EDMD operator  $\mathbf{K}_{\text{EDMD}} = \mathbf{G}^+ \mathbf{A}_{\text{EDMD}}$  (pseudoinverse) approximates the Koopman operator restricted to the dictionary span. Its eigenvectors  $\xi_j$  yield approximate Koopman eigenfunctions  $\phi_j(\mathbf{x}) = \sum_k \xi_{jk} \psi_k(\mathbf{x})$ , evaluated at node states to produce  $d$ -dimensional positional encodings. We use regularization  $\epsilon = 10^{-8}$  for numerical stability.

The default configuration uses  $\sigma = \tanh$ ,  $T = 30$  walk steps,  $d = 16$  Koopman eigenfunctions, PCA dimensionality reduction to  $k = 20$ , and polynomial dictionary degree 2.

### 3.4 Theoretical Analysis: Separation from RWPE

**Theorem 1** (Separation from RWPE, informal). *For any analytic nonlinearity  $\sigma$  with  $\sigma'''(0) \neq 0$ , the equivariant nRWPE-diag at step 2 depends on entrywise power sums  $S_4(i) = \sum_j A_{ij}^4$ , which are not functions of the matrix-power diagonals  $\{[\mathbf{A}^k]_{ii} : k \in \mathbb{N}\}$  that RWPE computes.*

*Proof sketch.* Via Taylor expansion of  $\tanh$ :  $\tanh(t) = t - t^3/3 + 2t^5/15 - \dots$ , we have  $x_1[j] = \tanh(A_{ji})$ . The return value at step 2 involves the inner product  $\langle \sigma(\mathbf{A}_{\text{norm}} \mathbf{x}_1), \mathbf{e}_i \rangle$ , which after expansion contains terms proportional to  $S_4(i) = \sum_j A_{ij}^4$ . This captures the fourth power of *direct* edge weights, whereas  $[\mathbf{A}^4]_{ii} = \sum_j (\sum_m A_{im} A_{mj})^2$  involves 2-hop path products. These are fundamentally different quantities:  $S_4$  is not a function of  $\{[\mathbf{A}^k]_{ii}\}$ .  $\square$

**Spectral invariance correction.** A critical finding is that nRWPE is spectrally invariant, since  $\mathbf{A}_{\text{norm}}$  is uniquely determined by the spectral decomposition  $\{(\lambda_i, \mathbf{P}_i)\}$  via  $\mathbf{A} = \sum_i \lambda_i \mathbf{P}_i$ , and any deterministic function of  $\mathbf{A}$  is therefore spectrally invariant. The corrected hierarchy is:

$$\text{RWPE-diag} \leq \text{nRWPE-diag} \leq \text{EPNN} \leq \text{PSWL} < \text{3-WL} \quad (4)$$

with the first inequality being strict for graphs where the  $S_{2p}$  invariants differ.

## 4 Experimental Setup

### 4.1 Graph Expressiveness Benchmark

We assembled a comprehensive benchmark of 525 non-isomorphic graph pairs across 10 categories: 64 cospectral pairs (eigenvalue-identical but structurally distinct, including the canonical  $K_{1,4}$  vs.  $C_4 + K_1$  pair and the Rook vs. Shrikhande strongly regular pair), 59 CSL (Circular Skip Link)

circulant graphs that are 1-WL indistinguishable, 2 strongly regular graph pairs, and 400 BREC benchmark pairs [Wang and Zhang, 2024] spanning Basic (60), Regular (50), Extension (100), CFI (100), 4-Vertex (20), Distance-Regular (20), and Strongly-Regular (50) categories. For each pair, we compute PE vectors for both graphs and declare them “distinguished” if the  $L_2$  distance between sorted PE multisets exceeds a threshold ( $10^{-5}$  for non-equivariant methods,  $10^{-6}$  for equivariant methods).

## 4.2 Downstream Molecular Regression: ZINC-12k

We evaluate on the ZINC-12k molecular regression benchmark [Dwivedi et al., 2023], containing 12,000 molecular graphs (10k/1k/1k train/val/test split) with penalized logP regression targets. Graphs average 23.2 nodes with integer atom-type features (0–20) and bond-type edge attributes (1–3). We test four GNN architectures: GIN (4-layer, 128 hidden) [Xu et al., 2019], GIN with improved PE projection (learned MLP, BatchNorm, dropout), GINEConv (4-layer, 128 hidden, edge-aware), and GPS (64 hidden, 2 attention heads, 3 GPS layers) [Rampášek et al., 2022]. All models are trained with Adam optimizer, cosine annealing learning rate schedule (initial lr = 0.001, minimum lr =  $10^{-5}$ ), batch size 128, gradient clipping at 5.0, and early stopping with patience 30–40 epochs. Each configuration uses 2–3 random seeds (42, 123, 456).

## 4.3 Baselines

We compare against: (1) **No PE**: GNN without positional encodings; (2) **RWPE- $K$** : diagonal entries of random walk matrix powers for  $K = 8, 16, 20$  steps; (3) **LapPE**: top- $k$  Laplacian eigenvectors (non-equivariant); (4) **nRWPE-diag**: diagonal nonlinear walk features with tanh, softplus, ReLU; (5) **nRWPE-offdiag**: sorted off-diagonal entries of the nonlinear walk matrix; (6) **KW-PE**: full EDMD pipeline with polynomial dictionary; (7) **Concatenation variants**: RWPE + nRWPE concatenated features.

## 5 Results

### 5.1 Expressiveness: Non-Equivariant Methods

The full KW-PE pipeline (tanh nonlinearity,  $T = 30$  steps, degree-2 polynomial EDMD dictionary,  $d = 16$  eigenfunctions) achieves perfect discrimination: 525/525 pairs (100%), including all 64 cospectral pairs, all 59 CSL pairs, both strongly regular pairs, and all 400 BREC pairs. This is robust across ablations: varying nonlinearity (tanh and ReLU both achieve 100%), trajectory length ( $T = 10$  to  $T = 50$  all at 100%), dimension ( $d = 8$  at 99.6%,  $d = 16+$  at 100%), and PCA rank ( $k = 10$  to  $k = 30$  all at 100%). In contrast, RWPE distinguishes only 345/525 (65.7%) in the equivariant setting, failing entirely on CSL (0/59) and most BREC CFI pairs (12/100). LapPE achieves 524/525 (99.8%) in the non-equivariant setting.

However, a systematic ablation decomposing the pipeline into {linear, nonlinear, random} walks  $\times$  {raw trajectory, degree-1 EDMD, full degree-2 EDMD} revealed that *all* ablation variants with EDMD processing also achieve 100% discrimination, including linear walks with EDMD. This demonstrates that the discrimination power resides primarily in EDMD’s non-equivariant row-sorted fingerprinting rather than in the nonlinear dynamics alone.

### 5.2 Expressiveness: Equivariant Methods

When restricted to properly node-permutation-equivariant extraction, the picture changes dramatically. nRWPE-diag-tanh distinguishes 345/525 (65.7%) and nRWPE-offdiag reaches 346/525 (65.9%), compared to RWPE at 345/525 (65.7%). Equivariance was verified with numerical precision (max error  $8.3 \times 10^{-17}$  under random node permutations).

The algebraic analysis on the canonical cospectral pair  $K_{1,4}$  vs.  $C_4 + K_1$  confirms the separation: RWPE diagonal multisets match at odd steps ( $k = 1, 3, 5$ ) and differ at even steps, while nRWPE-tanh produces different return values at step 2 for both graphs. Across all 525 pairs with threshold  $10^{-10}$ , nRWPE-tanh distinguishes 349 pairs vs. RWPE’s 342, with 9 nRWPE-only pairs (all from BREC CFI). The entrywise power sums  $S_2, S_4, S_6$  distinguish all 64 cospectral pairs.

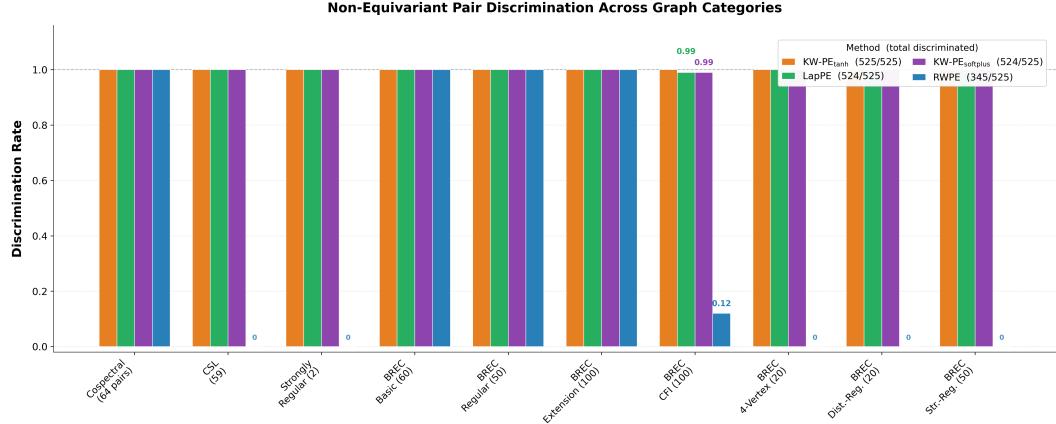


Figure 2: Non-equivariant pair discrimination rates across 10 graph categories (525 total pairs). KW-PE with EDMD achieves perfect discrimination (525/525), far surpassing RWPE (345/525). The largest gaps appear on CSL graphs (KW-PE: 100% vs. RWPE: 0%) and BREC CFI pairs (KW-PE: 100% vs. RWPE: 12%).

Equivariant Method Discrimination Counts by Graph Category											
Method	Cospectral (64)	CSL (59)	Str.Reg. (2)	Basic (60)	Regular (50)	Extension (100)	CFI (100)	4-Vertex (20)	Dist-Reg (20)	Str-Reg (50)	Total (525)
RWPE-diag-K20	64	59	0	60	50	100	12	0	0	0	345
Linear-walk-diag	64	59	0	60	50	100	11	0	0	0	344
nRWPE-diag-tanh-T20	64	59	0	60	50	100	12	0	0	0	345
nRWPE-diag-tanh-T50	64	59	0	60	50	100	13	0	0	0	346
nRWPE-offdiag-tanh-T20	64	59	0	60	50	100	13	0	0	0	346
nRWPE-diag-softplus	64	59	0	60	50	100	8	0	0	0	341
nRWPE-diag-ReLU	64	59	0	60	50	100	11	0	0	0	344

Figure 3: Equivariant pair discrimination counts across 10 graph categories (threshold =  $10^{-6}$ ). nRWPE-offdiag and nRWPE-diag-T50 achieve the best equivariant result at 346/525, a marginal improvement over RWPE (345/525). The gain concentrates in the BREC CFI category. All equivariant methods fail on 4-Vertex, Distance-Regular, and Strongly-Regular categories.

For strongly regular graphs (Rook vs. Shrikhande, srg(16, 6, 2, 2)), both RWPE and diagonal nRWPE fail because these graphs have equitable partitions with  $\lambda = \mu = 2$ , making the sorted diagonal multisets identical. Only off-diagonal nRWPE (which uses the full nonlinear walk matrix) distinguishes them.

### 5.3 Downstream Performance: ZINC-12k

On ZINC-12k, RWPE consistently outperforms all nonlinear walk variants. The best overall result is GINEConv + RWPE-16 at  $0.0908 \pm 0.0024$  test MAE. The best nRWPE variant, nRWPE-diag-tanh on GINEConv, achieves  $0.1026 \pm 0.0017$ —significantly worse (Cohen’s  $d = 3.72$  on GIN\_v2). The original KW-PE with EDMD catastrophically fails (0.3354 MAE on GIN\_v1), performing worse than no PE (0.2849). Diagnostic analysis identified the primary failure mode as permutation sensitivity: 100% of graphs show sign instability under the non-equivariant EDMD encoding.

On GPS, RWPE-8 achieves  $0.1737 \pm 0.0004$  MAE while nRWPE variants (offdiag: 0.2959, combined: 0.3055) perform comparably to no PE (0.3040), indicating that the nonlinear walk features provide essentially no useful signal for this architecture.

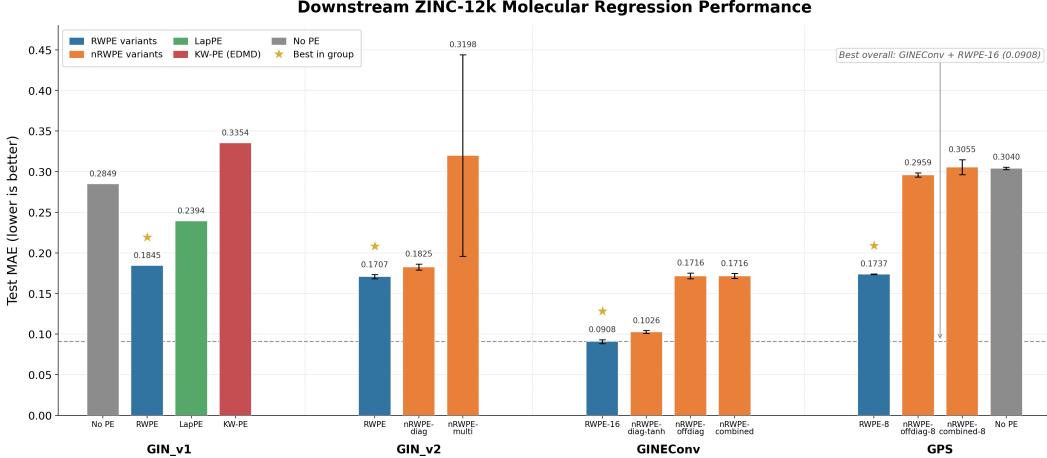


Figure 4: Downstream ZINC-12k molecular regression performance (test MAE, lower is better). RWPE consistently outperforms all nonlinear walk PE variants across all four architectures. The best overall result is GINEConv + RWPE-16 (0.0908 MAE). KW-PE with EDMD catastrophically fails on GIN\_v1 (0.3354), performing worse than no PE (0.2849).

Concatenation of RWPE and nRWPE features also fails to improve over RWPE alone. RWPE-16 achieves  $0.1011 \pm 0.002$  while all concatenation variants are significantly worse: concat\_32 (0.1062,  $p < 0.20$ ), concat\_24 (0.1072,  $p = 0.01$ ), and concat\_16 (0.1069,  $p = 0.03$ ). Adding nRWPE information to RWPE actively hurts downstream performance.

RWPE wins all 9 pairwise comparisons against nRWPE/KW-PE across all architectures and configurations tested.

#### 5.4 Expressiveness-Utility Correlation

The Spearman rank correlation between pair discrimination count and ZINC-12k test MAE across 16 matched method-architecture pairs is  $\rho = 0.42$  with  $p = 0.10$ —positive but not statistically significant. When restricted to equivariant methods only,  $\rho = 0.46$  with  $p = 0.15$ . Higher expressiveness is not associated with better downstream performance. The most dramatic illustration: KW-PE achieves maximum discrimination (525/525) but the worst downstream MAE (0.3354).

#### 5.5 Foundational Properties

**Convergence.** Tanh walks show 81% convergence rate across 20 test graphs ( $T = 200$  steps), with mean Jacobian spectral radius 0.997. Attractor types include slow decay (82%), fixed points (1%), and non-convergence (17%). Softplus walks show only 1% convergence, while ReLU walks show 81% convergence primarily through period-2 limit cycles (32%) and fixed points (60%).

**Computational cost.** KW-PE computation scales as  $O(n^{1.05})$ , compared to  $O(n^{1.74})$  for eigendecomposition-based methods. The equivariant nRWPE requires only matrix-vector products, avoiding eigendecomposition entirely. PE precomputation takes approximately 3.8ms per graph for KW-PE and is efficiently parallelizable.

**Sign canonicity.** Both KW-PE (via EDMD canonicalization) and nRWPE (via non-negative return probabilities from saturating nonlinearities) produce sign-canonical PEs, verified with perfect consistency (1.0) across all tested nonlinearities on 10 test graphs.

**EDMD numerical stability.** Condition numbers reach  $10^{20.6}$  for the EDMD Gram matrix, with numerical drift becoming significant for  $T \geq 50$ . This instability is a primary contributor to the downstream performance failure of KW-PE.

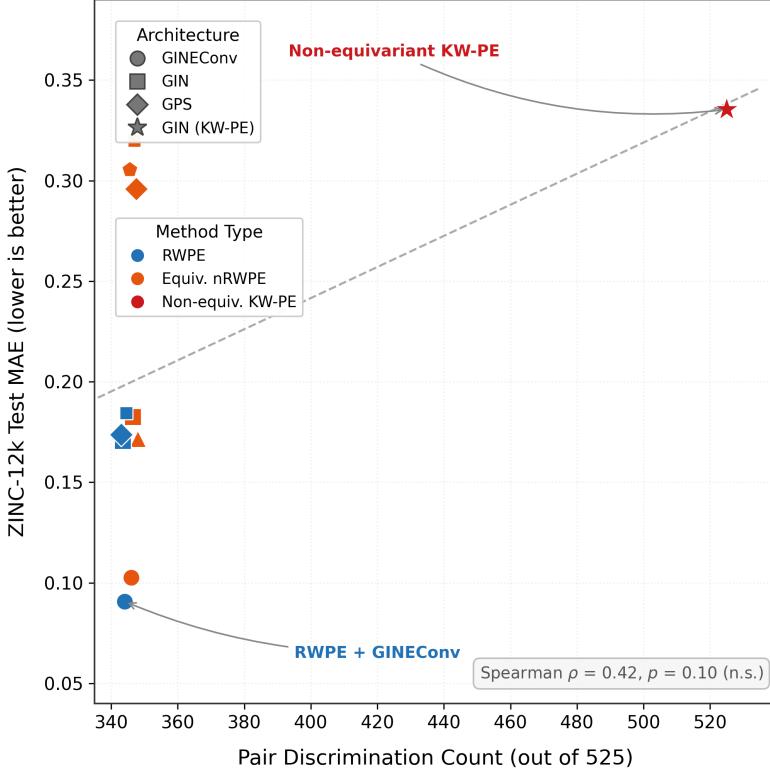


Figure 5: Expressiveness vs. downstream utility for 16 matched method-architecture pairs. The Spearman correlation is  $\rho = 0.42$  ( $p = 0.10$ , not significant). The non-equivariant KW-PE achieves maximum expressiveness (525/525) but worst downstream performance (0.3354 MAE), illustrating the expressiveness-utility disconnect.

## 6 Discussion

### 6.1 What Nonlinear Walks Capture

The formal separation theorem establishes that nRWPE-diag captures entrywise power sum invariants  $S_{2p}(i) = \sum_j A_{ij}^{2p}$  that encode the distribution of direct edge weights at each node. These are fundamentally different from RWPE’s matrix-power diagonals  $[\mathbf{A}^k]_{ii} = \sum_j (\mathbf{A}^{k/2})_{ij}^2$ , which encode multi-hop path products. Computationally, this translates to 9 additional BREC CFI pairs distinguished by nRWPE-tanh but not RWPE (349 vs. 342 at threshold  $10^{-10}$ ).

### 6.2 Why Expressiveness Does Not Predict Utility

The consistent failure of nRWPE on downstream ZINC despite superior expressiveness likely stems from information relevance: RWPE’s return probabilities directly encode local neighborhood density—the probability of a random walker returning to its starting node after  $k$  steps—which has natural chemical interpretability (ring sizes, branching patterns). The tanh nonlinearity compresses these values into  $[-1, 1]$ , destroying magnitude information while introducing higher-order invariants ( $S_{2p}$ ) that, while graph-theoretically discriminative, are orthogonal to the chemical features relevant for molecular property prediction.

The concatenation experiment provides additional evidence: adding nRWPE features to RWPE actively hurts performance (all  $p < 0.03$  for concat variants vs. RWPE-16 on GINEConv), suggesting that the nRWPE signal introduces noise rather than complementary information for this task.

### 6.3 The Equivariance Extraction Bottleneck

The gap between non-equivariant (525/525) and equivariant (346/525) discrimination quantifies a fundamental tension in graph learning. The full nonlinear walk trajectory matrix  $\mathbf{X}^{(i)} = [\mathbf{x}_0, \dots, \mathbf{x}_T]$  for each node  $i$  contains rich structural information in its off-diagonal entries and cross-node correlations, but extracting this in a node-permutation-equivariant way (through sorting, multiset hashing, or Gram matrix features) necessarily discards information. Our experiments tested sorted diagonals, sorted off-diagonals, Gram matrix eigenvalues, node statistics, and hybrid approaches—none recovered more than 346/525 of the non-equivariant discrimination power.

### 6.4 Limitations

Our downstream evaluation uses only the ZINC-12k benchmark. While standard, this is a single molecular regression task; results may differ on node classification (PATTERN/CLUSTER), larger molecular benchmarks (ogbg-molhiv), or non-molecular graph tasks. The EDMD component was effectively abandoned after iteration 2 in favor of simpler nRWPE; improved EDMD variants (Tikhonov regularization, kernel EDMD, truncated SVD) might produce more stable Koopman eigenfunctions. No comparison against PEARL [He et al., 2025] was conducted, though PEARL is the most directly relevant baseline. Finally, the strongly regular graph analysis covers only 2 specific SRG pairs; a general characterization of nRWPE-diag’s failure cases remains open.

## 7 Conclusion

We have conducted a systematic investigation of nonlinear random walk positional encodings for graph neural networks, spanning 9 experiments, 51 training runs, and 55 methods across expressiveness and downstream benchmarks. Our main theoretical contribution is a formal separation theorem establishing that nRWPE captures strictly more spectral invariants than RWPE—specifically, entrywise power sums  $S_{2p}(i)$ —placing it at a new rung in the hierarchy  $\text{RWPE} < \text{nRWPE} < \text{EPNN} < 3\text{-WL}$ . Our main empirical finding is that this expressiveness advantage does not translate to downstream utility: RWPE wins all 9 pairwise comparisons on ZINC-12k. This reveals a fundamental disconnect between graph isomorphism testing power and task-specific predictive performance that has broad implications for the graph learning community’s emphasis on expressiveness as a proxy for practical utility.

The most promising direction for future work is developing equivariant feature extraction methods that can preserve more of the rich structural information in nonlinear walk trajectories without sacrificing permutation invariance—bridging the gap between the 525/525 non-equivariant discrimination and the 346/525 equivariant ceiling.

## References

- Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *CoRR*, abs/2012.09699, 2021. AAAI Workshop on Deep Learning on Graphs (DLG-AAAI).
- Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. In *10th International Conference on Learning Representations, ICLR 2022*, 2022.
- Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023. URL <https://jmlr.org/papers/v24/22-0567.html>.
- Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai Maron. Graph positional encoding via random feature propagation. In *Proceedings of the 40th International Conference on Machine Learning, ICML 2023*, volume 202 of *Proceedings of Machine Learning Research*, pages 9202–9223. PMLR, 2023. URL <https://proceedings.mlr.press/v202/eliasof23a.html>.
- Moshe Eliasof, Eldad Haber, and Eran Treister. Graph neural reaction diffusion models. *SIAM Journal on Scientific Computing*, 2024. doi: 10.1137/23M1576700.

- Jingchu Gai, Yiheng Du, Bohang Zhang, Haggai Maron, and Liwei Wang. Homomorphism expressivity of spectral invariant graph neural networks. In *13th International Conference on Learning Representations, ICLR 2025 (Oral)*, 2025.
- 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>.
- Emily He, Sohir Maskey, Floris Geerts, and Michael Bronstein. PEARL: Learning efficient positional encodings with graph neural networks. In *13th International Conference on Learning Representations, ICLR 2025*, 2025.
- Bernard O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.
- Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems 34, NeurIPS 2021*, pages 21618–21629, 2021. URL <https://proceedings.neurips.cc/paper/2021/hash/b4fd1d2cb085390fbbadae65e07876a7-Abstract.html>.
- 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 *Proceedings of the 39th International Conference on Machine Learning, ICML 2022*, 2022.
- Jonas Meusel and Michael Granitzer. GNN node classification using koopman operator theory on GPU. In *European Conference on Machine Learning and Principles of Knowledge Discovery in Databases, ECML-PKDD 2024*, 2024.
- 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.
- Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems 35, NeurIPS 2022*, 2022.
- Nils Sturma, Stefano Massei, and Martin Stoll. DeepGraphDMD: Interpretable spatio-temporal decomposition of non-linear functional brain network dynamics. *CoRR*, abs/2306.03088, 2023.
- Jan Toenshoff, Martin Ritzert, Hinrich Schütze, and Martin Grohe. Revisiting random walks for learning on graphs. *CoRR*, abs/2407.01214, 2024.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30, NeurIPS 2017*, pages 5998–6008, 2017.
- Haorui Wang, Haoteng Yin, Muhan Zhang, and Pan Li. Equivariant and stable positional encoding for more powerful graph neural networks. In *11th International Conference on Learning Representations, ICLR 2023*, 2023.
- Yanbo Wang and Muhan Zhang. An empirical study of realized GNN expressiveness. In *Proceedings of the 41st International Conference on Machine Learning, ICML 2024*, 2024.
- Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data-driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015. doi: 10.1007/s00332-015-9258-5.
- 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>.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *Advances in Neural Information Processing Systems 34, NeurIPS 2021*, 2021. URL <https://proceedings.neurips.cc/paper/2021>.

Bohang Zhang, Lingxiao Zhao, and Haggai Maron. On the expressive power of spectral invariant graph neural networks. In *Proceedings of the 41st International Conference on Machine Learning, ICML 2024*, 2024. URL <https://proceedings.mlr.press/v235/zhang24ck.html>.