MAGRESPE: FOREST-NORMALIZED MAGNITUDE RESIDUAL POSITIONAL ENCODING FOR GRAPH TRANSFORMERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Transformers rely on positional encodings (PEs) to expose graph structure to attention, yet practical positional encodings provide only a finite structural interface. We introduce MAGRESPE, a deterministic multi-scale node PE derived from graph magnitude by measuring the ridge-regularized residual from an exact forest reference. This forest normalization gives every forest an explicit zero baseline while retaining non-forest structure across metric scales. We derive an all-orders BFS-shell expansion of the forest-prediction defect, whose coefficients are determined by same-shell edges and redundant shortest-path parents. Its second-order coefficient captures triangles and redundant two-hop geodesics, and the expansion yields a girth-controlled first nonzero order for the positive-ridge residual. We further prove a regular-subdivision separation: for arbitrarily prescribed finite horizons, connected graphs can share identical RRWP rooted row-profile multisets while exhibiting distinct small-q MAGRESPE onset orders. An exact cycle folding law further gives the sharp RRWP separation horizon. Empirically, on ZINC-12K, augmenting RRWP with MAGRESPE reduces test MAE from 0.0638 ± 0.0036 to 0.0600 ± 0.0021. Matched formulation controls and a SignNet spectral-node control do not reproduce this gain, indicating that it is not explained by raw magnitude or by the tested spectral-node augmentation. Using the same fixed MAGRESPE configuration, the hybrid also improves mean RRWP performance on Peptides-struct and PascalVOC-SP. Standalone replacement is not consistently competitive, supporting MAGRESPE as a complementary rather than universal PE. Together, these results suggest that forest-normalized graph magnitude is a promising basis for complementary positional encoding in Graph Transformers.

1 INTRODUCTION

Graph-structured data arise naturally in domains ranging from social and knowledge networks to molecular systems (Otte & Rousseau, 2002; Wang et al., 2024; Nguyen et al., 2024). Message-passing graph neural networks have become a standard framework for learning from such data, but their reliance on repeated local aggregation can make long-range dependencies difficult to capture (Alon & Yahav, 2021). Graph Transformers (GTs) provide an appealing alternative by allowing each node to interact through global self-attention (Ying et al., 2021; Kreuzer et al., 2021; Rampášek et al., 2022; Ma et al., 2023). Global attention alone, however, does not tell the model how nodes are situated within the graph. Unlike sequences or images, graphs have neither a canonical node ordering nor a fixed coordinate system, so graph structure must be supplied through additional inductive biases.

Positional and structural encodings have therefore become a central component of Graph Transformers. Following GRIT (Ma et al., 2023), we use *positional encoding* (PE) as an umbrella term for graph encodings often described as positional or structural, since many practical methods contain both kinds of information. Early and widely used approaches encode graph structure through shortest-path distances, Laplacian eigenvectors, or random walk statistics (Ying et al., 2021; Kreuzer et al., 2021; Rampášek et al., 2022). More recent work has developed sign- and basis-invariant spectral representations, learned transferable PEs, simple path statistics, and motif or homomorphism-based encodings (Lim et al., 2023; Cantürk et al., 2024; Airale et al., 2025). These developments reflect a common goal: expose structural relations that self-attention cannot infer reliably from node attributes

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.

alone. At the same time, controlled benchmarking shows that no PE family dominates uniformly; its usefulness depends on both the task and the host architecture (Grötschla et al., 2026).

Among these approaches, relative random walk probabilities (RRWP) provide a practical structural encoding based on finite sequences of random walk statistics. GRIT (Ma et al., 2023) uses RRWP through both node-level return probabilities and node-pair transition profiles, allowing structural information to influence node representations and attention. Simple path structural encoding (SPSE) subsequently showed that replacing random walk statistics with path-sensitive information can expose cyclic structures that finite random walk summaries may represent less directly (Airale et al., 2025). More broadly, recent theory emphasizes an important distinction between the information contained in a complete PE and the information made available by a finite practical interface (Black et al., 2024; Flora et al., 2026). We therefore do not argue that random walk information is fundamentally insufficient. Instead, we ask a more practical question: *can a compact node PE derived from the global shortest-path metric provide information complementary to a finite positional RRWP interface?*

We approach this question through *metric magnitude* (Leinster, 2013; 2019), which assigns a scale-dependent weighting to the shortest-path metric. On every forest, this weighting has an exact degree-local form, providing a natural reference for general graphs. We introduce MAGRESPE, which encodes the ridge-regularized node-wise correction to this forest-exact reference across multiple metric scales. A breadth-first search (BFS) shell analysis shows that non-tree-like relations–same-shell edges and excess shortest-path parents–determine the small-scale expansion of the forest-prediction defect, while graph girth controls the first nonzero order of the positive-ridge residual. Shorter cycles therefore appear earlier in the hierarchy, although the onset does not uniquely recover girth. The same analysis shows that scales concentrated near zero collapse toward the leading statistic, motivating a genuinely multi-scale representation.

MAGRESPE also differs from RRWP in how structural information is exposed to the model: RRWP materializes a finite sequence of random walk statistics, whereas MAGRESPE processes the complete shortest-path metric into a fixed-width node PE. We formalize this distinction through connected graph families whose truncated RRWP rooted row-profile summaries coincide while their MAGRESPE residual signatures differ. This representation-level separation motivates our hybrid design, which retains the original RRWP information and adds MAGRESPE through the node-input route. We test the theoretical and empirical contributions of MAGRESPE through controlled synthetic experiments and real-world benchmarks. On two theory-aligned synthetic classification tasks, the hybrid achieves perfect test accuracy on the cycle family and 99.17% on connected regular subdivisions, while the finite-horizon RRWP baseline remains at chance under the tested settings. On ZINC, augmenting RRWP with MAGRESPE reduces test MAE from 0.0638 ± 0.0036 to 0.0600 ± 0.0021. Matched raw-magnitude and a SignNet control do not reproduce this gain, providing evidence beyond raw magnitude and the tested spectral-node augmentation. The hybrid also improves mean RRWP performance on Peptides-struct and PascalVOC-SP using the same fixed scale bank. Nevertheless, standalone MAGRESPE is not consistently competitive on the evaluated benchmarks. Together, these theoretical and empirical results support MAGRESPE as a complementary structural encoding rather than a universal replacement for existing PEs.

Contributions. Our main contributions are:

- We introduce MAGRESPE, a deterministic multi-scale node positional encoding derived from graph magnitude by measuring the residual from an exact forest reference. This construction gives forests an explicit zero baseline and converts global shortest-path geometry into a compact node-level signal that can be combined with, rather than replace, a PE such as RRWP.
- We characterize MAGRESPE by proving that every forest maps exactly to zero, deriving an all-orders BFS-shell expansion of the forest-prediction defect and its second-order interpretation via triangles and redundant two-hop geodesics, establishing a girth-controlled onset for the positive-ridge residual, and proving a general regular-subdivision separation from finite-horizon RRWP rooted row-profile summaries. For cycles, an exact folding law yields the sharp separation horizon.
- We evaluate MAGRESPE on real-world molecular and non-molecular benchmarks and theory-aligned synthetic tasks using matched RRWP comparisons, magnitude and spectral controls, and scale and formulation ablations. The results show that MAGRESPE provides complementary information to finite-horizon RRWP.

2 Related Work

2.1 Structural encodings and finite PE interfaces

Graph Transformers incorporate structural information through node encodings, pairwise representations, and local message passing. Graphormer uses centrality and shortest-path biases (Ying et al., 2021), SAN introduces spectral attention (Kreuzer et al., 2021), GraphGPS combines local message passing with global attention (Rampášek et al., 2022), and GRIT incorporates finite-horizon RRWP through node-pair transition profiles and node-level return-probability features (Ma et al., 2023). Subsequent positional encodings address spectral symmetries and stability (Lim et al., 2023; Huang et al., 2024), transferability and scalability (Cantürk et al., 2024; Kanatsoulis et al., 2025), and multiresolution spectral and spatial structure through WavePE (Ngo et al., 2023). Complementary approaches capture simple paths (Airale et al., 2025), homomorphism counts (Bao et al., 2025), persistent homology (Verma et al., 2025), local topological structure through LEAP (Amboage et al., 2026), and cycle-aware topology through TIGT (Choi et al., 2024). Kernel-based encodings provide another route to longer-range structural information (Lee et al., 2025).

Complete and truncated interfaces should be distinguished. Black et al. (2024) compare absolute and relative PEs under sufficiently expressive transformer constructions, while Flora et al. (2026) show that truncation can break complete-positional encoding equivalences and that combinations of truncated families can be useful. PE ensembling has also been explored directly (Yeom et al., 2024; Feng et al., 2026). Our novelty is therefore not that two PEs can be combined, but the forest-normalized magnitude residual, its structural hierarchy, and its complementary use alongside the existing finite-horizon RRWP representation.

2.2 Magnitude and graph learning

Magnitude is a scale-dependent invariant of finite metric spaces with a graph specialization and vertex weightings (Leinster, 2013; 2019). Prior graph-learning work already uses metric magnitude, notably for geometry-aware graph pooling (Limbeck et al., 2025). Our construction instead subtracts an exact forest weighting on arbitrary graphs and exposes the resulting node-wise correction as a multi-scale PE. Magnitude homology is a distinct categorified object; prior work connects its diagonality to graph girth (Asao et al., 2024), whereas our girth theorem concerns the Taylor onset of an ordinary magnitude-weighting residual. Recent hierarchical low-rank work on metric magnitude suggests a plausible direction for scaling the exact preprocessing (Huntsman et al., 2026), but no approximate MagResPE solver is claimed here.

3 Preliminaries and Motivation

3.1 Graph notation and positional-encoding interfaces

Let $G = (V, E, X)$ be a finite simple undirected graph with $n = |V|$ nodes, edge set $E \subseteq V \times V$, and node features $X \in \mathbb{R}^{n \times d_x}$. We write $A \in \{0,1\}^{n \times n}$ for the adjacency matrix, $d_v = \sum_u A_{vu}$ for the degree of node v, and $\Delta = \mathrm{diag}(d_1, \ldots, d_n)$ for the degree matrix. Let I_n denote the $n \times n$ identity matrix and $\mathbf{1} \in \mathbb{R}^n$ the all-ones vector. Although the predictive model may additionally use node or edge attributes, MagResPE itself depends only on the graph topology.

For $u, v \in V$, let $D(u, v)$ denote their shortest-path distance, with $D(u, v) = \infty$ when they lie in different connected components. A *forest* is an acyclic graph, and the *girth* g of a cyclic graph is the length of its shortest cycle.

A positional encoding supplies structural information beyond the ordinary input attributes. It may be *node-wise*, $p : V \to \mathbb{R}^{d_p}$, or *pairwise*, $\psi : V \times V \to \mathbb{R}^{d_r}$. In this terminology, MagResPE is a node PE, while RRWP provides both node-level and node-pair structural information. Node-level features are constructed from random walk return probabilities, while node-pair features capture transition probabilities between vertices. Accordingly, our "RRWP interface" refers to the complete finite RRWP information supplied by the implementation, including its diagonal.

3.2 FINITE RRWP AND THE GRIT INTERFACE

Define $P = \Delta^{\dagger} A$, where $\Delta^{\dagger}_{vv} = d_v^{-1}$ for $d_v > 0$ and 0 otherwise. Isolated vertices therefore have zero transition rows, and for $d_u > 0$, $(P^k)_{uv}$ gives the k-step random walk transition probability from u to v. At a finite horizon h, the relative random walk probability profile of an ordered node pair (u, v) is

$$\Psi_h(u, v) = \left[I_n, P, P^2, \ldots, P^h\right]_{uv} \in \mathbb{R}^{h+1}. \tag{1}$$

Our GRIT implementation (Ma et al., 2023) uses RRWP at both the node and node-pair levels. The diagonal profiles $\Psi_h(v, v)$ provide node-level return-probability features, while the full profiles $\Psi_h(u, v)$ initialize learned representations for ordered node pairs. GRIT subsequently updates its node and node-pair representations through attention. Thus, the RRWP encoding considered in this work comprises both uses of the finite tensor $\{\Psi_h(u, v)\}_{u,v \in V}$. The horizon h determines the maximum walk length represented by the encoding; truncating this sequence can change the structural information available to the model (Flora et al., 2026).

This provides the comparison point for MAGRESPE. Whereas RRWP captures random walk statistics over a finite horizon, MAGRESPE processes the complete shortest-path distance matrix into a compact multi-scale node PE. Our theoretical comparisons in Section 5 concern rooted row-profile summaries derived from finite-horizon RRWP, including their diagonal return-probability information, rather than unrestricted random walk representations.

3.3 METRIC MAGNITUDE AND THE FOREST REFERENCE

Metric magnitude provides a scale-dependent summary of a finite metric space (Leinster, 2013; 2019). For the graph shortest-path metric and scale $q \in (0, 1)$, define the magnitude kernel

$$Z_q(u, v) = q^{D(u,v)}, \qquad q^{\infty} := 0. \tag{2}$$

Small q rapidly suppresses contributions from distant nodes, whereas larger q retains information from progressively more distant metric shells. A raw magnitude weighting $w_q \in \mathbb{R}^n$ satisfies $Z_q w_q = \mathbf{1}$ whenever the system is solvable.

The property that motivates MAGRESPE is that this weighting has an exact degree-local form on every forest:

$$b_q(v) = \frac{1 - (d_v - 1)q}{1 + q}. \tag{3}$$

Hence $Z_q b_q = \mathbf{1}$ on every forest. Rather than interpreting b_q as a decomposition of an arbitrary graph into "tree" and "cyclic" components, we use it only as an *exact forest reference*: on a general graph, we ask how much the observed metric departs from the prediction made by this reference.

A rooted breadth-first search gives the intuition. In a forest, every non-root vertex has exactly one parent in the preceding distance shell, and no edges join vertices in the same shell. Cycles force departures from this tree-like organization through excess shortest-path parents or same-shell edges. Section 4 first quantifies this discrepancy through the forest-prediction defect e_q, then computes the ridge-regularized correction $r_{q,\lambda}$ used to construct MAGRESPE. Section 5 characterizes their small-scale structural properties.

4 MAGRESPE: FOREST-NORMALIZED MAGNITUDE RESIDUAL ENCODING

Given the shortest-path magnitude kernel Z_q and the exact forest reference b_q from Section 3, MAGRESPE measures the node-wise correction required for the observed graph, normalizes this correction across metric scales, and supplies the resulting signature through the node-input route of a Graph Transformer.

4.1 FOREST-NORMALIZED RESIDUAL

For an arbitrary graph, we evaluate the forest reference b_q and define its *forest-prediction defect*

$$e_q = \mathbf{1} - Z_q b_q. \tag{4}$$

Thus e_q measures how far the forest-exact weighting fails to satisfy the magnitude equations on the observed graph. Rather than solving for a new weighting from scratch, we seek a correction r to this reference:

$$r_{q,\lambda} = \arg\min_{r\in\mathbb{R}^n} \|Z_q(b_q + r) - \mathbf{1}\|_2^2 + \lambda \|r\|_2^2, \qquad \lambda > 0. \tag{5}$$

Equivalently,

$$r_{q,\lambda} = (Z_q^\top Z_q + \lambda I)^{-1} Z_q^\top e_q. \tag{6}$$

The positive ridge makes the correction uniquely defined even when the unregularized magnitude system is singular or poorly conditioned. The forest normalization is structural rather than learned: if G is a forest, then $Z_q b_q = \mathbf{1}$ exactly and hence $e_q = r_{q,\lambda} = 0$ for every q and every $\lambda > 0$. On a general graph, $r_{q,\lambda}$ therefore records the regularized correction to the forest-exact reference rather than the raw magnitude weighting itself.

4.2 Multi-scale node signature

A single scale q emphasizes one range of the shortest-path metric. Small q concentrates strongly on nearby metric shells, whereas larger q retains progressively more distant contributions. We therefore evaluate the residual over a fixed bank $\mathcal{Q} = \{q_1, \ldots, q_K\} \subset (0, 1)$ of $K = |Q|$ metric scales.

Section 5 shows that the earliest possible graph-dependent residual occurs at second order in q. We use this universal onset to normalize the scale-dependent magnitude of the residual, followed by a signed logarithm:

$$p_v(q) = \operatorname{sgn}(r_{q,\lambda}(v)) \log\left(1 + \left|\frac{r_{q,\lambda}(v)}{q^2}\right|\right). \tag{7}$$

The q^{-2} factor aligns the universal leading scale without assuming that every graph first becomes nonzero at second order; longer-girth structures can enter later. The signed logarithm preserves the direction of the correction while compressing its dynamic range. The final MagResPE representation of node v is the deterministic multi-scale signature

$$p_v = \big[p_v(q_1), \ldots, p_v(q_K)\big] \in \mathbb{R}^K. \tag{8}$$

4.3 Node-level integration with GRIT

RRWP provides GRIT with both node-level return-probability features and node-pair transition profiles, whereas MagResPE produces a fixed-width, metric-derived structural vector p_v for each node. We map this signature through a learned node encoder f_{PE} and concatenate it with the ordinary dataset embedding:

$$h_v^{(0)} = \big[h_{\mathrm{data}}(v) \,\|\, h_{\mathrm{PE}}(v)\big], \qquad h_{\mathrm{PE}}(v) = f_{\mathrm{PE}}(p_v). \tag{9}$$

In the hybrid configuration, node-level RRWP features are incorporated after this initialization, while the original node-pair RRWP profiles are also preserved. The subsequent GRIT layers remain unchanged.

We evaluate two configurations. The standalone configuration retains the ordinary dataset features and the proposed MagResPE node PE while removing both node-level and node-pair RRWP information. The hybrid preserves both components of the original RRWP encoding and additionally incorporates the same MagResPE node branch. Comparing these configurations with the RRWP baseline allows us to assess MagResPE as a standalone representation and its complementary contribution to finite-horizon random walk information. Exact encoder dimensions and parameter accounting are provided in Appendix B.

Computation. For each graph, MagResPE requires all-pairs shortest-path distances and, for each $q \in \mathcal{Q}$, an $n \times n$ ridge solve. With direct dense linear algebra this gives cubic worst-case solve cost per scale and quadratic working memory, although the resulting node representation has only $O(nK)$ size. Consequently, the current exact implementation is suitable for collections of small and medium graphs but is not intended for very large individual graphs. Controlled runtime, memory, and numerical diagnostics are reported in Appendix E.

5 What Structural Information Does MagResPE Encode?

The forest normalization yields three complementary views of MagResPE: a girth-controlled structural hierarchy, an explicit local obstruction that explains the need for multiple scales, and a finite-interface complementarity result with RRWP.

5.1 Forest null class and girth-controlled onset

For node sets $S, T \subseteq V$, let $E(S,T)$ denote the set of edges with one endpoint in S and the other in T. When $S = T$, each edge induced by S is counted once. For a root v, let $L_j(v) = \{u : D(v,u) = j\}$ denote its jth BFS shell. Define

$$\begin{aligned} x_j(v) &= |E(L_{j-1}(v), L_j(v))| - |L_j(v)|, \\ m_j(v) &= |E(L_j(v), L_j(v))|, \\ a_j(v) &= x_j(v) + x_{j+1}(v) + 2m_j(v), \end{aligned} \tag{10}$$

where x_j counts excess shortest-path parents and m_j counts same-shell edges.

Theorem 1. *Fix $\lambda > 0$. If G is a forest, then $r_{q,\lambda} \equiv 0$ for every $q \in (0,1)$. Otherwise, let g be the girth of G and $k = \lceil g/2 \rceil$. Then $a_j \equiv 0$ for $j \le k-2$, while a_{k-1} is nonnegative and nonzero on vertices of a shortest cycle. As $q \to 0$,*

$$e_q = q^k a_{k-1} + O(q^{k+1}), \qquad r_{q,\lambda} = \frac{q^k}{1+\lambda} a_{k-1} + O(q^{k+1}). \tag{11}$$

Exact finite-scale defect. Unlike the asymptotic ridge expansion in Theorem 1, the forest-prediction defect admits the exact identity $e_q(v) = \frac{1}{1+q} \sum_{j=1}^{R_v} a_j(v) q^{j+1}$ for every $q \in (0,1)$, where R_v is the eccentricity of v within its connected component. Since $a_j(v) \ge 0$, the defect is nonnegative at every scale and is strictly positive precisely when v belongs to a component containing a cycle. Expanding $(1+q)^{-1}$ recovers the all-orders coefficient identity and the defect onset in Theorem 1. The complete derivation is given in Appendix A.3.

5.2 Local obstruction and why multiple scales matter

The leading nonzero coefficient has an explicit combinatorial interpretation. Let $B_2 = \mathbf{1}[D(u,v) = 2]$, $d = A\mathbf{1}$, and $n_2 = B_2\mathbf{1}$. Let T_v denote the number of triangles incident to v and let $\gamma_{vx} = |N(v) \cap N(x)|$. Then

$$r_{q,\lambda} = \frac{q^2}{1+\lambda} c + O(q^3), \qquad c_v = 2T_v + \sum_{x : D(v,x)=2} (\gamma_{vx} - 1). \tag{12}$$

Thus, the universal second-order obstruction counts incident triangles and redundant two-hop shortest paths. This term alone does not determine the multi-scale representation. Writing $\phi(t) = \operatorname{sgn}(t) \log(1 + |t|)$, the normalized channel satisfies

$$p_v(q) \longrightarrow \phi\left(\frac{c_v}{1+\lambda}\right) \qquad (q \to 0). \tag{13}$$

Hence a scale bank concentrated entirely near zero becomes asymptotically redundant. This gives a theoretical reason to retain finite scale diversity rather than using only the leading coefficient.

Appendix A.4 gives the corresponding local distance-layer expansion: the unregularized residual can be organized as rooted distance-labelled path-homomorphism terms, while Appendix A.5 gives the positive-ridge local analytic counterpart.

5.3 Complementarity with finite-horizon RRWP

Recent work emphasizes that complete and truncated PEs can have different distinguishing power (Black et al., 2024; Flora et al., 2026). We therefore compare MagResPE with the finite RRWP interface actually materialized by GRIT rather than with unrestricted random walk information.

For a root u, let $\mathcal{R}_h(u) = \{\{\Psi_h(u, v) : v \in V\}\}$ and $\mathcal{S}_h(G) = \{\{\mathcal{R}_h(u) : u \in V\}\}$. We study the permutation-invariant rooted row-profile summary $\mathcal{S}_h(G)$ under constant original node attributes. Each rooted summary contains the transition profiles $\Psi_h(u, v)$ for all $v \in V$, including the diagonal return-probability profile $\Psi_h(u, u)$.

Theorem 2. *Let G_1 and G_2 be finite connected simple d-regular graphs on the same number of vertices, with girths $g_1 < g_2$. For $L \geq 2$, let $S_i = \mathrm{Sub}_L(G_i)$ and $i \in \{1, 2\}$, where every base edge is replaced by a path of L edges. If $2h < g_1 L$ then $\mathcal{S}_h(S_1) = \mathcal{S}_h(S_2)$. Nevertheless, for every fixed $\lambda > 0$, the* MAGRESPE *residual multisets of S_1 and S_2 differ for all sufficiently small positive q.*

Uniform subdivisions of same-order d-regular graphs share identical horizon-h RRWP rooted row-profile summaries when their girths exceed $2h$, yet MAGRESPE distinguishes those with different girths for sufficiently small positive q and fixed $\lambda > 0$. This result concerns rooted row-profile summaries and does not establish indistinguishability under the complete RRWP pair representation or arbitrary RRWP-equipped Graph Transformers. The proof is given in Appendix A.6.

A sharp cycle family. For C_{2m} versus $2C_m$, the two-fold covering gives an exact folding identity for every random walk power. Their RRWP row-profile multisets coincide exactly when $h < h_\star(m) = \left\lceil \frac{m}{2} \right\rceil$ and differ at every later horizon. At the first separating horizon, the newly folded probability mass is exactly $2^{-h_\star(m)}$. In contrast, for every $q \in (0, 1)$ and $0 < \lambda < 1$, a single MAGRESPE channel separates the pair. Appendix A.7 gives the exact folding law and proof.

6 EXPERIMENTS

6.1 EXPERIMENTAL SETUP

We evaluate MAGRESPE on three real-world benchmarks with complementary roles. ZINC-12K (Dwivedi et al., 2023) serves as the primary graph regression and development benchmark. We evaluate the fixed MAGRESPE construction on two Long Range Graph Benchmark (LRGB) datasets (Dwivedi et al., 2022): Peptides-struct (molecular graph regression) and PascalVOC-SP (non-molecular superpixel node classification). We additionally consider two theory-aligned synthetic classification tasks: a paired cycle family testing the sharp RRWP separation threshold and a connected regular-subdivision family testing generalization to unseen base isomorphism classes. All experiments use GRIT as the Graph Transformer backbone. Unless explicitly varied in an ablation, we use the same fixed MAGRESPE hyperparameters $\mathcal{Q} = \{0.05, 0.08, 0.13, 0.21, 0.34, 0.52, 0.74, 0.95\}$ and $\lambda = 10^{-3}$. Checkpoints are selected by validation performance, and the test split is evaluated once at the selected checkpoint for each run. We report mean and sample standard deviation over five seeds. Dataset specifications, model configurations, and optimization details are provided in Appendix B.

6.2 REAL-WORLD BENCHMARK VALIDATION

Molecular benchmarks. Table 1 compares MAGRESPE with existing structural encodings on ZINC and Peptides-struct. On ZINC, RRWP+MAGRESPE achieves the lowest test MAE of 0.0600 ± 0.0021, improving over the RRWP baseline (0.0638 ± 0.0036), SPSE (0.0611 ± 0.0031), and RRWP+SignNet (0.0681 ± 0.0032). On Peptides-struct, the hybrid also achieves the lowest reported mean MAE of 0.2457 ± 0.0022, compared with 0.2479 ± 0.0047 for RRWP and 0.2483 ± 0.0016 for SPSE. Although the improvements over SPSE and RRWP on Peptides-struct are modest, the results show consistent mean improvements from augmenting RRWP with MAGRESPE across both molecular tasks. In contrast, standalone MAGRESPE does not outperform RRWP on either dataset, suggesting that its principal benefit lies in complementing rather than replacing existing random walk encodings.

Non-molecular benchmark. Table 2 extends the evaluation to PascalVOC-SP, a non-molecular node classification benchmark. RRWP+MAGRESPE achieves a test F1 of 0.4070 ± 0.0041, compared with 0.3968 ± 0.0108 for RRWP, yielding an absolute improvement of 0.0102. The hybrid also exhibits lower variability across seeds. In contrast, standalone MAGRESPE achieves only 0.2363 ± 0.0022, substantially below RRWP. Together with the molecular results, these findings suggest that MAGRESPE provides a complementary structural signal that can benefit different prediction tasks, while its standalone performance highlights the importance of retaining the RRWP encoding.

Table 1: Test performance on two molecular benchmarks. RRWP, SPSE, and RRWP+SignNet are independently retrained over five random seeds, with results reported as mean ± sample standard deviation. Differences from published results may reflect variations in experimental protocols.

Model	ZINC MAE ↓	Peptides-struct MAE ↓
GINE	–	0.3547 ± 0.0045
SAN (+RRWP)	0.139 ± 0.006	0.2545 ± 0.0012
GraphViT	0.073 ± 0.001	0.2475 ± 0.0015
Exphormer	–	0.2481 ± 0.0007
Drew	–	0.2536 ± 0.0015
GPS+GPSE	0.065 ± 0.003	0.2464 ± 0.0025
RRWP	0.0638 ± 0.0036	0.2479 ± 0.0047
SPSE	0.0611 ± 0.0031	0.2483 ± 0.0016
RRWP+SignNet	0.0681 ± 0.0032	0.2527 ± 0.0033
MAGRESPE	0.0645 ± 0.0045	0.2544 ± 0.0006
RRWP+MAGRESPE	$\mathbf{0.0600 \pm 0.0021}$	$\mathbf{0.2457 \pm 0.0022}$

Table 2: PascalVOC-SP node classification performance. RRWP is independently retrained over five random seeds, with results reported as mean ± sample standard deviation. Differences from published results may reflect variations in experimental protocols.

Model	Train F1	Test F1 ↑
GINE	0.2088 ± 0.0268	0.1265 ± 0.0076
GatedGCN	0.3552 ± 0.0451	0.2873 ± 0.0219
GatedGCN+LapPE	0.3512 ± 0.0167	0.2860 ± 0.0085
Transformer+LapPE	0.7170 ± 0.0048	0.2694 ± 0.0098
SAN+LapPE	0.5723 ± 0.0427	0.3230 ± 0.0039
SAN+RWSE	0.5819 ± 0.0331	0.3216 ± 0.0027
RRWP	0.7148 ± 0.0887	0.3968 ± 0.0108
MAGRESPE	0.3250 ± 0.0119	0.2363 ± 0.0022
RRWP+MAGRESPE	0.7511 ± 0.0227	$\mathbf{0.4070 \pm 0.0041}$

6.3 THEORY-ALIGNED SYNTHETIC VALIDATION

We evaluate whether the theoretical separations in Section 5.3 are accessible to a learned GRIT model using two controlled binary classification tasks.

Sharp cycle boundary. Following Theorem 14, we construct C_{2m} and $2C_m$ for $m \in \{41, \dots, 80\}$, yielding 80 graphs with paired examples assigned to the same data split. Their exact RRWP separation horizon is $h_\star(m) = \lceil m/2 \rceil$. At $h = 20$, the rooted RRWP row-profile summaries of all pairs coincide: NoPE (no positional encoding) and RRWP remain at chance, whereas MAGRESPE and the hybrid achieve perfect test accuracy across five seeds (Table 3). Increasing the RRWP horizon to 30 and 40 makes the rooted row-profile summaries mathematically distinguishable for half and all of the cycle pairs, respectively. Nevertheless, the trained RRWP model remains at chance on the held-out test pairs. This distinguishes structural accessibility from successful learning under the tested architecture and optimization protocol; it is not a further expressivity limitation implied by the theorem.

Connected regular subdivisions. To evaluate generalization beyond the cycle family, we select 240 non-isomorphic connected cubic graphs on 22 vertices, equally divided between girths 5 and 6, and uniformly subdivide their edges with $L = 9$. The resulting graphs have 286 vertices, identical degree and subdivision-type histograms, and girths 45 and 54. By Theorem 2, all graphs share the same horizon-20 RRWP rooted row-profile summary, since $2h = 40 < 45$. In the learned experiment, NoPE and RRWP both remain at chance, whereas MAGRESPE and the hybrid achieve 0.9833 and 0.9917 mean test accuracy, respectively, on unseen base isomorphism classes. The small difference between these two conditions does not establish a consistent hybrid advantage. Together, these experiments demonstrate that a learned GRIT model can exploit the MAGRESPE signal under controlled finite-horizon RRWP collisions, without implying superiority over complete RRWP.

Table 3: Theory-aligned synthetic classification results. Test accuracy is reported as mean ± sample standard deviation over five seeds.

Condition	RRWP horizon	Test accuracy ↑
Cycle family: C_{2m} versus $2C_m$		
NoPE	–	0.5000 ± 0.0000
RRWP	20	0.5000 ± 0.0000
RRWP	30	0.5000 ± 0.0000
RRWP	40	0.5000 ± 0.0000
MAGRESPE	–	$\mathbf{1.0000 \pm 0.0000}$
RRWP+MAGRESPE	20	$\mathbf{1.0000 \pm 0.0000}$
RRWP+MAGRESPE	30	$\mathbf{1.0000 \pm 0.0000}$
RRWP+MAGRESPE	40	$\mathbf{1.0000 \pm 0.0000}$
Connected regular-subdivision family		
NoPE	–	0.5000 ± 0.0000
RRWP	20	0.5000 ± 0.0000
MAGRESPE	–	0.9833 ± 0.0174
RRWP+MAGRESPE	20	$\mathbf{0.9917 \pm 0.0114}$

6.4 WHICH PARTS OF THE FORMULATION MATTER?

We examine three formulation choices under the matched ZINC-12K hybrid configuration: forest normalization, onset normalization, and the ridge correction. RawMagnitude uses the raw ridge magnitude weighting without forest subtraction or q^{-2} normalization. ResidualCommon introduces the forest-centered ridge residual without onset normalization, while the standard MAGRESPE configuration uses both components. To isolate the role of the ridge solve, we additionally evaluate DirectDefect, which uses the same forest-prediction defect $e_q = \mathbf{1} - Z_q b_q$ but replaces the regularized correction $r_{q,\lambda}$ with e_q before applying the same q^{-2} normalization and signed-log map.

Table 4: Matched ZINC-12K formulation ablations. Results are mean ± sample standard deviation.

Variant	Forest	q^{-2}	Ridge	Test MAE ↓
RawMagnitude	×	×	✓	0.0648 ± 0.0026
ResidualCommon	✓	×	✓	0.0634 ± 0.0039
DirectDefect	✓	✓	×	0.0610 ± 0.0035
MAGRESPE	✓	✓	✓	$\mathbf{0.0600 \pm 0.0021}$

The matched formulation ablations show that forest subtraction reduces mean test MAE by 0.0014, while q^{-2} normalization yields a further reduction of 0.0034. The ridge correction additionally reduces test MAE from 0.0610 to 0.0600 MAE. These results provide evidence for forest subtraction and onset normalization, while the larger observed mean change is associated with the q^{-2} onset normalization. Additional ablations are reported in Appendix C.

7 CONCLUSION

We introduced MAGRESPE, a deterministic multi-scale node encoding based on the ridge-regularized correction to an exact forest magnitude reference. An exact BFS-shell identity characterizes the forest-prediction defect in terms of same-shell edges and excess shortest-path parents, while our girth theorem establishes the first nonzero order of the ridge-regularized correction. We established connected graph families with identical finite-horizon RRWP rooted row-profile summaries but distinct MAGRESPE residuals. Empirically, RRWP+MAGRESPE improves mean performance across the evaluated real-world benchmarks, while the standalone and formulation ablations show which components matter and where MAGRESPE falls short. Although MAGRESPE is not a complete graph invariant, its main practical limitation is the cost of all-pairs shortest paths and dense ridge solves, motivating efficient approximations.

Reproducibility Statement

We provide deterministic theory certificates, experimental configurations, experimental results, and computational analysis. All reported results use checkpoints selected on validation performance, with a single test evaluation per run. Appendices B–G document the complete training configurations, model integration, per-seed results, structural certificates, and computational analyses.

AI Use Statement

Generative AI tools were used to assist with literature search and critical review, checking of mathematical statements and proof presentation, development and verification of deterministic certificate code, and language editing and organization. All AI-assisted mathematical claims, proofs, code, experimental interpretations, and written content were independently reviewed and verified by the authors, who retain full responsibility for the manuscript.

References

Louis Airale, Antonio Longa, Mattia Rigon, Andrea Passerini, and Roberto Passerone. Simple path structural encoding for graph transformers. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 857–873, 2025.

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

Juan P Garcia Amboage, Ernst Röell, Patrick Schnider, and Bastian Rieck. LEAP: Local ECT-based learnable positional encodings for graphs. In *The Fourteenth International Conference on Learning Representations*, 2026.

Yasuhiko Asao, Yasuaki Hiraoka, and Shu Kanazawa. Girth, magnitude homology and phase transition of diagonality. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 154(1):221–247, 2024.

Linus Bao, Emily Jin, Michael Bronstein, İsmail İlkan Ceylan, and Matthias Lanzinger. Homomorphism counts as structural encodings for graph learning. In *International Conference on Learning Representations*, 2025.

Mitchell Black, Zhengchao Wan, Gal Mishne, Amir Nayyeri, and Yusu Wang. Comparing graph transformers via positional encodings. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 4103–4139, 2024.

Semih Cantürk, Renming Liu, Olivier Lapointe-Gagné, Vincent Létourneau, Guy Wolf, Dominique Beaini, and Ladislav Rampášek. Graph positional and structural encoder. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5533–5566, 2024.

Yun Young Choi, Sun Woo Park, Minho Lee, and Youngho Woo. Topology-informed graph transformer. In Sharvaree Vadgama, Erik Bekkers, Alison Pouplin, Sekou-Oumar Kaba, Robin Walters, Hannah Lawrence, Tegan Emerson, Henry Kvinge, Jakub Tomczak, and Stephanie Jegelka (eds.), *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pp. 20–34. PMLR, 29 Jul 2024.

Vijay Prakash Dwivedi, Ladislav Rampášek, Mikhail Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long range graph benchmark. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 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.

Jinjia Feng, Danlin Li, Zhen Wang, and Zhewei Wei. Graph context-aware positional encoding for graph representation learning. *Frontiers of Computer Science*, 2026.

James Flora, Mitchell Black, Weng-Keen Wong, and Amir Nayyeri. Understanding truncated positional encodings for graph neural networks. In *International Conference on Machine Learning*, 2026.

Florian Grötschla, Jiaqing Xie, and Roger Wattenhofer. Benchmarking positional encodings for GNNs and graph transformers. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2663–2674, 2026.

Yinan Huang, William Lu, Joshua Robinson, Yu Yang, Muhan Zhang, Stefanie Jegelka, and Pan Li. On the stability of expressive positional encodings for graphs. In *International Conference on Learning Representations*, 2024.

Steve Huntsman, Jewell B. Thomas, and Cynthia Ukawu. Scalably computing metric magnitude. In *Proceedings of the 2nd Conference on Topology, Algebra, and Geometry in Data Science*, volume 334 of *Proceedings of Machine Learning Research*, pp. 4–24, 2026.

Charilaos I. Kanatsoulis, Evelyn Choi, Stefanie Jegelka, Jure Leskovec, and Alejandro Ribeiro. Learning efficient positional encodings with graph neural networks. In *International Conference on Learning Representations*, 2025.

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*, volume 34, pp. 21618–21629, 2021.

Yoon Hyeok Lee, Jaemin Park, Taejin Paik, Doyun Kim, and Bosun Hwang. From theory to practice: Rethinking Green and Martin kernels for unleashing graph transformers. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 33688–33714, 2025.

Tom Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013.

Tom Leinster. The magnitude of a graph. *Mathematical Proceedings of the Cambridge Philosophical Society*, 166:247–264, 2019.

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations*, 2023.

Katharina Limbeck, Lydia Mezrag, Guy Wolf, and Bastian Rieck. Geometry-aware edge pooling for graph neural networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H. S. Torr, and Ser-Nam Lim. Graph inductive biases in transformers without message passing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23321–23337, 2023.

Nhat Khang Ngo, Truong Son Hy, and Risi Kondor. Multiresolution graph transformers and wavelet positional encoding for learning long-range and hierarchical structures. *The Journal of Chemical Physics*, 159(3), July 2023.

Long D Nguyen, Quang H Nguyen, Quang H Trinh, and Binh P Nguyen. From SMILES to enhanced molecular property prediction: A unified multimodal framework with predicted 3D conformers and contrastive learning techniques. *Journal of Chemical Information and Modeling*, 64(24): 9173–9195, 2024.

Evelien Otte and Ronald Rousseau. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6):441–453, December 2002.

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*, volume 35, 2022.

Yogesh Verma, Amauri H. Souza, and Vikas K. Garg. Positional encoding meets persistent homology on graphs. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 61323–61343, 2025.

Yu Wang, Nedim Lipka, Ryan A Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. Knowledge graph prompting for multi-document question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 19206–19214, 2024.

Jeyoon Yeom, Taero Kim, Rakwoo Chang, and Kyungwoo Song. Structural and positional ensembled encoding for graph transformer. *Pattern Recognition Letters*, 183:104–110, 2024.

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*, volume 34, 2021.

A Proofs and Additional Structural Theory

This appendix provides proofs for the structural results stated in Section 5, together with additional properties of the forest-normalized residual that support the main analysis.

A.1 Basic properties of the magnitude residual

Lemma 3. *Let Z be symmetric. If $Zw = \mathbf{1}$ and $Zw' = \mathbf{1}$, then $\mathbf{1}^\top w = \mathbf{1}^\top w'$.*

Proof. Let $x = w - w'$. Then $Zx = 0$. Since $Zw = \mathbf{1}$ and Z is symmetric, then $\mathbf{1}^\top x = w^\top Zx = 0$. Hence $\mathbf{1}^\top w = \mathbf{1}^\top w'$. □

Proposition 4. *For every forest and every $q \in (0, 1)$,*

$$(Z_q^{-1})_{vv} = \frac{1 + (d_v - 1)q^2}{1 - q^2}, \qquad (Z_q^{-1})_{uv} = -\frac{q}{1 - q^2} \quad \text{if } \{u, v\} \in E, \tag{14}$$

with all other off-diagonal entries zero. Consequently, $Z_q^{-1}\mathbf{1} = b_q$.

Proof. It suffices to work on one tree component. Let T denote the matrix defined by Eq. (14). For fixed u, v, only v and its neighbors contribute to $(Z_q T)_{uv}$.

If $u = v$, the numerator is $1+(d_v-1)q^2-d_vq^2 = 1-q^2$. If $u \neq v$, let $d = D(u, v)$. The unique path from v to u starts through one neighbor x satisfying $D(u, x) = d-1$, while every other neighbor y of v satisfies $D(u, y) = d+1$. Hence the numerator is $q^d[1+(d_v-1)q^2]-q[q^{d-1}+(d_v-1)q^{d+1}] = 0$. Therefore $Z_qT = I$. Finally,

$$(T\mathbf{1})_v = \frac{1 + (d_v - 1)q^2 - d_vq}{1 - q^2} = \frac{1 - (d_v - 1)q}{1 + q} = b_q(v).$$

□

Proposition 5. *For every $\lambda > 0$, the optimization problem in Eq. (5) has a unique minimizer. Under a node relabeling represented by a permutation matrix Π, $r'_{q,\lambda} = \Pi r_{q,\lambda}$. Consequently, the node-wise multi-scale* MagResPE *encoding is permutation equivariant.*

Proof. The Hessian of Eq. (5) is $2(Z_q^\top Z_q + \lambda I)$. For every nonzero vector x,

$$x^\top(Z_q^\top Z_q + \lambda I)x = \|Z_qx\|_2^2 + \lambda \|x\|_2^2 > 0,$$

so the objective is strictly convex and Eq. (6) is its unique minimizer.

Under relabeling,

$$Z'_q = \Pi Z_q \Pi^\top, \qquad b'_q = \Pi b_q, \qquad e'_q = \Pi e_q.$$

Substitution into Eq. (6) and $\Pi^\top\Pi = I$ yield $r'_{q,\lambda} = \Pi r_{q,\lambda}$. □

Proposition 6. *If G is a forest, then $r_{q,\lambda} = 0$ for every $q \in (0,1)$ and $\lambda > 0$. If Z_q is invertible, then*

$$\lim_{\lambda \to 0^+} r_{q,\lambda} = Z_q^{-1}\mathbf{1} - b_q. \tag{15}$$

Proof. For a forest, Proposition 4 gives $Z_q b_q = \mathbf{1}$, hence $e_q = 0$ and the unique ridge solution is zero.

If Z_q is invertible, then $Z_q^\top Z_q$ is positive definite and continuity of matrix inversion gives

$$(Z_q^\top Z_q + \lambda I)^{-1} Z_q^\top \longrightarrow Z_q^{-1} \qquad (\lambda \to 0^+).$$

Applying this operator to $e_q = \mathbf{1} - Z_q b_q$ proves Eq. (15). □

A.2 Second-order obstruction

Let

$$B_2 = \mathbf{1}[D(u,v) = 2], \qquad d = A\mathbf{1}, \qquad n_2 = B_2\mathbf{1},$$

and let T_v be the number of triangles incident to v and $\gamma_{vx} = |N(v) \cap N(x)|$.

Proposition 7. *For every fixed graph and fixed $\lambda \geq 0$,*

$$r_{q,\lambda} = \frac{q^2}{1+\lambda} c + O(q^3), \qquad c = Ad - d - n_2, \tag{16}$$

as $q \to 0$. Moreover,

$$c_v = 2T_v + \sum_{x : D(v,x)=2} (\gamma_{vx} - 1) \geq 0. \tag{17}$$

For $\lambda = 0$, the statement is understood locally around $q = 0$, where Z_q is invertible.

Proof. Using the distance layers,

$$Z_q = I + qA + q^2 B_2 + O(q^3).$$

Since $Z_0 = I$, the kernel is invertible for sufficiently small q and

$$Z_q^{-1}\mathbf{1} = \mathbf{1} - qd + q^2(Ad - n_2) + O(q^3).$$

The forest reference satisfies $b_q = \mathbf{1} - qd + q^2 d + O(q^3)$ so the unregularized correction is $q^2(Ad - d - n_2) + O(q^3)$. Equivalently, $e_q = q^2 c + O(q^3)$.

For fixed $\lambda > 0$,

$$(Z_q^\top Z_q + \lambda I)^{-1} = \frac{1}{1+\lambda} I + O(q), \qquad Z_q^\top = I + O(q),$$

which yields Eq. (16).

For the combinatorial interpretation, $(Ad)_v$ counts length-two walks starting at v. Exactly d_v return to v; walks ending at a neighbor contribute $2T_v$ in total; and each vertex x at distance two contributes γ_{vx} walks rather than the single unit subtracted by $(n_2)_v$. Therefore Eq. (17) follows. □

A.3 BFS-shell identity and proof of the forest–girth theorem

Lemma 8. *With the notation of Eq. (10), let R_v denote the eccentricity of v within its connected component. Then, for every $q \in (0,1)$,*

$$e_q(v) = \frac{1}{1+q} \sum_{j=1}^{R_v} a_j(v) q^{j+1}.$$

Consequently, the constant and linear coefficients vanish, and for every $s \geq 2$,

$$[q^s] e_q(v) = \sum_{j=1}^{s-1} (-1)^{s-1-j} a_j(v).$$

Moreover, $e_q(v) \geq 0$, with equality if and only if the connected component containing v is acyclic.

Proof. Fix a root v and write $n_j = |L_j(v)|$ and $s_j = \sum_{u \in L_j(v)} d_u$. The degree accounting across BFS shells gives

$$s_0 = n_1, \qquad s_j = n_j + n_{j+1} + a_j(v), \quad j \geq 1,$$

where shells beyond R_v are empty.

Define the finite generating polynomials

$$N_v(q) = \sum_{j \geq 0} n_j q^j, \qquad U_v(q) = \sum_{j \geq 0} s_j q^j.$$

Summing the shell-degree identities yields

$$qU_v(q) = (1+q)(N_v(q) - 1) + \sum_{j \geq 1} a_j(v) q^{j+1}.$$

The forest reference can be written as

$$b_q(u) = 1 - \frac{q}{1+q} d_u.$$

Consequently,

$$(Z_q b_q)(v) = N_v(q) - \frac{q}{1+q} U_v(q).$$

Substituting the preceding generating identity gives

$$e_q(v) = 1 - (Z_q b_q)(v) = \frac{1}{1+q} \sum_{j \geq 1} a_j(v) q^{j+1}.$$

Expanding $(1+q)^{-1}$ as a geometric series and collecting coefficients proves the all-orders identity.

Finally, $x_j(v)$ and $m_j(v)$ are nonnegative, so $a_j(v) \geq 0$. If the component containing v is a tree, every vertex has a unique BFS parent and there are no same-shell edges; hence every $a_j(v)$ vanishes.

Conversely, if the component contains a cycle, any BFS spanning tree leaves at least one non-tree edge. Such an edge either joins two vertices in the same shell or contributes an excess parent between consecutive shells. Thus some $a_j(v) > 0$, implying $e_q(v) > 0$ for every $q \in (0, 1)$. □

Proof of Theorem 1. Forest annihilation follows from Proposition 4. Suppose now that G contains a cycle, has girth g, and let $k = \left\lceil \frac{g}{2} \right\rceil$. If $m_j(v) > 0$, two shortest paths from v together with a same-shell edge contain a cycle of length at most $2j + 1$. If $x_j(v) > 0$, some vertex in $L_j(v)$ has at least two parents in $L_{j-1}(v)$, and the corresponding shortest paths contain a cycle of length at most $2j$.

Hence, for every $j \leq k - 2$,

$$x_j(v) = x_{j+1}(v) = m_j(v) = 0,$$

and therefore $a_j(v) = 0$.

Now let v lie on a shortest cycle C of length g. The cycle C is isometric: if two of its vertices admitted a path shorter than their shorter cycle arc of length $s \leq g/2$, the path and that arc would contain a cycle of length less than $2s \leq g$, contradicting the girth. If $g = 2k$, the vertex opposite v on C therefore lies in $L_k(v)$ and has two distinct predecessors in $L_{k-1}(v)$. Hence $x_k(v) > 0$. If $g = 2k - 1$, the two vertices at cyclic distance $k - 1$ from v in opposite directions are adjacent and both lie in $L_{k-1}(v)$. Hence $m_{k-1}(v) > 0$. In either case, $a_{k-1}(v) > 0$ for every vertex v of a shortest cycle.

Lemma 8 now gives

$$e_q = q^k a_{k-1} + O(q^{k+1}).$$

For fixed $\lambda > 0$,

$$(Z_q^\top Z_q + \lambda I)^{-1} Z_q^\top = \frac{1}{1+\lambda} I + O(q),$$

hence

$$r_{q,\lambda} = \frac{q^k}{1+\lambda} a_{k-1} + O(q^{k+1}),$$

which proves Eq. (11). □

Corollary 9. *Let $\phi(t) = \operatorname{sgn}(t)\log(1+|t|)$. Then*

$$p_v(q) \longrightarrow \phi\left(\frac{c_v}{1+\lambda}\right) \qquad (q \to 0). \tag{18}$$

Consequently, if all scales in a bank approach zero, the scale coordinates become asymptotically at most one-dimensional; if $c = 0$, they all converge to zero.

Proof. Divide Eq. (16) by q^2 and apply the continuous map ϕ coordinate-wise. □

A.4 Finite-scale distance-layer series

Let $B_j = \mathbf{1}[D(u,v) = j]$ and $Y_q = \sum_{j\geq 1} q^j B_j$ so that $Z_q = I + Y_q$. In particular, $B_1 = A$.

Relation to the classical magnitude series. The underlying alternating vertex-weight expansion is classical; see, e.g., Leinster (2019, Prop. 3.9). Our use of this expansion is different: we subtract the exact forest weighting and reorganize the resulting node-wise residual in terms of the distance-layer matrices $\{A_j\}$.

Theorem 10. *For $\lambda = 0$ and sufficiently small q, the exact residual is analytic and, for every node v,*

$$r_{q,0}(v) = \sum_{m\geq 2} q^m \left[\sum_{k=1}^{m} (-1)^k \sum_{\substack{r_1+\cdots+r_k=m \\ r_i\geq 1}} e_v^\top B_{r_1}\cdots B_{r_k}\mathbf{1} - (-1)^m d_v \right]. \tag{19}$$

Proof. For sufficiently small q, it holds that $(I+Y_q)^{-1} = \sum_{k\geq 0}(-1)^k Y_q^k$. Taking coordinate v and collecting all terms of total degree m gives the first term in brackets in Eq. (19). Moreover, one obtains $b_q(v) = 1 + d_v \sum_{m\geq 1}(-1)^m q^m$. Subtracting this forest series cancels the constant and linear terms. Each product $e_v^\top B_{r_1}\cdots B_{r_k}\mathbf{1}$ counts rooted walks whose consecutive endpoint pairs lie in prescribed distance layers, equivalently rooted homomorphisms from a distance-labelled path. □

Writing the coefficient of q^m in Eq. (19) as $h_m(v)$ gives

$$q^{-2} r_{q,0}(v) = c_v + \sum_{m\geq 3} q^{m-2} h_m(v). \tag{20}$$

Thus the second-order statistic is only the leading term of the local unregularized expansion. The theorem is local around $q = 0$ and does not by itself characterize arbitrary finite scales or the positive-ridge channels used in the experiments.

A.5 Positive-ridge local expansion

Proposition 11. *Fix a finite graph and $\lambda > 0$. Then $q \mapsto r_{q,\lambda}$ is real analytic in a neighborhood of $q = 0$. Consequently,*

$$r_{q,\lambda} = \sum_{m\geq 2} q^m h_m^{(\lambda)} = \frac{q^2}{1+\lambda} c + \sum_{m\geq 3} q^m h_m^{(\lambda)} \tag{21}$$

for sufficiently small q.

Proof. Because G is finite, $Z_q = I + \sum_{j\geq 1} q^j B_j$ contains only finitely many nonzero distance layers and is analytic near $q = 0$. The forest reference b_q and therefore $e_q = \mathbf{1} - Z_q b_q$ are analytic there as well.

Define $M_\lambda(q) = Z_q^\top Z_q + \lambda I$. Since $M_\lambda(0) = (1+\lambda)I$ is invertible, matrix inversion is analytic in a neighborhood of $q = 0$. Hence $r_{q,\lambda} = M_\lambda(q)^{-1} Z_q^\top e_q$ is analytic.

By Proposition 7, $e_q = q^2 c + O(q^3)$, while

$$M_\lambda(q)^{-1} = \frac{1}{1+\lambda} I + O(q), \qquad Z_q^\top = I + O(q).$$

Substitution yields the stated leading coefficient, and analyticity collects all remaining terms into $\{h_m^{(\lambda)}\}_{m\geq 3}$. □

Proposition 12. *For fixed Z and $\lambda > 0$, the linear map $e \longmapsto (Z^\top Z + \lambda I)^{-1} Z^\top e$ has operator norm at most $\frac{1}{2\sqrt{\lambda}}$.*

Proof. If σ is a singular value of Z, the corresponding singular value of the ridge response is $\frac{\sigma}{\sigma^2+\lambda}$. Its maximum over $\sigma \geq 0$ is attained at $\sigma = \sqrt{\lambda}$ and equals $1/(2\sqrt{\lambda})$. □

A.6 Regular-subdivision separation

Proof of Theorem 2. Let G_1 and G_2 be connected d-regular graphs on the same n vertices. Since each has $M = \frac{dn}{2}$ edges, uniform L-subdivision gives

$$|V(S_i)| = n + M(L-1) = n + \frac{dn}{2}(L-1), \qquad i \in \{1, 2\}.$$

Thus S_1 and S_2 have the same order.

For the proof only, assign each subdivision vertex an auxiliary type. Original vertices have type B, while an internal vertex at offset $s \in \{1, \ldots, L-1\}$ along a subdivided edge has type $I_{\min\{s, L-s\}}$. These types are not supplied to the model; they are used only to couple corresponding vertices in the two subdivisions. Their global multiplicities depend only on n, d, and L, and hence are identical in S_1 and S_2.

Because every internal subdivision vertex has degree two, every cycle in S_i contracts to a cycle of G_i, and every base cycle of length ℓ becomes a cycle of length ℓL. Therefore $g(S_i) = g_i L$. Under the hypothesis $2h < g_1 L$, both subdivided graphs satisfy $g(S_i) > 2h$. The universal cover of any connected d-regular graph is the infinite d-regular tree T_d. After uniform subdivision, both S_1 and S_2 therefore have the same typed universal cover $T_{d,L} = \mathrm{Sub}_L(T_d)$.

Fix a root $u_i \in S_i$. Distinct lifted root-to-target paths of length at most h cannot project to the same target: otherwise their union would contain a cycle of length at most $2h$, contradicting $g(S_i) > 2h$. Consequently, all vertices and walks relevant to $P_i^0, \ldots, P_i^h$ are represented faithfully in the common subdivided-tree cover.

When $g(S_i) = 2h + 1$, an edge may join two vertices in the depth-h shell of the induced radius-h subgraph. This does not affect the argument: reaching either endpoint already requires h steps, so traversing the additional edge requires step $h + 1$ and cannot contribute to $P_i^0, \ldots, P_i^h$.

Corresponding walks in $T_{d,L}$ also have identical probabilities. Indeed, every walk $(v_0, \ldots, v_k)$ has probability $\prod_{j=0}^{k-1} \frac{1}{\deg(v_j)}$ and the degree sequence is determined by the auxiliary types: $\deg(B) = d$ and $\deg(I_t) = 2$. Hence corresponding targets of the same type have identical horizon-h RRWP profiles.

Targets at graph distance greater than h have zero positive-power coordinates through horizon h. Since S_1 and S_2 have identical global type multiplicities and identical reachable type multiplicities from coupled roots, they also have identical numbers of such zero profiles. Thus roots of the same auxiliary type have identical RRWP row-profile multisets. Taking the multiset over all roots and then forgetting the auxiliary types gives $\mathcal{S}_h(S_1) = \mathcal{S}_h(S_2)$. The coupling also matches the diagonal profiles $\Psi_h(u_i, u_i)$ for roots of the same auxiliary type.

It remains to show the MagResPE separation. Since $g(S_i) = g_i L$, Theorem 1 gives the first nonzero residual orders $k_i = \left\lceil \frac{g_i L}{2} \right\rceil$. Because $g_2 \geq g_1 + 1$ and $L \geq 2$, then $k_2 \geq k_1 + 1$. Moreover, the leading coefficient at order k_i is nonnegative and nonzero on vertices of a shortest cycle. Hence, for some $c_1 > 0$,

$$\|r_{q,\lambda}^{(1)}\|_\infty = c_1 q^{k_1} + O(q^{k_1+1}),$$

whereas

$$\|r_{q,\lambda}^{(2)}\|_\infty = O(q^{k_2}) = o(q^{k_1}).$$

Therefore, for every fixed $\lambda > 0$ and all sufficiently small $q > 0$, we obtain $\|r_{q,\lambda}^{(1)}\|_\infty > \|r_{q,\lambda}^{(2)}\|_\infty$, so their residual multisets differ. The fixed-q signed-log channel map is injective, and therefore the corresponding MagResPE channel multisets also differ. This argument establishes separation for

each fixed pair, without asserting that a fixed finite scale bank separates the entire arbitrary-horizon subdivision family. □

Family-wide consequence. Fix n, d, L, h and let $\mathcal{G}$ be the class of connected d-regular graphs on n vertices satisfying $g(G)L > 2h$. The proof above shows that

$$\mathcal{S}_h(\mathrm{Sub}_L(G)) = \mathcal{S}_h(\mathrm{Sub}_L(G')) \qquad \text{for all } G, G' \in \mathcal{G}.$$

Thus, whenever $\mathcal{G}$ contains non-isomorphic members, their uniform subdivisions form a common finite-horizon RRWP row-profile-summary equivalence class.

Interpretation. Uniform subdivision pushes cycle closure beyond any prescribed finite walk horizon while preserving the local branching pattern. The resulting graphs therefore share the same rooted RRWP row-profile summaries, whereas MAGRESPE retains differences in their girth-controlled onset orders.

Scope. The theorem establishes equality of permutation-invariant RRWP row-profile summaries, not equality of the complete vertex-indexed RRWP tensors. It consequently makes no claim about the indistinguishability of these graphs by arbitrary RRWP-equipped Graph Transformers.

Corollary 13. *Let H be the 14-vertex cubic Heawood graph of girth 6, and let X_{14} be the fixed connected cubic graph of girth 5 defined in Appendix D.2. For any finite horizon h, choose $L \geq 2$ such that $2h < 5L$ and define $G_h = \mathrm{Sub}_L(H)$ and $H_h = \mathrm{Sub}_L(X_{14})$. Then G_h and H_h are connected graphs of the same order and satisfy $\mathcal{S}_h(G_h) = \mathcal{S}_h(H_h)$. Moreover, the two graphs have girths $6L$ and $5L$, respectively, so their* MAGRESPE *onset orders are $3L$ and $\lceil \frac{5L}{2} \rceil$. Since $\lceil \frac{5L}{2} \rceil < 3L$, their* MAGRESPE *residual multisets differ for every fixed $\lambda > 0$ and all sufficiently small positive q.*

Proof. Apply Theorem 2 with $d = 3$, $n = 14$, $g(H) = 6$, and $g(X_{14}) = 5$. □

A.7 SHARP CYCLE-FAMILY SEPARATION

Theorem 14. *Let $m \geq 3$. Let P_m and P_{2m} denote the random walk transition matrices of C_m and C_{2m}, respectively, and let $\pi : V(C_{2m}) \to V(C_m)$ be the natural two-fold covering map. Fix a root $u \in V(C_m)$ and a lift $\widetilde{u} \in V(C_{2m})$ satisfying $\pi(\widetilde{u}) = u$. For every $v \in V(C_m)$, write $\pi^{-1}(v) = \{\widetilde{v}_1, \widetilde{v}_2\}$. Then, for every $k \geq 0$, $(P_m^k)_{uv} = (P_{2m}^k)_{\widetilde{u}\widetilde{v}_1} + (P_{2m}^k)_{\widetilde{u}\widetilde{v}_2}$. Thus the RRWP profile of (u, v) in C_m is the coordinate-wise sum of the profiles of $(\widetilde{u}, \widetilde{v}_1)$ and $(\widetilde{u}, \widetilde{v}_2)$ in C_{2m}.*

Let u' denote the copy of u in one component of $2C_m = C_m \sqcup C_m$. Then the rooted horizon-h RRWP row-profile multisets of C_{2m} rooted at $\widetilde{u}$ and $2C_m$ rooted at u' are equal if and only if $h < h_\star(m) := \lceil \frac{m}{2} \rceil$. Hence $h_\star(m)$ is the exact first RRWP horizon at which C_{2m} and $2C_m$ become distinguishable. At $h = h_\star(m)$, the newly folded contribution from the second reachable lift has probability exactly $2^{-h_\star(m)}$.

Proof. We first prove the folding identity. Because $\pi : C_{2m} \to C_m$ is a graph covering, every length-k walk in C_m starting at u has a unique lift to a length-k walk in C_{2m} starting at the fixed lift $\widetilde{u}$.

Fix $v \in V(C_m)$. A lifted walk whose projection ends at v must end at exactly one of the two vertices $\pi^{-1}(v) = \{\widetilde{v}_1, \widetilde{v}_2\}$. Conversely, every walk in C_{2m} from $\widetilde{u}$ to either $\widetilde{v}_1$ or $\widetilde{v}_2$ projects to a walk in C_m from u to v. These two sets of lifted walks are disjoint.

Both cycles are 2-regular, so every length-k walk has probability 2^{-k}. Summing the probabilities of the two disjoint classes of lifted walks therefore gives $(P_m^k)_{uv} = (P_{2m}^k)_{\widetilde{u}\widetilde{v}_1} + (P_{2m}^k)_{\widetilde{u}\widetilde{v}_2}$.

Applying this equality for $k = 0, \ldots, h$ gives the corresponding coordinate-wise folding identity for the RRWP profiles defined in Section 3.2.

We next establish the collision below the threshold. For every $v \in V(C_m)$, the two vertices $\widetilde{v}_1$ and $\widetilde{v}_2$ are antipodal in C_{2m}. Consequently, $D(\widetilde{u}, \widetilde{v}_1) + D(\widetilde{u}, \widetilde{v}_2) = m$. Suppose $h < \lceil \frac{m}{2} \rceil$. Since h is an integer, this is equivalent to $2h < m$. Therefore the two vertices in a fiber cannot both lie within distance h of $\widetilde{u}$.

For each $v \in V(C_m)$, at least one vertex in $\pi^{-1}(v)$ thus has distance greater than h from $\widetilde{u}$. Assume without loss of generality that this vertex is $\widetilde{v}_2$. Then no walk of length at most h can reach $\widetilde{v}_2$, so $(P_{2m}^k)_{\widetilde{u}\widetilde{v}_2} = 0$ and $0 \le k \le h$. The folding identity then shows that the RRWP profile of (u, v) in C_m is identical to the profile of $(\widetilde{u}, \widetilde{v}_1)$ in C_{2m}, while $\widetilde{v}_2$ contributes the all-zero profile.

In $2C_m$, the copy of v in the component containing u' has the same profile as (u, v) in C_m, whereas the copy of v in the other component is unreachable from u' and therefore contributes the all-zero profile. Hence every covering fiber of C_{2m} contributes exactly the same pair of profiles as the corresponding two copies of v in $2C_m$. Taking the multiset union over all $v \in V(C_m)$ gives equality of the rooted horizon-h row-profile multisets.

It remains to determine the first horizon at which equality fails.

First suppose that $m = 2a$. Then $h_\star(m) = a$. Let v be the vertex antipodal to u in C_m. There are exactly two shortest paths of length a from u to v, one in each direction around the cycle. Therefore $(P_m^a)_{uv} = 2 \cdot 2^{-a} = 2^{1-a}$, while $(P_m^k)_{uv} = 0$ with $k < a$.

The two vertices $\widetilde{v}_1, \widetilde{v}_2 \in \pi^{-1}(v)$ are both at distance a from $\widetilde{u}$ in $C_{2m} = C_{4a}$. Each is connected to $\widetilde{u}$ by a unique shortest path of length a, and hence $(P_{2m}^a)_{\widetilde{u}\widetilde{v}_i} = 2^{-a}$ with $i \in \{1, 2\}$. Any target in C_{2m} having zero coordinates before time a and a nonzero coordinate at time a must lie at distance exactly a from $\widetilde{u}$. The only such targets are $\widetilde{v}_1$ and $\widetilde{v}_2$, whose ath coordinates are 2^{-a} rather than 2^{1-a}. Thus the RRWP profile of (u, v) in C_m is absent from the rooted row-profile multiset of C_{2m}, and the two multisets differ at horizon a.

Now suppose that $m = 2a + 1$. Then $h_\star(m) = a + 1$. Choose $v \in V(C_m)$ with $D(u, v) = a$. The two simple paths from u to v around C_m have lengths a and $a + 1$. Consequently, we obtain $(P_m^a)_{uv} = 2^{-a}$, $(P_m^{a+1})_{uv} = 2^{-(a+1)}$ and $(P_m^k)_{uv} = 0$ with $k < a$.

The two vertices in $\pi^{-1}(v) = \{\widetilde{v}_1, \widetilde{v}_2\}$ lie at distances a and $a + 1$ from $\widetilde{u}$, respectively. Thus the two nonzero coordinates above arise from different vertices of the same covering fiber.

Suppose that some $w \in V(C_{2m})$ had the same horizon-$(a + 1)$ RRWP profile as (u, v) in C_m. Since this profile vanishes before time a and is nonzero at time a, necessarily $D(\widetilde{u}, w) = a$. However, $C_{2m} = C_{4a+2}$ is bipartite. A vertex at distance a from $\widetilde{u}$ can therefore be reached only at walk lengths having the same parity as a. Hence $(P_{2m}^{a+1})_{\widetilde{u}w} = 0$, contradicting the nonzero $(a + 1)$st coordinate of the profile of (u, v) in C_m. Thus this profile is absent from the rooted row-profile multiset of C_{2m}, and the two multisets differ at horizon $a + 1$.

We have shown that the rooted row-profile multisets agree for every $h < h_\star(m)$ and differ at $h = h_\star(m)$. They cannot become equal again at a later horizon: if equality held for some $h > h_\star(m)$, truncating every horizon-h profile to its first $h_\star(m) + 1$ coordinates would imply equality at $h_\star(m)$, contradicting the result above. Therefore equality holds if and only if $h < h_\star(m)$.

Finally, consider the newly folded probability contribution at the first separating horizon. If $m = 2a$, both lifts of the antipodal vertex first contribute at horizon $a = h_\star(m)$, each with probability 2^{-a}. If $m = 2a+1$, the farther lift first contributes at horizon $a+1 = h_\star(m)$ with probability $2^{-(a+1)}$. Hence in both cases the newly folded contribution at the first separating horizon is exactly $2^{-h_\star(m)}$. □

Diagonal RRWP consequence. The diagonal RRWP profiles remain indistinguishable beyond the pairwise separation threshold. For the root $u \in V(C_m)$ and its chosen lift $\widetilde{u} \in V(C_{2m})$,

$$(P_m^k)_{uu} = (P_{2m}^k)_{\widetilde{u}\widetilde{u}}, \qquad 0 \le k < m.$$

Indeed, before m steps, a closed walk in C_m cannot wrap around the cycle, so the return walks are the same as those in C_{2m}. By vertex transitivity, the same equality holds for every corresponding pair of vertices. Hence the diagonal return-probability profiles of C_{2m} and $2C_m$ are identical through every horizon $h < m$. Since

$$h_\star(m) = \left\lceil \frac{m}{2} \right\rceil < m \qquad (m \ge 3),$$

the first separation at $h_\star(m)$ occurs in the off-diagonal pairwise RRWP profiles rather than in the diagonal return-probability profiles.

For a cycle C_L, $L \geq 3$, and any $v \in V(C_L)$, define

$$S_L(q) = \sum_{u \in V(C_L)} q^{D(v,u)}, \qquad \beta_q = \frac{1-q}{1+q}, \tag{22}$$

where $q \in (0,1)$. By vertex transitivity, $S_L(q)$ is independent of v, and β_q is the degree-two forest reference. Consequently, $Z_q \mathbf{1} = S_L(q)\mathbf{1}$ and $e_q = (1 - \beta_q S_L(q))\mathbf{1}$. Since Z_q is symmetric, the unique ridge correction is constant across vertices, with value

$$\rho_L(q,\lambda) = \frac{S_L(q)(1-\beta_q S_L(q))}{S_L(q)^2 + \lambda}. \tag{23}$$

The same value applies to every vertex of a disjoint union of identical L-cycles.

For fixed q and λ, define

$$F_{q,\lambda}(S) = \frac{S(1-\beta_q S)}{S^2 + \lambda}, \tag{24}$$

so that $\rho_L(q,\lambda) = F_{q,\lambda}(S_L(q))$.

Theorem 15. *For every $m \geq 3$, $q \in (0,1)$, and $0 < \lambda < 1$,*

$$\rho_m(q,\lambda) > \rho_{2m}(q,\lambda). \tag{25}$$

Hence, a single MAGRESPE *scale separates C_{2m} from $2C_m$, including at every horizon where their truncated RRWP rooted row-profile summaries coincide (Theorem 14).*

Proof. For $S \geq 1$,

$$F'_{q,\lambda}(S) = \frac{\lambda(1 - 2\beta_q S) - S^2}{(S^2+\lambda)^2} < 0,$$

since $\beta_q > 0$ and $0 < \lambda < 1$. Moreover, the cycle distance-shell counts give

$$S_m(q) < S_{2m}(q) \qquad (q \in (0,1)).$$

Therefore,

$$\rho_m(q,\lambda) = F_{q,\lambda}(S_m(q)) > F_{q,\lambda}(S_{2m}(q)) = \rho_{2m}(q,\lambda).$$

Multiplication by $q^{-2} > 0$ and the strictly increasing signed-log map preserve this inequality, establishing separation by a single MAGRESPE channel. □

Corollary 16. *Let*

$$S_\infty(q) = \frac{1+q}{1-q}.$$

For every $m \geq 3$, $q \in (0,1)$, and $0 < \lambda < 1$,

$$\rho_m(q,\lambda) - \rho_{2m}(q,\lambda) \geq \frac{1-\lambda}{(S_\infty(q)^2 + \lambda)^2} q^{\lceil m/2 \rceil} > 0. \tag{26}$$

Proof. For every cycle C_L,

$$1 \leq S_L(q) < S_\infty(q).$$

Furthermore, for $1 \leq S \leq S_\infty(q)$,

$$-F'_{q,\lambda}(S) = \frac{S^2 - \lambda + 2\lambda\beta_q S}{(S^2+\lambda)^2} \geq \frac{1-\lambda}{(S_\infty(q)^2+\lambda)^2}.$$

It remains to bound the difference between the cycle row sums. If $m = 2a$, then

$$S_{2m}(q) - S_m(q) = q^a + 2\sum_{j=a+1}^{2a-1} q^j + q^{2a} \geq q^a.$$

If $m = 2a+1$, then

$$S_{2m}(q) - S_m(q) = 2\sum_{j=a+1}^{2a} q^j + q^{2a+1} \geq q^{a+1}.$$

Thus, in both cases,

$$S_{2m}(q) - S_m(q) \geq q^{\lceil m/2 \rceil}.$$

Applying the mean-value theorem to $F_{q,\lambda}$ on $[S_m(q), S_{2m}(q)]$ gives

$$\rho_m(q,\lambda) - \rho_{2m}(q,\lambda) \geq \frac{1-\lambda}{(S_\infty(q)^2+\lambda)^2} q^{\lceil m/2 \rceil},$$

as claimed. □

Scope of the finite-scale witness. For any cubic graph, the forest reference satisfies $b_{1/2} = 0$. Consequently, $e_{1/2} = \mathbf{1}$ and the regularized correction coincides with the raw ridge magnitude weighting at this scale. Complementary certificates at the scales $q = 1/3$ and $q = 1/2$ are provided in Appendix D.2.

A.8 Additional separations, stability, and ceilings

Proposition 17. *With constant initial node features, 1-WL does not distinguish C_6 from $2C_3$. At $q = 1/2$ and $\lambda = 0$, however, their* MAGRESPE *residual multisets are $\{1/21\}^{\times 6}$ and $\{1/6\}^{\times 6}$, respectively. The separation persists for sufficiently small positive λ.*

Proof. Both graphs are 2-regular on six vertices, so constant-feature 1-WL assigns the same color multiset at every iteration. By vertex transitivity, the exact magnitude weighting is the reciprocal of the kernel row sum. At $q = 1/2$, these row sums are $21/8$ for C_6 and 2 inside each triangle. Subtracting the common forest reference $1/3$ gives $1/21$ and $1/6$, respectively. Continuity in λ yields the positive-ridge statement. □

Proposition 18. *Let G_1 and G_2 be connected d-regular graphs on the same number of vertices. Suppose there exist nonnegative integers $n_0, \ldots, n_D$ such that, for every vertex v in either graph and every $j \in \{0, \ldots, D\}$: $\big|\{u : D(v,u) = j\}\big| = n_j$. Thus every vertex in both graphs has the same distance-shell histogram. Then their* MAGRESPE *residuals are the same constant node vector for every $q \in (0,1)$ and $\lambda > 0$. In particular, $K_{3,3}$ and the triangular prism have identical* MAGRESPE *residuals at every scale and ridge parameter despite having different girths.*

Proof. Since both graphs are d-regular, their forest references are constant:

$$b_q = \beta_{q,d}\mathbf{1}, \qquad \beta_{q,d} = \frac{1-(d-1)q}{1+q}.$$

The common distance-shell histogram implies that every row of either magnitude kernel has the same sum $s_q = \sum_{j=0}^{D} n_j q^j$. Therefore, in both graphs,

$$Z_q\mathbf{1} = s_q\mathbf{1}, \qquad e_q = (1 - \beta_{q,d}s_q)\mathbf{1}.$$

Because Z_q is symmetric, $\mathbf{1}$ is also an eigenvector of $Z_q^\top Z_q$ with eigenvalue s_q^2. Consequently, the ridge residual is $r_{q,\lambda} = \frac{s_q(1-\beta_{q,d}s_q)}{s_q^2+\lambda}\mathbf{1}$. This expression depends only on the common degree and distance-shell histogram. Both $K_{3,3}$ and the triangular prism are 3-regular graphs on six vertices, and every vertex in either graph has the distance-shell histogram $(1,3,2)$. The proposition therefore applies to this pair. □

Scope of the finite-RRWP results. The separation results above concern the finite horizon supplied to the model. They do not imply that MAGRESPE contains information absent from an unrestricted or complete random walk representation. This distinction is consistent with prior work showing that the expressive relationships between positional encodings can change under truncation (Black et al., 2024; Flora et al., 2026).

B EXPERIMENTAL DETAILS AND IMPLEMENTATION

This appendix gives the complete predictive configurations and the implementation details omitted from the main text. Unless stated otherwise, checkpoints are selected using validation performance and the test split is evaluated once at the selected checkpoint. The deterministic preprocessing that produces the MAGRESPE signature contains no learned parameters.

B.1 DATASETS AND EXPERIMENTAL CONFIGURATIONS

B.1.1 REAL-WORLD BENCHMARKS

Table 6 summarizes the predictive configurations for ZINC-12K, Peptides-struct, and PascalVOC-SP. All datasets use the fixed MAGRESPE encoding configuration, with task-specific model architectures and optimization settings.

Table 5: Real-world benchmark descriptions.

Configuration	ZINC-12K	Peptides-struct	PascalVOC-SP
Graphs	12,000	15,535	11,355
Split	10,000/1,000/1,000	10,873/2,331/2,331	8,498/1,428/1,429
Task	graph regression	graph regression	node classification

Table 6: Real-world benchmark configurations. `ksteps` includes the identity matrix, so the corresponding RRWP horizon $h = \texttt{ksteps} - 1$. d_p denotes MAGRESPE projection width. MAE denotes mean absolute error.

Configuration	ZINC-12K	Peptides-struct	PascalVOC-SP
Metric	MAE ↓	MAE ↓	macro-F1 ↑
RRWP `ksteps`	21	24	24
GRIT width	56	96	96
Layers / heads	10 / 8	4 / 8	4 / 8
MAGRESPE d_p	28	28	28
Batch	32	16	$4 \times$ accum. 4
Optimizer	AdamW	AdamW	AdamW
Learning rate	10^{-3}	3×10^{-4}	3×10^{-4}
Weight decay	10^{-5}	0	0
Warmup	50 epochs	10 epochs	10 epochs
Training epochs	1,500 epochs	200 epochs	200 epochs
Seeds	0–4	0–4	0–4
RRWP params.	364,169	444,971	439,029
RRWP+MAGRESPE params.	363,849	440,367	438,877

B.1.2 THEORY-ALIGNED SYNTHETIC DATASETS

We construct two balanced binary classification tasks to evaluate the finite-horizon separation results of Section 5.3. Both use constant original node features and controlled structural statistics, with separate training, validation, and test partitions defined by their underlying graph-generating units.

Synthetic dataset construction. We construct two balanced binary classification datasets for cycle family and connected regular-subdivision family.

Cycle family. For each $m \in \{41, \dots, 80\}$, we construct C_{2m} and $2C_m$, labelled as classes 0 and 1, respectively. The resulting 80 graphs are split by m, so both members of each pair belong to the same partition. We use $24/8/8$ values of m for training, validation, and testing, respectively, giving $48/16/16$ graphs. The validation values are $\mathcal{M}_{\text{val}} = \{42, 49, 51, 54, 68, 69, 78, 79\}$, and the test values are $\mathcal{M}_{\text{test}} = \{46, 48, 52, 57, 61, 67, 71, 76\}$; the remaining values form the training partition.

We evaluate RRWP at horizons $h \in \{20, 30, 40\}$. By Theorem 14, all 40 pairs are RRWP-indistinguishable at $h = 20$, whereas 20 and 40 pairs become distinguishable at $h = 30$ and $h = 40$,

respectively. At $h = 30$, each data partition contains equal numbers of RRWP-indistinguishable and distinguishable pairs.

Connected regular-subdivision family. We enumerate non-isomorphic connected cubic graphs on 22 vertices using nauty `geng`. The census contains $90{,}553$ graphs of girth 5 and 385 of girth 6. Within each class, we select the first 120 graphs. Using the same ordering, we assign the first 72 selected graphs to training, the next 24 to validation, and the remaining 24 to testing. This yields 240 distinct base isomorphism classes and $144/48/48$ train/validation/test graphs.

Every base edge is replaced by a path of $L = 9$ edges. Since each cubic base has 22 vertices and 33 edges, every subdivided graph has 286 vertices and 297 edges, including 22 degree-3 branch vertices and 264 degree-2 subdivision vertices. The two classes have girths 45 and 54, respectively, but identical degree and subdivision-type histograms. We use RRWP horizon $h = 20$, for which $2h = 40 < 45$. Thus all 240 graphs belong to the same horizon-20 RRWP row-profile equivalence class under Theorem 2. The split is disjoint at the base-isomorphism-class level.

Synthetic model and training protocol. Both tasks use the same reduced GRIT architecture and optimization configuration, summarized in Table 7. We compare NoPE, RRWP, standalone MAGRESPE, and RRWP+MAGRESPE. All original node features are constant, with no explicit degree, subdivision-type, graph-size, or class-dependent edge attributes supplied to the model.

NoPE and standalone MAGRESPE both use the plain GRIT without RRWP. The RRWP and hybrid conditions retain both RRWP routes. The synthetic MAGRESPE encoder uses $d_p = 16$, rather than the reference ZINC projection width $d_p = 28$; the eight-scale bank and $\lambda = 10^{-3}$ remain unchanged.

Table 7: Synthetic dataset and training configurations. Both tasks use identical GRIT architecture and optimization settings.

Configuration	Cycle family	Regular subdivisions
Split	48/16/16	144/48/48
Nodes	82–160	286
Task	graph classification	graph classification
Metric	Accuracy	Accuracy
RRWP `ksteps`	$21, 31, 41$	21
GRIT width	40	40
Layers / heads	4 / 8	4 / 8
MAGRESPE d_p	16	16
Batch size	10	10
Optimizer	AdamW	AdamW
Learning rate	10^{-3}	10^{-3}
Weight decay	10^{-5}	10^{-5}
Warmup	5 epochs	5 epochs
Training epochs	200 epochs	200 epochs
Seeds	0–4	0–4
RRWP params. (`ksteps`=21)	79,001	79,001
RRWP+MAGRESPE params. (`ksteps`=21)	79,129	79,129
RRWP params. (`ksteps`=31)	79,801	–
RRWP+MAGRESPE params. (`ksteps`=31)	79,929	–
RRWP params. (`ksteps`=41)	80,601	–
RRWP+MAGRESPE params. (`ksteps`=41)	80,729	–

Validation accuracy is evaluated every epoch, and the checkpoint with the highest validation accuracy is selected; ties retain the earliest checkpoint. Each test split is evaluated exactly once at the selected checkpoint. Consequently, the cycle and connected experiments comprise 40 and 20 learned runs, respectively.

B.2 BASELINES AND PARAMETER BUDGETS

Baselines and controls. We evaluate MAGRESPE against several structural encodings using task-specific GRIT configurations and matched training protocols within each dataset. RRWP serves as

the reference baseline and uses both node-level return probabilities and node-pair transition profiles. SPSE replaces RRWP node and node-pair encoding with simple path structural encoding, while standalone MAGRESPE removes the RRWP encoding and uses the proposed node representation. The RRWP+MAGRESPE hybrid retains the original RRWP interface and adds the MAGRESPE node branch. We additionally consider RRWP combined with SignNet (Lim et al., 2023), a spectral-node augmentation using eight normalized Laplacian eigenvectors. RawMagnitude, ResidualCommon, and DirectDefect provide formulation-specific controls, as described in Appendix C.

Node integration and parameter budgets. The MAGRESPE encoder maps the deterministic signature $p_v \in \mathbb{R}^K$ to a learned representation of width d_p, which is concatenated with the ordinary dataset embedding of width $d - d_p$, as defined in Eq. (9). No additional projection is applied after concatenation.

We use approximately matched, task-specific parameter budgets for the RRWP baseline and RRWP+MAGRESPE hybrid rather than imposing a single architecture across all benchmarks. The corresponding trainable parameter counts are reported in Tables 6 and 7. The hybrid uses slightly fewer parameters than RRWP on all three real-world benchmarks. We follow the same parameter-matching principle for the remaining datasets and controls, with their task-specific configurations reported in Appendix B.

Published reference results. Where available, we also report published results from SPSE (Airale et al., 2025) and LRGB (Dwivedi et al., 2022) to provide broader benchmark context. These results are identified separately from our matched experiments because architectures, training settings, and evaluation protocols may differ. Only results obtained under our experimental protocol are used for paired comparisons.

B.3 PREPROCESSING ALGORITHM AND CACHING

Algorithm 1 summarizes the deterministic preprocessing used to construct the MAGRESPE node representation. The shortest-path metric is computed once per graph and reused across all scales. For disconnected vertex pairs, $D(u, v) = \infty$ and we use $q^{D(u,v)} = 0$. The v-th row of $\mathbf{S}_{\text{MAGRESPE}}$ is exactly the multi-scale node signature p_v defined in Eq. (8). Thus $\mathbf{S}_{\text{MAGRESPE}}$ stacks the deterministic MAGRESPE signatures of all nodes before neural training.

For the standard MAGRESPE configuration, $K = 8$. Once constructed, the representation is cached and reused across runs with the same graph topology and MAGRESPE configuration, so the shortest-path computation and ridge systems are not repeated during training. Cache identity is determined by the canonical graph topology together with the parameters defining the representation, including $\mathcal{Q}$ and λ. Computational cost is analyzed separately in Appendix E.

B.4 SCALE BANK AND RIDGE SETTING

Unless explicitly modified by the ablations in Appendix C, the standard MAGRESPE configuration uses the fixed scale bank

$$\mathcal{Q} = \{0.05, 0.08, 0.13, 0.21, 0.34, 0.52, 0.74, 0.95\}, \qquad \lambda = 10^{-3}. \tag{27}$$

The parameter q controls the metric range of the magnitude kernel $Z_q(u, v) = q^{D(u,v)}$: smaller values strongly attenuate distant shortest-path shells, whereas larger values retain progressively more long-range contributions. This scale dependence is intrinsic to graph magnitude (Leinster, 2019). We therefore use a bank spanning small, intermediate, and large q rather than a single metric scale.

This choice is also consistent with the analysis in Section 5. As $q \to 0$, the normalized channels collapse to the common leading statistic in Eq. (13), whereas finite positive scales can retain higher-order distance-layer contributions. A scale bank concentrated only near zero would therefore discard part of the multi-scale information that MAGRESPE is designed to expose.

The ridge guarantees a unique regularized correction while leaving the scale-dependent residual as the primary signal. Both $\mathcal{Q}$ and λ are kept unchanged across ZINC, Peptides-struct, and PascalVOC-SP. Residual-based scale ablations retain $\lambda = 10^{-3}$ and modify only the scale bank as specified in

Algorithm 1 MAGRESPE preprocessing

Require: Graph $G = (V, E)$, scale bank $\mathcal{Q} = \{q_1, \dots, q_K\}$, ridge $\lambda > 0$
Ensure: Signature matrix $\mathbf{S}_{\text{MAGRESPE}} \in \mathbb{R}^{|V| \times K}$

Canonicalize G as a loop-free undirected graph
Compute node degrees $\{d_v\}_{v \in V}$
Compute all-pairs shortest-path distances $D(u, v)$
for $i = 1, \dots, K$ **do**
  $q \leftarrow q_i$
  Construct the magnitude kernel

$$Z_q(u, v) \leftarrow q^{D(u,v)}$$

  Evaluate the forest reference

$$b_q(v) \leftarrow \frac{1 - (d_v - 1)q}{1 + q}$$

  Compute the forest-prediction defect

$$e_q \leftarrow \mathbf{1} - Z_q b_q$$

  Solve the ridge correction

$$r_{q,\lambda} \leftarrow (Z_q^\top Z_q + \lambda I)^{-1} Z_q^\top e_q$$

  for $v \in V$ **do**
    $(\mathbf{S}_{\text{MAGRESPE}})_{v,i} \leftarrow \operatorname{sgn}(r_{q,\lambda}(v)) \log\left(1 + \left|\frac{r_{q,\lambda}(v)}{q^2}\right|\right)$
  end for
end for
return $\mathbf{S}_{\text{MAGRESPE}}$

Appendix C. Because Eq. (5) uses an unnormalized fidelity term, a fixed λ does not imply identical relative shrinkage across graph sizes; its effect also depends on the spectrum of Z_q. We retain $\lambda = 10^{-3}$ unchanged across datasets, while size-adaptive or normalized regularization is left for future work.

C Additional Experimental Results and Ablations

C.1 Ablation variants

We use the standard MAGRESPE configuration from Appendix B.4 as the reference representation, with eight finite metric scales $\mathcal{Q} = \{0.05, 0.08, 0.13, 0.21, 0.34, 0.52, 0.74, 0.95\}$. The following variants modify either the scale coverage or the formulation while otherwise retaining the same node-encoding mechanism and matched host configuration.

Small-4. Small-4 retains four scales concentrated in the low-q regime: $\mathcal{Q}_{\text{small}} = \{0.05, 0.08, 0.13, 0.21\}$. It is designed as a theory-motivated negative control for Corollary 9: if all scales are concentrated near zero, the normalized channels increasingly approach the same leading second-order statistic. Small-4 therefore tests whether densely sampling only short metric ranges is sufficient.

Compact-4. Compact-4 uses the four-scale bank $\mathcal{Q}_{\text{compact}} = \{0.08, 0.21, 0.52, 0.95\}$, selected from the standard MAGRESPE bank $\mathcal{Q} = \{0.05, 0.08, 0.13, 0.21, 0.34, 0.52, 0.74, 0.95\}$ to retain coverage across the finite-q range while reducing the representation from eight channels to four. Unlike Small-4, which concentrates its scales near $q = 0$, Compact-4 preserves both smaller and larger metric scales and therefore tests whether the standard multi-scale representation can be compressed without collapsing toward the common small-q limit.

C2. C2 discards the finite-q scale bank and uses only the closed-form second-order coefficient

$$c_v = 2T_v + \sum_{x:D(v,x)=2} (\gamma_{vx} - 1), \tag{28}$$

from Eq. (12). It tests the stronger hypothesis that the universal leading small-q statistic alone is sufficient, without the higher-order distance-layer information retained at finite scales.

Zero-info. Zero-info retains the RRWP interface and the trainable auxiliary node-encoder architecture of the RRWP+MAGRESPE hybrid, but replaces the eight MAGRESPE input channels with constant zeros. It therefore controls for the additional trainable node branch without supplying graph-dependent information through that branch.

Standalone and hybrid MAGRESPE. Standalone MAGRESPE uses the standard node representation while removing the RRWP encoding. The RRWP+MAGRESPE hybrid retains the finite-horizon RRWP route and adds the same MAGRESPE node branch. Their comparison tests replacement versus complementarity.

Formulation controls. RawMagnitude replaces the forest-normalized residual with the raw ridge magnitude weighting and omits q^{-2} normalization. ResidualCommon uses the forest-normalized ridge residual but omits q^{-2} normalization. These two controls examine forest subtraction and onset normalization. DirectDefect instead retains the forest reference, the standard scale bank, and q^{-2} normalization, but omits the ridge solve. Its node channels are defined directly from the forest-prediction defect. The exact formulations are given in Appendix C.2.

C.2 MATCHED FORMULATION DECOMPOSITION

The formulation ablation separates forest normalization from the q^{-2} onset normalization. Define the raw ridge magnitude weighting by

$$w_{q,\lambda}^{\mathrm{raw}} = (Z_q^\top Z_q + \lambda I)^{-1} Z_q^\top \mathbf{1}. \tag{29}$$

Let

$$\phi(t) = \mathrm{sgn}(t) \log(1 + |t|).$$

The four matched channel constructions are

$$p_v^{\mathrm{raw}}(q) = \phi\big(w_{q,\lambda}^{\mathrm{raw}}(v)\big)\,, \tag{30}$$

$$p_v^{\mathrm{res}}(q) = \phi(r_{q,\lambda}(v))\,, \tag{31}$$

$$p_v^{\mathrm{direct}}(q) = \phi\left(\frac{e_q(v)}{q^2}\right). \tag{32}$$

$$p_v^{\mathrm{MAGRESPE}}(q) = \phi\left(\frac{r_{q,\lambda}(v)}{q^2}\right). \tag{33}$$

RawMagnitude therefore uses neither forest subtraction nor onset normalization; ResidualCommon introduces the forest-normalized residual but omits q^{-2}; DirectDefect uses the same forest-prediction defect e_q but replaces the regularized correction $r_{q,\lambda}$ with e_q before applying the same q^{-2} normalization; final MAGRESPE uses all components.

All four variants use the same five seeds, host architecture, scale bank, ridge, signed-log map, and optimization protocol.

Table 8: Matched ZINC-12K formulation ablations. Results are mean ± sample standard deviation.

Variant	Forest	q^{-2}	Ridge	Test MAE ↓
RawMagnitude	×	×	✓	0.0648 ± 0.0026
ResidualCommon	✓	×	✓	0.0634 ± 0.0039
DirectDefect	✓	✓	×	0.0610 ± 0.0035
MAGRESPE	✓	✓	✓	0.0600 ± 0.0021

The paired RawMagnitude-minus-ResidualCommon differences are $[0.0008, 0.0006, 0.0046, 0.0005, 0.0003]$, favoring forest subtraction on all five seeds, with a mean test MAE reduction of 0.0014. However, both variants have the same mean validation MAE of 0.0725. The ResidualCommon-minus-MAGRESPE differences are $[0.0068, 0.0043, -0.0002, 0.0026, 0.0036]$, favoring the q^{-2} normalization on four of five seeds, with a mean test MAE reduction of 0.0034.

To isolate the ridge solve, the paired DirectDefect-minus-MAGRESPE test differences are $[-0.0018, 0.0026, 0.0013, 0.0071, -0.0045]$. The ridge-based formulation improves three of five seeds, with a modest mean test MAE reduction of 0.0010 and a bootstrap confidence interval of $[-0.0024, 0.0044]$. Together, these results indicate contributions from forest subtraction and onset normalization, while the ridge solve provides a modest test improvement.

C.3 SCALE AND LEADING-ORDER ABLATIONS

The three-seed mechanistic experiment (Table 9) compares the standard standalone MAGRESPE representation with three variants: Small-4, which concentrates its scales near zero; Compact-4, which retains four spread finite scales; and C2, which uses only the closed-form second-order coefficient. All conditions are evaluated on aligned seeds 0–2.

Small-4 has worse validation MAE than the full eight-scale representation on all three seeds, although its mean test MAE is lower. Compact-4 achieves a lower mean validation MAE but a higher mean test MAE and greater test variability than the eight-scale reference. Its test variability is driven primarily by seed 0. C2 performs substantially worse on both validation and test, indicating that the leading second-order coefficient alone is insufficient as a standalone replacement under the tested configuration.

Table 9: Three-seed scale and leading-order ablations on ZINC.

Condition	Val MAE	Test MAE
MAGRESPE	0.0698 ± 0.0034	0.0657 ± 0.0050
Small-4	0.0760 ± 0.0069	0.0631 ± 0.0035
C2	0.1403 ± 0.0121	0.1194 ± 0.0025
Compact-4	0.0665 ± 0.0072	0.0674 ± 0.0097

C.4 PAIRED COMPARISONS ACROSS BENCHMARKS

Table 10: ZINC test MAE (mean ± SD over five seeds) for the principal matched conditions. Lower is better.

Condition	Test MAE
RRWP	0.0638 ± 0.0036
Zero-info	0.0670 ± 0.0039
SPSE	0.0611 ± 0.0031
RRWP+SignNet	0.0681 ± 0.0032
MAGRESPE	0.0645 ± 0.0045
RRWP+MAGRESPE	$\mathbf{0.0600 \pm 0.0021}$

Table 11 reports seed-aligned comparisons between the hybrid and the available baselines. On ZINC, the hybrid improves mean test MAE over RRWP by 0.0038, with 4/5 paired wins. It also outperforms standalone MAGRESPE and RRWP+SignNet on all five aligned seeds, whereas its advantage over SPSE is smaller and less consistent (3/5 wins).

On Peptides-struct, the hybrid achieves mean paired improvements of 0.0022 and 0.0025 MAE over RRWP and SPSE, respectively, with 3/5 and 4/5 wins. It also improves over standalone MAGRESPE and RRWP+SignNet on all five seeds. On PascalVOC-SP, the hybrid improves macro-F1 over RRWP by 0.0102 on average, with all five paired differences positive.

Table 11: Five-seed paired test comparisons against RRWP+MAGRESPE across three benchmarks. Positive gains favor the hybrid. For ZINC and Peptides-struct, gain is baseline MAE minus hybrid MAE; for PascalVOC-SP, gain is hybrid macro-F1 minus baseline macro-F1. Results use aligned seeds 0–4. Confidence intervals are 95% percentile bootstrap intervals. W/T/L denotes hybrid wins, ties, and losses. Sign-test p-values are exact, two-sided, and unadjusted.

Comparator	Paired gain $\pm$ SD	95% bootstrap CI	W/T/L	Sign p
ZINC (test MAE)				
RRWP	$+0.0038 \pm 0.0045$	$[0.0005, 0.0074]$	4/0/1	0.3750
Zero-info	$+0.0070 \pm 0.0028$	$[0.0047, 0.0091]$	5/0/0	0.0625
SPSE	$+0.0010 \pm 0.0034$	$[-0.0017, 0.0036]$	3/0/2	1.0000
MAGRESPE	$+0.0044 \pm 0.0050$	$[0.0008, 0.0086]$	5/0/0	0.0625
RRWP+SignNet	$+0.0080 \pm 0.0041$	$[0.0048, 0.0112]$	5/0/0	0.0625
Peptides-struct (test MAE)				
RRWP	$+0.0022 \pm 0.0047$	$[-0.0009, 0.0061]$	3/0/2	1.0000
SPSE	$+0.0025 \pm 0.0029$	$[0.0005, 0.0049]$	4/0/1	0.3750
MAGRESPE	$+0.0086 \pm 0.0024$	$[0.0070, 0.0106]$	5/0/0	0.0625
RRWP+SignNet	$+0.0070 \pm 0.0037$	$[0.0038, 0.0095]$	5/0/0	0.0625
PascalVOC-SP (test macro-F1)				
RRWP	$+0.0102 \pm 0.0087$	$[0.0058, 0.0180]$	5/0/0	0.0625

Overall, the results indicate positive mean improvements from augmenting RRWP with MAGRESPE across all three benchmarks. The consistent directional advantages over standalone MAGRESPE and the tested spectral control further support its complementary role. However, none of the exact two-sided sign tests reaches the conventional 0.05 significance threshold, and the bootstrap intervals should be interpreted cautiously given the small number of seeds. Because only five seeds are available, the confidence intervals and sign tests have limited resolution. We therefore interpret the real-world results as evidence of positive mean improvements rather than conclusive statistical superiority.

C.5 ZINC STRUCTURAL AUDIT

We additionally examine the topology of the ZINC benchmark without using predictive labels. Among all 12,000 graphs, only 66 (0.55%) are forests. Among cyclic graphs, the observed girth distribution is dominated by girths five and six: 735 graphs have girth three, 181 have girth four, 7,139 have girth five, 3,868 have girth six, and 11 have girth at least seven.

The practical ZINC RRWP setting uses horizon $h = 20$. This already exceeds the graph diameter for 11,963 of 12,000 graphs (99.69%); 25 graphs have diameter exactly 20 and only 12 exceed 20. The finite-horizon separation theorem should therefore be interpreted as an accessibility result rather than as evidence that the ZINC improvement is simply caused by RRWP failing to reach distant vertices.

The MAGRESPE representation nevertheless varies strongly with non-forest structure. Spearman correlations between cycle rank and the node-representation norm are $0.6860/0.6660/0.6940$ on the train/validation/test splits, respectively. The corresponding correlations with triangle count are $0.4120/0.4560/0.4010$, whereas correlations with graph diameter are only $0.0230/-0.0090/0.0650$. These statistics are descriptive and are not used for model selection or as evidence that any individual structural quantity causally explains the predictive improvement.

D DETERMINISTIC STRUCTURAL CERTIFICATES

D.1 CONNECTED SUBDIVISION CERTIFICATES AND FIXED-BANK NUMERICAL ACCESSIBILITY

Arbitrary-horizon certificates. We verify the connected separation theorem at representative horizons $h \in \{2, 5, 10, 20\}$. For each horizon, we choose the smallest $L \geq 2$ satisfying $2h < 5L$, giving $L \in \{2, 3, 5, 9\}$ and graph orders $\{35, 56, 98, 182\}$. In all four cases, the typed rooted RRWP row-profile summaries coincide exactly, while the predicted onset order $\lceil 5L/2 \rceil$ of X_{14} is strictly

smaller than the Heawood onset order $3L$. These deterministic certificates verify the arbitrary-finite-horizon construction independently of predictive experiments.

Numerical accessibility of the fixed scale bank. We further examine the actual 240-graph connected subdivision benchmark, comprising 120 graphs of girth 45 and 120 of girth 54, each with 286 vertices. Using the fixed eight-scale bank and $\lambda = 10^{-3}$, we evaluate both the cached float32 MAGRESPE signatures used during training and fresh float64 recomputations.

For each scale q, we sort the node-channel values of each graph and measure the ℓ_∞ distance between the resulting vectors for all $120 \times 120 = 14{,}400$ cross-class graph pairs. Let $\delta_q^{\min}$ denote the minimum of these distances. We estimate an empirical numerical perturbation bound η_q from the largest observed discrepancy under random vertex relabeling and independent cache recomputation.

Table 12 reports the resulting separation margins. At $q \le 0.21$, the channels lie near the numerical cancellation floor and resolve no cross-class pairs at the $10\eta_q$ threshold. In contrast, each of $q \in \{0.34, 0.52, 0.74, 0.95\}$ resolves all $14{,}400$ pairs, with its minimum separation exceeding $10^3\eta_q$. Thus, the fixed bank provides numerically stable distinctions on the implemented benchmark without requiring scales outside the experimental configuration.

Table 12: Fixed-bank numerical accessibility on the 240-graph connected subdivision benchmark. η_q is the largest observed numerical perturbation, and $\delta_q^{\min}$ is the minimum sorted-channel ℓ_∞ distance over all $14{,}400$ cross-class pairs. A pair is resolved when its distance exceeds $10\eta_q$.

q	η_q	$\delta_q^{\min}$	Resolved pairs
0.05	9.1×10^{-14}	2.2×10^{-15}	0
0.08	3.7×10^{-14}	7.9×10^{-15}	0
0.13	2.6×10^{-14}	1.9×10^{-15}	0
0.21	1.3×10^{-14}	2.2×10^{-15}	0
0.34	2.3×10^{-14}	1.0×10^{-10}	14,400
0.52	1.1×10^{-13}	3.8×10^{-7}	14,400
0.74	5.7×10^{-11}	1.3×10^{-4}	14,400
0.95	3.0×10^{-8}	1.2×10^{-2}	14,400

The attenuation of small-q signals is consistent with the girth-controlled onset orders q^{23} and q^{27} before q^{-2} normalization. The empirical perturbation bounds quantify numerical stability under the tested computations; they are not formal floating-point error bounds. Moreover, these measurements concern the deterministic pre-encoder signatures and do not establish which channels the trained network exploits.

D.2 EXACT FINITE-SCALE CONNECTED CERTIFICATES

We consider the Heawood graph H and a connected cubic graph X_{14} on vertices $\{0, \dots, 13\}$, defined by

$$\begin{aligned} E(X_{14}) = \{&(0,5),(0,8),(0,9),(1,2),(1,8),(1,10),\\ &(2,3),(2,9),(3,12),(3,13),(4,7),(4,11),\\ &(4,12),(5,7),(5,10),(6,7),(6,8),(6,13),\\ &(9,11),(10,12),(11,13)\}. \end{aligned}$$

Both graphs are connected and 3-regular, with 14 vertices and 21 edges, but their girths are six and five, respectively.

Because both graphs have girth greater than four, every root has three neighbors and six distinct vertices at distance two, with no additional walks of length at most two. Consequently, their horizon-

two RRWP row-profile multisets coincide. More explicitly, for every root u in either graph,

$$\mathcal{R}_2(u) = \left\{\left(1, 0, \frac{1}{3}\right)\right\}^{\times 1} \uplus \left\{\left(0, \frac{1}{3}, 0\right)\right\}^{\times 3} \uplus \left\{\left(0, 0, \frac{1}{9}\right)\right\}^{\times 6} \uplus \{(0,0,0)\}^{\times 4}.$$

Thus, $\mathcal{S}_2(H) = \mathcal{S}_2(X_{14})$.

We evaluate their forest-prediction defects and ridge-corrected residuals at two finite metric scales, using $\lambda = 10^{-3}$. For a graph G, let $\mathcal{M}_e(G; q)$ and $\mathcal{M}_r(G; q, \lambda)$ denote the multisets of its node-wise defect and corrected-residual values, respectively.

Certificate at $q = 1/2$. For every cubic graph, the forest reference vanishes at this scale:

$$b_{1/2}(v) = \frac{1 - 2(1/2)}{1 + 1/2} = 0.$$

Hence $e_{1/2} = \mathbf{1}$, and the ridge-corrected residual coincides with the raw ridge magnitude weighting.

The exact corrected-residual multisets are

$$\mathcal{M}_r(H; 1/2, \lambda) = \left\{\frac{4500}{20251}\right\}^{\times 14}, \tag{34}$$

and

$$\mathcal{M}_r(X_{14}; 1/2, \lambda) = \left\{\frac{85352554000}{398997195887}\right\}^{\times 6} \uplus \left\{\frac{88838616500}{398997195887}\right\}^{\times 6} \uplus \left\{\frac{99439022000}{398997195887}\right\}^{\times 2}. \tag{35}$$

Here $\{x\}^{\times k}$ denotes k copies of x, and $\uplus$ denotes multiset union. The Heawood residual is constant, whereas X_{14} has three distinct residual values. Since the signed-log channel map is strictly increasing at fixed q, the corresponding MAGRESPE channel multisets also differ.

This certificate establishes exact finite-scale separation despite the horizon-two RRWP row-profile collision. However, because $b_{1/2} = 0$, it does not isolate the contribution of forest subtraction.

Certificate at $q = 1/3$. We next consider a scale at which the forest reference is nonzero. Since both graphs are cubic,

$$b_{1/3} = \frac{1}{4}\mathbf{1}.$$

Their exact forest-prediction defect multisets are

$$\mathcal{M}_e(H; 1/3) = \left\{\frac{8}{27}\right\}^{\times 14},$$

and

$$\mathcal{M}_e(X_{14}; 1/3) = \left\{\frac{8}{27}\right\}^{\times 12} \uplus \left\{\frac{49}{162}\right\}^{\times 2}.$$

Thus, their defects differ even before the regularized correction is applied.

At $\lambda = 10^{-3}$, the corrected residual is constant on the Heawood graph:

$$r^{H}_{1/3,\lambda}(v) = \frac{608000}{5776729} \qquad \text{for all } v \in V(H).$$

By contrast, vertex 2 of X_{14} has

$$r^{X_{14}}_{1/3,\lambda}(2) = \frac{69883344589295000}{596429760885606343}.$$

These values are unequal. Because the Heawood residual is constant, this single coordinate establishes that the complete corrected-residual multisets differ. The corresponding MAGRESPE channel multisets are therefore also distinct.

Interpretation. Both certificates establish finite-scale separation of the connected graph pair despite identical horizon-two RRWP row-profile multisets. The first uses a scale at which the forest reference vanishes; the second shows that separation persists with a nonzero forest reference and distinct forest-prediction defects.

Neither certificate establishes that forest subtraction is necessary for distinguishing this pair. Moreover, the collision concerns rooted RRWP row-profile multisets, not the full indexed RRWP representation.

D.3 Sharp Cycle Threshold and Fixed-Scale Separation

We independently verify the sharp RRWP separation horizon $h_\star(m) = \lceil m/2 \rceil$ for C_{2m} versus $2C_m$ across 13 tested cycle sizes. In particular, at the RRWP horizon $h = 20$ used on ZINC, the rooted row-profile summaries of C_{82} and $2C_{41}$ coincide, and they first differ at $h = 21$.

We measure numerical MagResPE separation using the sorted-channel ℓ_∞ distance defined in Appendix D.1. At $q = 0.95$ and $\lambda = 10^{-3}$, this distance is approximately 0.0112 for the cycle pair. This provides a fixed-scale numerical witness of the theoretical separation without requiring a larger RRWP horizon.

Interface-storage interpretation. For the cycle family, the first separating RRWP horizon is $h_\star(m) = \lceil m/2 \rceil = \Theta(n)$, where $n = 2m$ is the number of vertices. Explicitly materializing the complete dense RRWP tensor through this horizon requires $\Theta(n^2)$ ordered pairs per transition power and hence $\Theta(n^3)$ stored scalars. In contrast, the separating MagResPE witness uses one post-preprocessing scalar per vertex, requiring $\Theta(n)$ representation storage. This is an interface-storage comparison rather than an end-to-end complexity advantage: exact MagResPE preprocessing still requires all-pairs shortest-path distances and dense ridge solves.

D.4 High-precision girth-onset verification

We independently verify the small-q onset predicted by the girth theorem using 90-digit arithmetic on cycles and cycle-plus-tree graphs with girths 3–10. Across all 16 tested graphs, the fitted onset slopes agree with the predicted order $\lceil g/2 \rceil$, with maximum absolute error 9.37×10^{-4}. A random-tree control has residual norm below 8.5×10^{-91}, consistent with the exact zero residual on forests. This label-free diagnostic provides a numerical implementation check of the asymptotic result independently of the predictive experiments.

E Computational and Numerical Analysis

E.1 Asymptotic cost

Let $n = |V|$ and $K = |\mathcal{Q}|$. Exact preprocessing first computes all-pairs shortest-path distances, then performs K dense $n \times n$ ridge solves. With direct dense linear algebra, the solve stage has worst-case cost $O(Kn^3)$ and requires $O(n^2)$ working memory for the dense matrices. Once preprocessing is cached, the neural model consumes an $n \times K$ node feature tensor, i.e., $O(nK)$ storage.

E.2 Controlled runtime and memory

On ZINC-12K, we measure the steady-state computational cost of the structural encodings on an exclusive NVIDIA A100 after all required representations have been constructed and cached. Training cost is reported as mean epoch time over three warm repeats, while inference cost is measured by validation-forward throughput. We additionally report peak CUDA memory actively allocated by each process.

Once cached, adding MagResPE to the RRWP host incurs only a modest steady-state training cost: mean epoch time increases from 63.815 s to 65.116 s. Validation-forward throughput remains comparable, increasing from 1173.69 to 1204.31 graphs/s, while peak allocated CUDA memory remains of the same order as the RRWP baseline. Standalone MagResPE is substantially lighter because it does not maintain the RRWP representation.

Table 13: Controlled steady-state runtime and memory measurements on an exclusive A100. Structural representations are cached before timing. Alloc. MB denotes peak CUDA memory actively allocated by the process.

Condition	Warm epoch (s)	Alloc. MB	Val. graphs/s
RRWP	63.815 ± 0.463	696.1	1173.69 ± 78.60
SPSE	64.221 ± 0.174	696.1	1229.34 ± 23.85
MAGRESPE	57.273 ± 0.355	135.1	1562.16 ± 51.32
Compact-4	57.502 ± 0.127	134.7	1465.48 ± 125.78
RRWP+MAGRESPE	65.116 ± 0.597	665.7	1204.31 ± 64.25

Table 14 reports the preprocessing time of MAGRESPE across all three datasets, using the full eight-scale bank ($\lambda = 10^{-3}$) on a single CPU core. Total preprocessing takes 11.7 seconds on ZINC, 128.7 seconds on Peptides-struct, and approximately 18 minutes on PascalVOC-SP. The cost increases with graph size, reflecting the expense of exact shortest-path computation and dense ridge solves. Nevertheless, preprocessing is performed only once per dataset, and the resulting encodings can be cached and reused across training runs.

Table 14: MAGRESPE preprocessing time using the full eight-scale bank ($\lambda = 10^{-3}$) on a single CPU core.

Dataset	Graphs	Nodes	Total
ZINC	12,000	23/37	11.7 s
Peptides-struct	15,535	151/444	128.7 s
PascalVOC-SP	11,355	479/500	1,086 s

E.3 PREPROCESSING SCALING

To measure the cost that is hidden once MAGRESPE representations are cached, we separately profile the preprocessing pipeline on synthetic sparse connected graphs. The graphs contain approximately $2n$ edges, with $n \in \{20, 40, 80, 160, 320\}$, and each graph size is evaluated over three repetitions. The reported times are medians across repetitions.

Table 15: Empirical preprocessing scaling of MAGRESPE on synthetic sparse connected graphs. Times are median wall-clock preprocessing times over three repetitions.

Number of nodes n	20	40	80	160	320
Preprocessing time (ms)	2.93	4.75	11.84	41.97	156.50

Preprocessing becomes increasingly expensive as graph size grows, consistent with the dense algebra in the exact construction. For $n = 320$, the recorded median component times are 41.95 ms for all-pairs shortest paths, 31.99 ms for construction of the multi-scale magnitude kernels, and 82.27 ms for the eight ridge solves. Thus, at the largest tested size, the ridge systems account for the largest component of the preprocessing cost.

These measurements should be interpreted as an implementation scaling curve rather than an empirical complexity fit. The exact method still requires all-pairs distances and dense ridge algebra, consistent with the $O(Kn^3)$ worst-case solve cost described above. This explains why preprocessing is inexpensive for the small graphs used in ZINC but can become substantial on datasets containing larger individual graphs, as observed for PascalVOC-SP.

E.4 NUMERICAL CONDITIONING AND SOLVE ACCURACY

Positive ridge guarantees that $Z_q^\top Z_q + \lambda I$ is positive definite and uniquely solvable, but uniqueness does not imply favorable conditioning. We measure the spectral condition number for an invertible

matrix B

$$\kappa_2(B) = \frac{\sigma_{\max}(B)}{\sigma_{\min}(B)},$$

where larger values indicate poorer numerical conditioning.

On 120 ZINC graphs sampled across graph sizes, conditioning is poorest at the largest MAGRESPE scale, $q = 0.95$. At this scale,

$$\text{median}\, \kappa_2(Z_q) = 3.86 \times 10^3,$$

with 95th percentile 6.24×10^3 and maximum 7.60×10^3. For the regularized normal matrix,

$$\text{median}\, \kappa_2(Z_q^\top Z_q + \lambda I) = 3.21 \times 10^5,$$

with 95th percentile 4.99×10^5 and maximum 6.44×10^5.

Despite this conditioning, the relative residual of the solved normal system remains very small: at $q = 0.95$, the median is 1.23×10^{-16} and the maximum is 2.23×10^{-15}. Thus the implemented solver accurately satisfies the specified linear system for the tested graphs. This should not be interpreted as a claim that the underlying kernel is insensitive to perturbations. For two triangles joined by a bridge, the maximum coordinate discrepancy is 4.59×10^{-2} at $\lambda = 10^{-3}$.

Although Z_q is symmetric, solving the regularized least-squares problem through the normal equations can amplify numerical conditioning, since the singular values of Z_q are squared before regularization. The positive ridge parameter ensures a unique solution but does not eliminate this conditioning concern. The small measured linear-system residuals confirm accurate satisfaction of the normal equations in our tested setting, rather than forward accuracy or perturbation stability of the resulting encoding.

F LIMITATIONS AND BROADER IMPACT

Limitations. MAGRESPE is intentionally a compact structural summary rather than a complete graph invariant. In particular, every forest maps to zero, and distinct graphs can share the same residual signatures; MAGRESPE should therefore not be interpreted as solving graph isomorphism. Empirically, its strongest role is complementary: standalone replacement is not consistently competitive on Peptides-struct and performs poorly on PascalVOC-SP, so our claim is that MAGRESPE complements finite-horizon RRWP rather than universally replacing existing PEs. Our finite-horizon separation concerns rooted RRWP row-profile summaries and does not establish an indistinguishability result for the complete RRWP tensor or for arbitrary Graph Transformer architectures. Finally, exact preprocessing requires all-pairs shortest paths and dense ridge solves, which limits applicability to large individual graphs. Results are obtained with a fixed scale bank and the GRIT implementation.

Broader impact. This work develops a general-purpose graph representation technique and does not directly generate molecules, recommend interventions, or make clinical predictions. The proposed encoding may be useful in domains where graph structure is important, including molecular, biological, social, and knowledge-network applications. As with other learned graph models, downstream behavior can reflect biases or limitations in the training data and evaluation benchmarks. In safety-critical applications, particularly molecular and biomedical settings, model predictions should therefore be treated as computational evidence requiring appropriate domain-specific and experimental validation. The dense preprocessing required by the current implementation also introduces additional computational cost, motivating more efficient variants in future work.

G REPRODUCIBILITY INFORMATION

All predictive experiments use fixed configuration files and select the reported model using validation performance. The test split is evaluated once at the selected checkpoint for each run. The corresponding model architectures and optimization settings on all benchmark datasets are reported in Appendix B. Reported aggregate results are computed directly from the corresponding per-seed outputs. Paired comparisons use aligned seed identities, and bootstrap intervals and sign tests are treated as descriptive because of the small number of seeds.