Resolving Node Identifiability in Graph Neural Processes via Laplacian Spectral Encodings

Anonymous ACL submission

Abstract

Message-passing GNNs are 1-WL limited and can collapse distinct nodes on symmetric graphs, which in Graph Neural Processes leads to intrinsic posterior ambiguity and a non-vanishing Bayes-risk floor for node localization. We prove that Laplacian spectral positional information breaks this identifiability barrier, establishing a sample-complexity separation on random r-regular graphs: constant-shot identifiability is achievable with spectral coordinates, while WL-bounded GNPs fail in the sublogarithmic regime. The proof links shortest-path observations to diffusion geometry in a logarithmic tree-like window, applies constant-anchor trilateration, and uses quantitative spectral injectivity with logarithmic-size coordinates. Empirically, we adopt the practical choice of concatenating a few raw Laplacian eigenvectors and observe improved accuracy and faster optimization on drug-drug interaction prediction.

1 Introduction

Learning from graph-structured data underpins applications such as molecular chemistry and drug discovery (2017) and large-scale recommendation (2018). Message-passing graph neural networks are the dominant approach, building representations by iteratively aggregating local neighborhood information (2016). Probabilistic formulations such as graph neural processes (GNPs) extend this pipeline by modeling distributions over functions on graphs, enabling context-conditioned prediction with uncertainty estimates that are valuable when supervision is limited or reliability matters (2018; 2025b).

Most existing GNPs instantiate their encoder with a standard message-passing backbone, for example GCN- or GraphSAGE-style architectures (2019). These encoders produce node embeddings that are aggregated into a context-dependent representation consumed by the neural process decoder. This design is practical and effective, but it inherits the representational limits of the underlying message-passing encoder.

A key limitation is that standard message-passing GNNs are bounded by the 1-Weisfeiler-Lehman (1-WL) test (2019; 2019). On regular or highly symmetric graphs, many nodes share identical local structural signatures, so a 1-WL-bounded encoder can collapse distinct nodes into the same embedding. In a probabilistic encoder-decoder setting, this collapse has a direct statistical consequence: if the context does not allow the encoder to separate multiple candidate sources, the posterior remains ambiguous and the Bayes risk for node-localization cannot vanish.

This leads to the following question: *can we design a graph neural process that provably breaks the 1-WL identifiability barrier and avoids a Bayes-risk floor induced by structural symmetry, while remaining compatible with permutation invariance and practical message passing?*

We address this question by injecting global positional information derived from the graph Laplacian spectrum. We formalize a Laplacian-augmented graph neural process (Lap-GNP) and analyze it on random r-regular graphs, a regime where the 1-WL limitation is sharp. Our main result proves a sample-complexity separation: with Laplacian spectral positional information, a constant number of context observations can suffice for node identifiability, whereas any WL-bounded GNP requires growing context and still exhibits a nontrivial error floor in the sublogarithmic regime. The proof proceeds by converting shortest-path observations into diffusion geometry within a logarithmic tree-like window, recovering the source location via constant-anchor trilateration, and applying quantitative spectral injectivity with logarithmic-size coordinates. This chain makes explicit why global coordinates resolve ambiguities that local message passing cannot eliminate.

While the theory clarifies which spectral symmetries should be respected at the feature level, our empirical evaluation follows the standard practical choice of directly concatenating a small number of raw Laplacian eigenvectors to the node features. On drug-drug interaction (DDI) prediction, this simple spectral augmentation yields consistent gains and faster optimization relative to a no-PE baseline, supporting the practical value of explicit global positional information.

The primary contributions of this work are as follows:

1. We formalize node indistinguishability in Graph Neural Processes with 1-WL-bounded GNN encoders (WL-GNPs) and prove that, on highly symmetric graphs, it induces intrinsic posterior ambiguity and a nontrivial Bayes-risk lower bound for node identification with sublogarithmic context size.
2. We propose the Laplacian Graph Neural Process (Lap-GNP) by augmenting a GNP encoder with Laplacian spectral positional information, and we prove a sharp sample-complexity separation on random r-regular graphs: Lap-GNP achieves constant-shot node identifiability while WL-GNPs provably fail.
3. An empirical study on a real-world Drug–Drug Interaction (DDI) prediction task shows that concatenating a small number of raw Laplacian eigenvectors improves predictive performance and accelerates optimization relative to a no-PE baseline, supporting the practical value of spectral positional information in GNPs.

2 Literature Review

We briefly review prior work on message-passing expressiveness, neural processes on graphs, and positional encodings. Our focus is on how 1-WL limitations translate into posterior ambiguity and motivate injecting global positional information.

2.1 Expressiveness Limits of Message Passing GNNs

Message-Passing Graph Neural Networks (MPNNs) are a standard backbone for learning on graphs, spanning node-, edge-, and graph-level prediction (2017; 2017; 2021). Canonical architectures (e.g., GCN, GraphSAGE, GAT, and GIN) update node representations by aggregating neighborhood information over multiple rounds (2018; 2019). A central limitation is that many MPNNs are bounded by the 1-dimensional Weisfeiler-Lehman (1-WL) test (2019; 2019), which implies that structurally symmetric (1-WL-indistinguishable) nodes cannot be separated when initial identifiers are weak or absent. This phenomenon is especially pronounced on regular or highly symmetric graphs, and motivates augmenting permutation-invariant models with additional positional signals.

2.2 Neural Processes on Graphs and Posterior Ambiguity

Neural Processes (NPs) learn distributions over functions conditioned on a context set, combining amortized inference with uncertainty quantification (2018; 2019). Graph Neural Processes (GNPs) instantiate NP encoders with GNN backbones to model relational tasks with few-shot conditioning and predictive uncertainty. However, when the encoder inherits 1-WL limitations, distinct candidate nodes can induce identical representations under the available observations, yielding ambiguous posteriors on symmetric graphs (2019; 2019). Related probabilistic graph models (e.g., graph VAEs and Bayesian GNNs) highlight the importance of uncertainty for reliability, but they typically do not connect encoder design-especially positional design-to formal identifiability guarantees (2016; 2020; 2025a). Our work aims to make a modest step in this direction by analyzing how WL-bounded encoders translate into a Bayes-risk lower bound, and how adding appropriate positional information can mitigate posterior ambiguity.

2.3 Positional Encodings: Structural Signals and Invariant Spectral Coordinates

Positional encodings (PEs) provide permutation-invariant architectures with additional coordinates reflecting position or geometry. Structural and diffusion-based encodings (e.g., shortest-path distances, PPR, heat kernels) have been used to enrich representations beyond local message passing (2023; 2024). Spectral approaches construct global coordinates from Laplacian eigenpairs and can break symmetries that remain invisible to 1-WL message passing (2021; 2021; 2022). A practical challenge is that eigenvectors are not uniquely defined due to sign flips and rotations within degenerate eigenspaces; recent work addresses this via sign- and basis-invariant spectral constructions (2022).

In our setting, such invariance is not merely a robustness detail: it supports a coherent notion of positional information inside a probabilistic encoder. For completeness, we provide a longer technical comparison to Distance Encoding (DE) (2020) in Appendix A.

2.4 DDI Link Prediction

Drug-Drug Interaction (DDI) prediction has evolved from text-mining and similarity-based models to graph-centric pipelines operating on molecular graphs and biomedical networks (2013; 2018). Recent methods incorporate knowledge graphs, multimodal signals, and interaction-specific modules such as co-attention or gating (2024; 2023; 2025b). We adopt a strong GNP-based baseline aligned with widely used DrugBank protocols (2014; 2025), and use DDI as an applied benchmark to examine whether principled positional information yields measurable benefits under both inductive and transductive settings.

Overall, we do not view our results as a complete theory of positional design for probabilistic graph models. Rather, we aim to connect two lines of work, WL-bounded expressiveness and its implications for ambiguity, and invariant spectral positional encodings to evaluate this connection in a representative downstream task.

3 Preliminaries

This section sets up the few-shot node-localization problem and the associated observation model. The notation introduced here will be used to state identifiability and Bayes-risk guarantees.

3.1 Problem Setup and Notation

Let $G = (V, E)$ be a finite, connected, undirected graph with $|V| = n$. An unobserved source node $v_0 \sim \mathrm{Unif}(V)$ induces the target function

$$f_{G,v_0}(v) \;=\; \mathrm{SPD}(v, v_0), \qquad v \in V. \tag{1}$$

The learner observes a context set $\mathcal{C} = \{(v_i, y_i)\}_{i=1}^{k}$ where $v_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(V)$ and $y_i = f_{G,v_0}(v_i)$. The goal is to identify v_0 (equivalently, reconstruct f_{G,v_0} on V) from $(G, \mathcal{C})$ using as few context points as possible.

We use a Bayes-risk view to formalize identifiability under restricted encoder families. Let $Z := (G, \mathcal{C})$ and let an encoder produce a summary $R = \mathrm{Enc}(G, \mathcal{C})$. A decision rule δ outputs $\hat{v} = \delta(Z, R)$ and is evaluated by $\mathrm{Risk}(\delta) = \Pr[\hat{v} \neq v_0]$. If two nodes induce the same posterior given (Z, R) under an encoder class, they are indistinguishable; write $[v_0]$ for the induced equivalence class. Then

$$\Pr(\hat{v} \neq v_0 \mid Z, R) \;\geq\; 1 - \mathbf{E}[|[v_0]|^{-1} \mid Z, R]. \tag{2}$$

Thus, showing $|[v_0]| \geq 2$ with high probability implies a non-trivial error floor for that encoder family.

A Graph Neural Process (GNP) consists of an encoder $R = \mathrm{Enc}(G, \mathcal{C})$ and a decoder $\hat{f}(v) = \mathrm{Dec}(v, R)$. We compare two encoder families: WL-bounded encoders (denoted WL-GNP), and the same backbone augmented with Laplacian spectral positional information (denoted LAP-GNP). The theoretical analysis treats the Laplacian positional features as sign-/basis-invariant; a concrete invariant construction and implementation notes (including the practical eigenvector variants used in experiments) are deferred to Appendix B.1. We also discuss an optional connection to Distance Encoding in Appendix B.2.

3.2 Distances and Laplacian Basics

For $u, v \in V$, let $\mathrm{SPD}(u, v)$ denote the shortest-path distance. We write $B_G(u, R) = \{w \in V : \mathrm{SPD}(u, w) \leq R\}$ and $\mathrm{diam}(G)$ for the diameter.

Let A be the adjacency matrix and D the degree matrix. We use the normalized Laplacian

$$L_{\mathrm{sym}} := I - D^{-1/2} A D^{-1/2}. \tag{3}$$

For $t > 0$, define the heat kernel operator $K_t = e^{-tL_{\mathrm{sym}}}$ with entries $k_t(u, v) = (K_t)_{uv}$. The diffusion distance is

$$d_t(u, v)^2 \;=\; \sum_{j \geq 1} e^{-2t\lambda_j} \big(\phi_j(u) - \phi_j(v)\big)^2, \tag{4}$$

where $\{(\lambda_j, \phi_j)\}$ are Laplacian eigenpairs. We will use the kernel identity

$$d_t(u, v)^2 \;=\; k_{2t}(u, u) + k_{2t}(v, v) - 2k_{2t}(u, v). \tag{5}$$

Let $\{(\lambda_j, \phi_j)\}_{j \geq 1}$ denote the nonzero eigenpairs of L_{sym}. For $t > 0$ and an integer $m \geq 1$, define the truncated diffusion-map coordinates

$$\Phi_t^{(m)}(v) := \big(e^{-t\lambda_1}\phi_1(v), \ldots, e^{-t\lambda_m}\phi_m(v)\big) \in \mathbb{R}^m, \tag{6}$$

and the corresponding truncated diffusion distance

$$d_t^{(m)}(u, v)^2 := \sum_{j=1}^{m} e^{-2t\lambda_j} \big(\phi_j(u) - \phi_j(v)\big)^2. \tag{7}$$

3.3 Graph Distribution for the Separation Result

Our main separation theorem is stated for random r-regular graphs $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$, which form a canonical hard regime for WL-bounded encoders with trivial initial features. Further discussion is deferred to Appendix B.6.

4 Theoretical Analysis

This section establishes a sharp sample-complexity separation between 1-WL-bounded Graph Neural Processes (WLGNPs) with trivial initial features and Laplacian-PE-augmented GNPs (Lap-GNPs) on random r-regular graphs $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$. All graph notations, the single-source task $f_{G,v_0}(v) = \mathrm{SPD}(v, v_0)$, and the Laplacian PE $\Psi(\cdot)$ are as defined in Section 3.

4.1 Regime and Technical Ingredients

We work with $G \sim \mathcal{G}_{n,r}$ for fixed $r \geq 3$. Throughout, we only apply local couplings to a fixed (or polylog-many prescribed) set of seed vertices. Let

$$R \;\leq\; \left(\tfrac{1}{2} - \epsilon_0\right) \frac{\log n}{\log(r-1)}, \tag{8}$$

so that radius-R neighborhoods around such seeds are tree-like w.h.p. We also impose a logarithmic depth budget

$$L \;\leq\; \left\lceil \left(\tfrac{1}{2} + c\right) \frac{\log n}{\log(r-1)} \right\rceil, \tag{9}$$

for a fixed constant $c > 0$.

Proposition 1. *The spectral positional encoding $\Psi(\cdot)$ defined in Appendix B.1 is robust to the two inherent non-uniquenesses of Laplacian eigenvectors: the sign ambiguity of individual eigenvectors and orthogonal rotations within any degenerate eigenspace.*

Theorem 2. *There exist constants $t \in [t_-, t_+]$, $c_1, c_2 > 0$, and a strictly increasing function $\psi : [0, R] \to \mathbb{R}_+$ such that, with probability $1 - o(1)$, for the (polylog-many) node pairs (u, v) queried by our procedure within the treelike window (in particular pairs (a_i, v_0) with $\mathrm{SPD}(a_i, v_0) \leq R$),*

$$c_1\, \psi\big(\mathrm{SPD}(u,v)\big) \;\leq\; d_t(u,v) \;\leq\; c_2\, \psi\big(\mathrm{SPD}(u,v)\big). \tag{10}$$

Assumption 1. *Fix $m \geq 1$ and $t > 0$. With probability $1 - o(1)$ over $G \sim \mathcal{G}_{n,r}$,*

$$\Delta_m := \min_{u \neq v} \|\Phi_t^{(m)}(u) - \Phi_t^{(m)}(v)\|_2 \geq n^{-\beta} \tag{11}$$

for some $\beta > 0$.

Assumption 2. *There exists a fixed integer $m \geq 1$ such that for $k_0 = m + 1$ i.i.d. uniform anchors $\{a_i\}_{i=1}^{k_0}$, the points $\{\Phi_t^{(m)}(a_i)\}_{i=1}^{k_0} \subset \mathbb{R}^m$ are affinely independent with probability $1 - o(1)$.*

4.2 Main Theoretical Results

We state the core statements used in the separation theorem. All proofs are deferred to Appendix C.

Given the context set $\mathcal{C} = \{(v_i, y_i)\}_{i=1}^k$ with $y_i = \mathrm{SPD}(v_i, v_0)$, define the distance-key map

$$\begin{aligned} \mathbf{d}(u) \;:=\; &\big(\mathrm{SPD}(u, v_1), \ldots, \mathrm{SPD}(u, v_k)\big) \\ &\in \{0, 1, \ldots, \mathrm{diam}(G)\}^k. \end{aligned} \tag{12}$$

Lemma 3. *Let $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$ and $k = o(\log n)$. With probability $1 - o(1)$ over $(G, \{v_i\}_{i=1}^k)$, the source $v_0 \sim \mathrm{Unif}(V)$ lies in a bucket $\{u : \mathbf{d}(u) = \mathbf{d}(v_0)\}$ of size at least 2.*

Lemma 4. *On an r-regular graph with trivial initial features, any 1-WL-bounded encoder assigns identical embeddings to any u, u' with $\mathbf{d}(u) = \mathbf{d}(u')$. In particular, u and u' are indistinguishable to WLGNPs given the context.*

Proposition 5. *Let $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$ and $k = o(\log n)$. Then any WLGNP decision rule has Bayes risk*

$$\mathrm{Risk}(\delta) \;\geq\; \tfrac{1}{2} - o(1) \tag{13}$$

for identifying v_0.

Corollary 1. *As a consequence of Theorem 2, for any polylog-sized prescribed set of queried pairs (a, v_0) with $\mathrm{SPD}(a, v_0) \leq R$, the proxy radius $\widehat{r} = \psi(\mathrm{SPD}(a, v_0))$ satisfies*

$$\left|\widehat{r} - d_t(a, v_0)\right| = o(1) \tag{14}$$

with probability $1 - o(1)$.

Lemma 6. *Fix $t > 0$ and $m \geq 1$. If anchors $\{a_i\}_{i=1}^{m+1}$ satisfy Assumption 2, then the first m coordinates of $\Phi_t^{(m)}(v_0) \in \mathbb{R}^m$ are uniquely determined by the $m{+}1$ distances $\{d_t^{(m)}(a_i, v_0)\}_{i=1}^{m+1}$.*

Theorem 7. *Let $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$ and $v_0 \sim \mathrm{Unif}(V)$.*

(i) If $k = o(\log n)$, then any WL-GNP *decision rule has Bayes risk* $\mathrm{Risk}(\delta) \geq \frac{1}{2} - o(1)$ *for identifying v_0.*

(ii) With Proposition 1, Theorem 2, and Assumptions 1-2. Then there exist constants $m \in \mathbb{N}$, $k_0 = m+1$, a time $t > 0$, and $M = \Theta(\log n)$ such that the Lap-GNP *family admits a depth-L model (with L satisfying (9)) that identifies v_0 with probability $1 - o(1)$ using $k \geq k_0$ i.i.d. context nodes.*

Remark 8. *Once v_0 is recovered, the target function is obtained everywhere via $\hat{f}(v) = \mathrm{SPD}(v, v_0)$.*

Theorem 9. *Under the random-wave (delocalized) surrogate model for the first $M = \Theta(\log n)$ Laplacian eigenvectors, the invariant encoding Ψ is injective and satisfies a separation bound of the form $\min_{u \neq v} \|\Psi(u) - \Psi(v)\|_2 \geq n^{-\alpha}$ with probability $1 - o(1)$.*

The proof (anti-concentration for the quadratic invariants defining Ψ) is given in Appendix C.5.

5 Experimental Setup

We conduct a controlled comparison of Laplacian-based positional encodings under a unified GNP-style backbone for drug–drug interaction (DDI) prediction. Unless stated otherwise, implementation details and hyperparameters follow (2025b) and are provided in Appendix B.7.

Dataset and protocol. We evaluate inductive DDI link prediction on DrugBank, where nodes are drugs and edges are known interactions. Following prior work, we use the standard inductive split in which test drugs are unseen during training. Molecular graphs are constructed from SMILES using RDKit; nodes and edges use standard atom/bond features.

Compared variants. To isolate positional effects, we fix the backbone and compare three variants: **No PE** (baseline without explicit positional channels), **Laplacian PE** (concatenating Laplacian eigenvectors), and **Unsigned Laplacian PE** (concatenating eigenvectors of an unsigned Laplacian). All variants share identical architecture and training settings.

Positional channels. For each graph, Laplacian PE concatenates the first k *nontrivial* eigenvectors (excluding the constant eigenvector when applicable) of the normalized Laplacian L_{sym} to the input node features ($k = 16$). The unsigned variant uses the same concatenation strategy but computes eigenvectors from $L_{\text{unsigned}} = L + 2A$. Since raw eigenvectors have sign ambiguity, we keep the PE pipeline fixed across runs and separately probe sensitivity in a controlled analysis (Appendix B.7).

Model and optimization. All models are implemented in PyTorch and PyTorch Geometric. Node/edge features are projected to hidden dimension 64, and the model stacks three GNP blocks with two message-passing iterations per block. We train with AdamW (learning rate 10^{-4}, weight decay 5×10^{-5}) using cosine annealing over 50 epochs, batch size 32 with gradient accumulation over 4 steps (effective batch size 128). Full architectural and training details are in Appendix B.7.

Evaluation. DDI prediction is treated as binary classification. We report AUROC and F1 on the held-out test set, selecting the F1 threshold on the validation set. We use validation AUROC for early stopping and model selection (patience 10 epochs) and report the test performance of the best checkpoint.

6 Experimental Results and Analysis

This section evaluates Laplacian-based positional information along two axes: practical link prediction on DrugBank under inductive/transductive splits, and node identifiability in the synthetic random r-regular regime that matches our theoretical setting. Unless otherwise stated, we follow the training protocol in Appendix B.7.

6.1 Spectral PEs improve inductive generalization

We consider inductive DDI prediction on DrugBank. We compare a backbone without positional encodings (No PE), the same model with Laplacian eigenvector positional encodings (Laplacian PE), and a sign-/basis-robust variant (Unsigned Laplacian PE).

Convergence profile. For a test metric $m(e)$ at epoch e, define the threshold hitting time

$$e_\tau := \min\{e : \; m(e) \geq \tau\}, \qquad (15)$$

computed via linear interpolation. Lower e_τ indicates faster convergence at threshold τ.

Figure 1 reports e_τ profiles for test AUROC and test F1. Both Laplacian-based encodings reduce e_τ over a wide range of thresholds, and they reach higher thresholds within the training budget where the No-PE baseline becomes slow or fails to cross. Across thresholds, Laplacian PE and Unsigned Laplacian PE behave similarly, so we defer

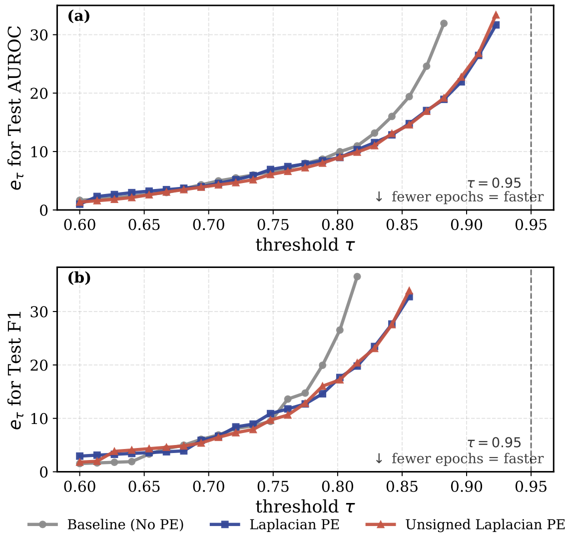

Figure 1: **Inductive DrugBank DDI: convergence profiles.** For each threshold τ, e_τ is the first epoch where the *test* metric exceeds τ (lower is faster). Laplacian-based positional encodings reduce e_τ and enable reaching higher thresholds compared to No PE.

targeted invariance checks (sign flips / eigenspace rotations) to Table 1.

6.2 Spectral PEs accelerate transductive optimization

We additionally evaluate a transductive setting where the model observes the full graph during training (including test nodes). Since final AUROC/F1 are comparable across variants, we focus on optimization efficiency.

Speedup over baseline. Let $e_\tau(\cdot)$ denote the threshold hitting time defined above. We report the speedup over No PE:

$$\Delta e_\tau := e_\tau(\text{No PE}) - e_\tau(\text{method}), \qquad (16)$$

where $\Delta e_\tau > 0$ means fewer epochs than the baseline to reach threshold τ.

Figure 2 shows Δe_τ for test AUROC and test F1 across high thresholds. Laplacian-based encodings yield positive speedups over a broad range of thresholds, indicating faster attainment of near-peak performance; the Unsigned variant is more stable at the highest thresholds, consistent with reduced sensitivity to spectral sign/basis choices.

6.3 Random r-regular graphs: empirical separation and bucket diagnostics

We now consider the synthetic regime $G \sim \mathcal{G}_{n,r}$, which matches the theoretical analysis in Section 4.

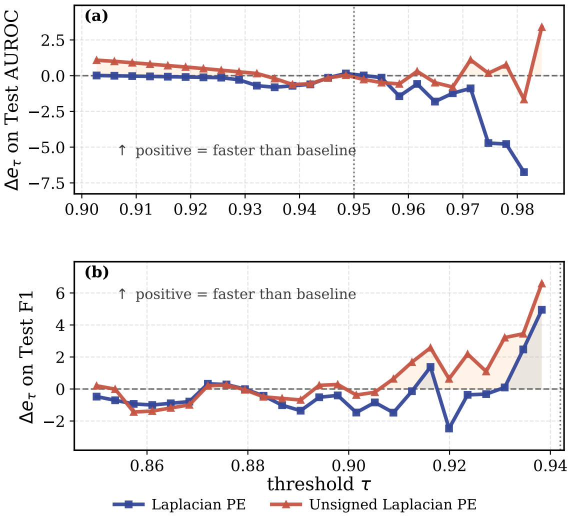

Figure 2: **Transductive DrugBank DDI: speedup profiles over No PE.** $\Delta e_\tau = e_\tau(\text{No PE}) - e_\tau(\text{method})$; positive values indicate fewer epochs than the baseline to reach threshold τ. Laplacian-based positional encodings accelerate convergence across high thresholds.

We evaluate the sample complexity and the bucket structure in two settings: the WL-bounded baseline (WLGNP) and the spectral model (Lap-GNP).

Figure 3 shows the separation between WLGNP and Lap-GNP. In Fig. 3(a), WLGNP requires progressively more contexts to achieve high accuracy, while Lap-GNP reaches near-perfect performance with a small number of contexts. This sharp transition is consistent with the constant-shot nature of Lap-GNP, as expected from the spectral localization mechanism (cf. Lemma 6).

To quantify the scaling behavior, Fig. 3(b) presents the threshold sample complexity k_τ across a range of thresholds τ. For WLGNP, k_τ grows with τ and n, indicating a logarithmic scaling in n. In contrast, Lap-GNP maintains a nearly constant k_τ, which is consistent with its constant-shot behavior.

Figure 4 shows the relationship between WL-GNP Top-1 accuracy and the bucket risk term $\mathbb{E}[1/|\text{bucket}|]$. The points align closely with the reference line $y = x$, supporting the hypothesis that WL-bounded decoders are constrained to select from an equivalence class induced by the distance keys. This leads to the conditional error model $1 - 1/|[v_0]|$ as in Eq. (2).

6.4 Corruption Robustness on DrugBank

We study robustness on DrugBank under corruptions with severity $\text{sev} \in \{0, \ldots, 5\}$, reporting AU-

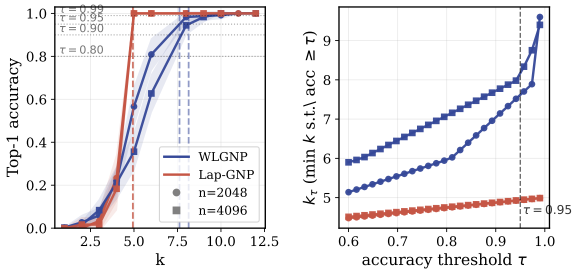

Figure 3: **Threshold profiles.** (a) Accuracy vs. k with threshold lines; vertical lines indicate $k_{\tau=0.95}$. (b) Threshold sample complexity k_τ vs. accuracy threshold τ.

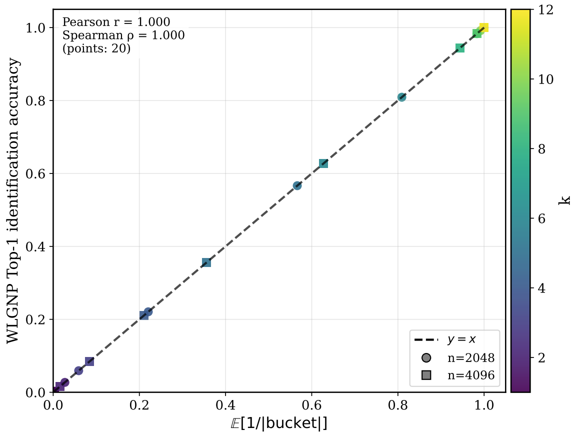

Figure 4: **Bucket-based mechanism check.** WLGNP accuracy vs. $\mathbb{E}[1/|\text{bucket}|]$ with color indicating k and markers for n. The dashed line represents $y = x$.

ROC/F1 and calibration (ECE/AURC). Table 1 separates two effects.

Task/graph corruptions. Under structural perturbation (EDGE_DROP) and label noise (LABEL_FLIP), Laplacian-based PEs retain a clear advantage over No PE at both clean and severe regimes, although all methods degrade when supervision becomes unreliable.

Representation perturbations (invariance checks). Under PE_SIGNFLIP and PE_SUBSPACE_ROT, all reported metrics change negligibly ($|\Delta| \approx 10^{-3}$). This controlled test supports the claim that the spectral features used by our models are effectively invariant to eigenvector sign flips and orthogonal rotations within eigenspaces, matching the symmetry considerations behind Proposition 1.

6.5 Ablation Study

Table 2 summarizes ablations on DrugBank and ChChMiner.

- **Hidden dimension.** Increasing the hidden size improves AUROC/F1 on both datasets, indicating that additional capacity helps integrate spectral coordinates with local features.
- **PE dimension (k).** Performance peaks at a moderate spectral dimension (DrugBank around k=16, ChChMiner around k=32); very small k under-represents global geometry, while very large k provides diminishing returns and can slightly degrade performance.
- **Propagation steps.** A small number of propagation steps is sufficient (two on DrugBank, three on ChChMiner), and deeper propagation yields limited gains, consistent with explicit PEs reducing the need to implicitly propagate long-range structure.

Table 2: Ablations on DrugBank and ChChMiner. Best within each block is shaded.

DrugBank

Dimension	**Value**	Val AUROC	Test AUROC	Val F1	Test F1
Hidden Dim					
	64	0.9382	0.9380	0.8695	0.8709
	128	0.9800	0.9802	0.9368	0.9372
PE Dim (k)					
	4	0.9269	0.9279	0.8594	0.8627
	8	0.9298	0.9299	0.8610	0.8632
	16	0.9386	0.9395	0.8715	0.8734
	32	0.9344	0.9359	0.8677	0.8707
	64	0.9366	0.9376	0.8680	0.8710
Propagation Steps (n_{iter})					
	1	0.9335	0.9339	0.8665	0.8676
	2	0.9383	0.9379	0.8689	0.8705
	3	0.9359	0.9360	0.8686	0.8690

ChChMiner

Dimension	**Value**	Val AUROC	Test AUROC	Val F1	Test F1
Hidden Dim					
	64	0.9405	0.9412	0.8699	0.8728
	128	0.9514	0.9505	0.8854	0.8827
PE Dim (k)					
	4	0.9416	0.9411	0.8717	0.8722
	8	0.9415	0.9416	0.8722	0.8736
	16	0.9405	0.9412	0.8699	0.8728
	32	0.9457	0.9454	0.8786	0.8781
	64	0.9401	0.9408	0.8720	0.8734
Propagation Steps (n_{iter})					
	1	0.9396	0.9402	0.8690	0.8709
	2	0.9405	0.9412	0.8699	0.8728
	3	0.9441	0.9442	0.8747	0.8773

Limitations

We note a few limitations and practical considerations of our theory and experiments.

Spectral computation and scalability. Our approach uses spectral coordinates, which in practice require (approximate) Laplacian eigendecomposition. This can be expensive on very large graphs

Table 1: Robustness on DDI (main metrics). Entries are mean±sd (reference values for formatting; replace with measured sd in the final version). Columns 0 and 5 denote severity 0 (clean) and 5 (worst). Δ is computed on the means (sev5−sev0). “Perturb.” reports the corruption magnitude at sev$=5$. Gray cells indicate the best mean at sev$=5$ within each corruption.

Corruption	Perturb. (sev=5)	Method	AUROC ↑			F1 ↑		
			0	5	Δ	0	5	Δ
EDGE_DROP	$p_{\text{drop}} = 0.50$	Baseline (No PE)	0.680±0.007	0.645±0.013	-0.034	0.672±0.012	0.612±0.028	-0.060
		Laplacian PE	0.886±0.014	0.823±0.019	-0.064	0.773±0.009	0.644±0.016	-0.129
		Unsigned Laplacian PE	0.892±0.015	0.837±0.012	-0.056	0.793±0.011	0.688±0.009	-0.105
LABEL_FLIP	$p_{\text{flip}} = 0.50$	Baseline (No PE)	0.891±0.009	0.735±0.012	-0.156	0.819±0.010	0.699±0.011	-0.120
		Laplacian PE	0.934±0.009	0.760±0.013	-0.175	0.868±0.008	0.724±0.021	-0.144
		Unsigned Laplacian PE	0.931±0.011	0.760±0.017	-0.171	0.864±0.015	0.726±0.017	-0.138
PE_SIGNFLIP	$p_{\text{sign}} = 1.0$	Laplacian PE	0.934±0.012	0.934±0.009	+0.000	0.867±0.008	0.865±0.018	-0.002
		Unsigned Laplacian PE	0.931±0.009	0.931±0.012	-0.000	0.865±0.021	0.863±0.027	-0.001
PE_SUBSPACE_ROT	$\gamma = 1.0$ (Haar)	Laplacian PE	0.767±0.003	0.769±0.011	+0.002	0.609±0.014	0.610±0.010	+0.001
		Unsigned Laplacian PE	0.750±0.012	0.749±0.008	-0.001	0.571±0.006	0.570±0.017	-0.001

Perturbation schedule. For EDGE_DROP/LABEL_FLIP, $p(s) = 0.1s$ so sev$=5$ corresponds to $p = 0.5$. For PE_SIGNFLIP, sev$=5$ uses i.i.d. random ±1 multipliers ($p_{\text{sign}} = 1$). For PE_SUBSPACE_ROT, sev$=5$ applies Haar-random orthogonal rotations within each degenerate eigenspace ($\gamma = 1$).

and may be a bottleneck in both time and memory, especially when the graph has millions of nodes. Although randomized and polynomial-filter approximations are natural alternatives, we do not provide a full analysis of how approximation error propagates through the trilateration pipeline or how much spectral accuracy is needed to preserve the separations required for reliable identification.

Design choices: spectral dimension and anchors. The pipeline depends on the choice of spectral dimension (k in implementation, M in theory) and the trilateration dimension/anchor count (m and $m+1$ anchors). We fix these choices for simplicity, but a principled or adaptive selection rule (e.g., based on spectral decay or empirical separation margins) is not developed here. Likewise, our theory assumes anchors in general position; we do not study data-driven anchor selection or anchor design criteria (e.g., maximizing the smallest singular value of the trilateration matrix) beyond the stylized random-anchor setting.

Empirical generality and failure modes. Empirically, on DrugBank, standard Laplacian PE performs competitively with our sign-/basis-invariant variant, suggesting that eigenvector sign ambiguity is not the dominant failure mode in this dataset. This limits the strength of empirical claims one can make about invariance in typical biomedical benchmarks; our invariant construction should be interpreted primarily as a robustness guarantee against worst-case symmetries rather than as a universally necessary ingredient. More broadly, we evaluate on DDI datasets and a specific GNP backbone; results may not fully transfer to other domains, labeling mechanisms, or architectures.

Data and evaluation caveats. DDI resources can be incomplete and subject to reporting/curation biases, and unobserved drug pairs used as negatives may include false negatives. Therefore, performance under standard splits may not fully reflect real-world pharmacovigilance settings, and the observed gains should be interpreted with these caveats in mind.

Future directions. We treat positional encodings and anchors as fixed, precomputed components. We leave joint end-to-end learning of anchors, radial distance transforms, and spectral regularizers-as well as scalable spectral coordinate constructions that avoid full eigendecomposition while preserving trilateration geometry-to future work.

Use of AI assistants. We used AI assistants to support code development (e.g., debugging and boilerplate). All experimental results, analyses, and claims were produced and verified by the authors.

Acknowledgments

The authors would like to express their sincere gratitude to all the anonymous reviewers for their careful reading and insightful suggestions.

References

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

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

Marta Garnelo, Johannes Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J Rezende, S. M. Ali Eslami, and Yee Whye Teh. 2018. Conditional neural processes. In *International Conference on Machine Learning (ICML)*, pages 1704–1713. PMLR.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. Pmlr.

Francesco Giuliari, Gianluca Scarpellini, Stefano Fiorini, Stuart James, Pietro Morerio, Yiming Wang, and Alessio Del Bue. 2024. Positional diffusion: Graph-based diffusion models for set ordering. *Pattern Recognition Letters*, 186:272–278.

William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Arman Hasanzadeh, Ehsan Hajiramezanali, Shahin Boluki, Mingyuan Zhou, Nick Duffield, Krishna Narayanan, and Xiaoning Qian. 2020. Bayesian graph neural networks with adaptive connection sampling. In *International conference on machine learning*, pages 4094–4104. PMLR.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. 2019. Attentive neural processes. *arXiv preprint arXiv:1901.05761*.

Thomas N. Kipf and Max Welling. 2016. Variational graph auto-encoders. In *NeurIPS Workshop on Bayesian Deep Learning*.

TN Kipf. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Devin Kreuzer, Dominique Beaini, Will Hamilton, V Liptchinsky, and R Pouliot. 2021. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 22635–22647.

Vivian Law, Craig Knox, Yifeng Djoumbou, Tim Jewison, and 1 others. 2014. Drugbank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, 42(D1):D1091–D1097.

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

Xinyue Li, Zhankun Xiong, Wen Zhang, and Shichao Liu. 2024. Deep learning for drug-drug interaction prediction: A comprehensive review. *Quantitative Biology*, 12(1):30–52.

Derek Lim, T. Konstantin Rusch, Yu Huang, and Felipe Tang. 2022. Sign and basis invariant networks for spectral graph representation. In *Advances in Neural Information Processing Systems (NeurIPS)*. If you cite SignNet/BasisNet, align to the exact venue/version you use.

Mei Ma and Xiujuan Lei. 2023. A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLOS Computational Biology*, 19(1):e1010812.

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

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and 1 others. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8026–8037.

Bethany Percha and Russ B. Altman. 2013. Discovery and explanation of drug-drug interactions via text mining. In *Pacific Symposium on Biocomputing (PSB)*, pages 410–421.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. 2022. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Zhenqian Shen, Mingyang Zhou, Yongqi Zhang, and Quanming Yao. 2025. Benchmarking drug-drug interaction prediction methods: a perspective of distribution changes. *Bioinformatics*, page btaf569.

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

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2021. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24.

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

Zimo Yan, Jie Zhang, Zheng Xie, Chang Liu, Yizhen Liu, and Yiping Song. 2025a. Metamolgen: A neural graph motif generation model for de novo molecular design. *Preprint*, arXiv:2504.15587.

Zimo Yan, Jie Zhang, Zheng Xie, Yiping Song, and Hao Li. 2025b. A multi-scale graph neural process with cross-drug co-attention for drug-drug interactions prediction. *Preprint*, arXiv:2509.15256.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Tie-Yan Shen, and Yanming Liu. 2021. Do transformers really perform bad for graph representation? In *Advances in Neural Information Processing Systems (NeurIPS)*. Commonly referred to as Graphormer.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 974–983.

Marinka Zitnik, Monica Agrawal, and Jure Leskovec. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466.

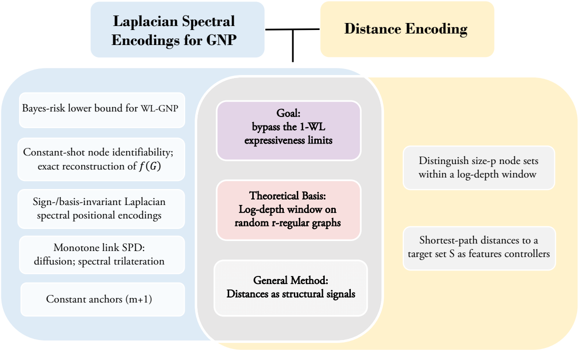

Figure 5: High-level comparison between Distance Encoding (DE) (2020) and our anchor-based spectral approach. Both operate in a logarithmic depth window on random r-regular graphs and exploit distance-derived signals to go beyond 1-WL message passing. DE focuses on distinguishing node sets of size p via distance features, while we study single-source identifiability ($p_{\text{task}}=1$) in a probabilistic (GNP) setting and use a constant number of anchors together with invariant spectral coordinates.

A Relation to Distance Encoding (DE) and the role of p

Distance Encoding (DE) (2020) constructs a permutation-invariant feature $\zeta(\cdot \mid S)$ based on distances to a target node set S and refers to *DE-p* when $|S| = p$ (cf. (2020, Def. 3.1)). A key result shows that, on (almost all) random r-regular graphs, a DE-GNN-p can distinguish two size-p node sets within a logarithmic depth window

$$L \;\le\; \left(\tfrac{1}{2}+\varepsilon\right)\frac{\log n}{\log(r-1)}, \tag{17}$$

whereas *DE-1* can fail on distance-regular graphs (see (2020, Thm. 3.7, Cor. 3.8)).

Task and mechanism differences. Our theoretical focus is *single-source node identifiability* ($p_{\text{task}} = 1$) under a probabilistic encoder class (GNPs), where the core difficulty is *posterior ambiguity* induced by WL-bounded expressiveness. Rather than encoding the hidden source set directly as in DE, we introduce a *constant-size* anchor set $S = \{a_i\}_{i=1}^{m+1}$ (thus $p_{\text{anchor}} = m+1$ fixed) and observe shortest-path distances from anchors to the hidden source. Conceptually, anchors provide a reference frame: shortest-path signals are mapped through a monotone linkage to a diffusion-like signal and combined with an invariant spectral coordinate system to support a reconstruction-style argument (trilateration).

How the role of p differs. In DE, p is the size of the target set to be distinguished and the result is stated in terms of distinguishing different size-p sets within the log-depth window. In our setting, p_{task} is fixed to one (a single hidden source), and the relevant design choice is instead the *anchor cardinality* $p_{\text{anchor}} = m{+}1$, which remains constant as n grows. This distinction matters because our guarantees are phrased as an identifiability (or Bayes-risk) statement for a probabilistic model class, rather than a purely discriminative distinguishability claim.

Takeaway. We view DE as complementary: it provides a principled way to enrich GNNs with distance-derived signals and gives strong distinguishability results in random regular graphs. Our analysis asks a different question-when a probabilistic graph neural process uses a WL-bounded encoder, what identifiability limitations follow, and what type of positional information suffices to remove ambiguity? The appendix comparison is intended to clarify that both approaches share the same "log-depth window" intuition but differ in objectives (set distinguishability vs. single-source identifiability) and in the role played by a fixed number of anchors together with invariant spectral coordinates.

B Supplementary Background and Experimental Details

This appendix collects technical background and implementation details that are referenced in the main text but omitted there for brevity. Full theoretical proofs are deferred to Appendix C.

B.1 A sign-/basis-invariant spectral positional encoding

Let $\{(\lambda_j, \phi_j)\}_{j=1}^{M}$ be the first M nontrivial eigenpairs of the chosen Laplacian. Define the truncated eigenvector coordinate

$$\phi^{(M)}(v) := (\phi_1(v), \dots, \phi_M(v))^\top \in \mathbb{R}^M. \tag{18}$$

Following a feature-level invariant construction, we form quadratic invariants

$$s(v) := \phi^{(M)}(v) \odot \phi^{(M)}(v) \in \mathbb{R}^M, \tag{19}$$

and

$$u(v) := \mathrm{vec}_{\triangle}\left(\left|\phi^{(M)}(v)\phi^{(M)}(v)^\top\right|\right) \in \mathbb{R}^{M(M-1)/2}, \tag{20}$$

where $\mathrm{vec}_{\triangle}(\cdot)$ stacks strict upper-triangular entries in a fixed order. The final encoding is

$$\Psi(v) := \mathrm{concat}\Big(\{\lambda_i\}_{i=1}^M,\ s(v),\ u(v)\Big). \quad (21)$$

This encoding depends only on eigenvalues and sign/rotation-invariant quadratic functions of eigenvector coordinates, and is therefore invariant to eigenvector sign flips and to orthogonal basis rotations within eigenspaces.

B.2 Implementation notes for spectral channels in DDI experiments

For the DDI ablations, we follow a common and lightweight practice: compute the first k nontrivial eigenvectors of the normalized Laplacian L_{sym} and concatenate them to the initial node features (we use $k = 16$). We also consider an "unsigned" spectral variant that uses eigenvectors of $L_{\mathrm{unsigned}} = L + 2A$ with the same concatenation strategy.

Because raw eigenvectors are not uniquely oriented (sign ambiguity), run-to-run variation may occur unless one enforces a deterministic sign convention or uses sign-invariant features (as in the theoretical Ψ construction). In our DDI ablations, we keep the PE pipeline fixed across runs, and we additionally probe sensitivity via controlled perturbations (sign flips / subspace rotations) in a separate analysis.

B.3 Graph notation and Laplacians

Let $G = (V, E)$ be a finite, simple, connected, undirected graph with $|V| = n$. Let $A \in \{0,1\}^{n\times n}$ be the adjacency matrix and D the diagonal degree matrix. We use the (symmetric) normalized Laplacian

$$L_{\mathrm{sym}} := I - D^{-1/2} A D^{-1/2}, \quad (22)$$

and the random-walk Laplacian $L_{\mathrm{rw}} := I - P$ with $P := D^{-1}A$. On r-regular graphs ($D = rI$), these coincide and reduce to

$$L_{\mathrm{sym}} = L_{\mathrm{rw}} = I - \tfrac{1}{r}A \ =:\ L, \quad (23)$$

which is the form used throughout our tree/random-regular arguments.

For $u, v \in V$, $\mathrm{SPD}(u, v)$ denotes the shortest-path distance and $\mathcal{N}(u)$ the neighbor set. For $R \in \mathbb{N}$, define the radius-R ball

$$B_G(u, R) := \{v \in V : \mathrm{SPD}(u, v) \le R\}. \quad (24)$$

We write $\mathrm{diam}(G)$ for the diameter and use w.h.p. for probability $1 - o(1)$ as $n \to \infty$.

B.4 Heat kernel and diffusion distance

For $t > 0$, define the heat semigroup and kernel

$$K_t := e^{-tL}, \quad k_t(u, v) := (K_t)_{uv}. \quad (25)$$

Let $\{(\lambda_j, \phi_j)\}_{j\ge1}$ be the nonzero eigenpairs of L, with $\{\phi_j\}$ orthonormal in $\ell^2(V)$. The diffusion distance is

$$d_t(u, v)^2 = \sum_{j\ge1} e^{-2t\lambda_j}\big(\phi_j(u) - \phi_j(v)\big)^2. \quad (26)$$

Using the spectral expansion $K_t = \sum_{j\ge1} e^{-t\lambda_j}\phi_j\phi_j^\top$, one obtains the standard kernel identity

$$d_t(u, v)^2 = k_{2t}(u, u) + k_{2t}(v, v) - 2k_{2t}(u, v). \quad (27)$$

On vertex-transitive graphs (e.g., the infinite r-regular tree), $k_{2t}(u, u)$ is constant in u, which simplifies (27) and is convenient for establishing monotone linkage between SPD and d_t on the tree; see Appendix C for full proofs used in the main theorem.

Poissonization identity (regular graphs). When $L = I - \frac{1}{r}A$ and $P := \frac{1}{r}A$, we use

$$K_t \ = \ e^{-tL} \ = \ e^{-t}\, e^{tP}. \quad (28)$$

This representation is useful to compare heat kernels (and thus diffusion distances) between a random regular graph and its tree cover within logarithmic radii.

B.5 Spectral positional encoding (PE) used by Lap-GNP

Let $\{(\lambda_j, \phi_j)\}_{j=1}^M$ be the first M nontrivial eigenpairs of the chosen Laplacian. Define $\phi^{(M)}(v) := (\phi_1(v), \ldots, \phi_M(v))^\top \in \mathbb{R}^M$ and

$$s(v) := \phi^{(M)}(v) \odot \phi^{(M)}(v) \in \mathbb{R}^M, \quad (29)$$

$$u(v) := \mathrm{vec}_{\triangle}\Big(\big|\phi^{(M)}(v)\,\phi^{(M)}(v)^\top\big|\Big) \in \mathbb{R}^{\frac{M(M-1)}{2}}, \quad (30)$$

where $\mathrm{vec}_{\triangle}(\cdot)$ stacks strict upper-triangular entries in a fixed order. The final encoding is

$$\Psi(v) := \mathrm{concat}\Big(\{\lambda_i\}_{i=1}^M,\ s(v),\ u(v)\Big). \quad (31)$$

This construction is invariant to eigenvector sign flips and to orthogonal basis rotations within eigenspaces because it depends only on eigenvalues and sign/rotation-invariant quadratic features of eigenvector coordinates. A formal statement and proof can be placed in Appendix C if desired.

B.6 Why random r-regular graphs are a hard regime

We analyze $G \sim \mathcal{G}_{n,r}$ (uniform random labeled r-regular graphs with fixed $r \geq 3$) as a canonical hard family for 1-WL-bounded message passing with trivial initial features: graphs are highly symmetric and locally tree-like up to logarithmic radius. Concretely, there exists $\varepsilon_0 > 0$ such that for

$$T \leq \left(\tfrac{1}{2} - \varepsilon_0\right) \frac{\log n}{\log(r-1)}, \tag{32}$$

a uniformly random root u has $B_G(u, T)$ being a tree w.h.p., and moreover all but $o(n)$ vertices satisfy this property w.h.p. This logarithmic treelike window underpins the stability arguments that compare diffusion geometry on G to that on the infinite r-regular tree.

B.7 Experimental setup for DDI link prediction

Task and dataset. We follow the inductive drug-drug interaction (DDI) link prediction protocol on the **DrugBank** dataset used by Yan et al. (2025b). The goal is to predict whether a candidate drug pair forms an interaction edge under an inductive split (test drugs unseen during training).

Implementation. All models are implemented in PyTorch (2019) and PyTorch Geometric (2019). We focus on ablations that isolate the impact of explicit positional information while keeping the backbone identical.

Compared variants (ablation on positional encodings).

- **Baseline:** MPNP-DDI (2025b) without explicit positional encodings.
- **Laplacian PE:** Baseline + $k = 16$ eigenvectors of the normalized Laplacian L_{sym}, concatenated to initial node features.
- **Unsigned Laplacian PE:** Baseline + $k = 16$ eigenvectors of the unsigned Laplacian $L_{\text{unsigned}} = L + 2A$ (same concatenation strategy).

PE computation details (for reproducibility). For each molecular graph, we compute the first k *nontrivial* eigenvectors (excluding the constant eigenvector when applicable) via a standard symmetric eigensolver. Eigenvectors are used as node-level positional channels and concatenated to the original node feature vector before message passing. If eigenvectors are not uniquely oriented (sign ambiguity), models may experience run-to-run variation unless a deterministic sign convention is imposed or sign-invariant features are used (as in our theoretical Ψ construction in (31)). In the DDI ablation above, we follow the common practice in the cited baseline and keep the PE pipeline fixed across runs.

Architecture. All variants share the same architecture for fair comparison: node and edge features are projected to a hidden dimension of 64. The model stacks three Graph Neural Process blocks; each block performs two message-passing iterations to capture multi-scale representations.

Optimization. We use AdamW with learning rate 1×10^{-4} and weight decay 5×10^{-5}. A cosine annealing schedule is applied over 50 epochs. We train with batch size 32 and gradient accumulation over 4 steps (effective batch size 128), which stabilizes training under memory constraints.

Evaluation and model selection. We use validation AUROC for early stopping and model selection; training stops if validation AUROC does not improve for 10 consecutive epochs. We report final performance of the best checkpoint on a held-out test set.

Code availability. We provide an anonymized code repository for reproducibility: `https://anonymous.4open.science/r/LapGNP-2EB6`.

C Proofs for Section 4

This appendix contains the full proofs deferred from Section 4. Throughout, $r \geq 3$ is fixed and $G \sim \mathcal{G}_{n,r}$ denotes a uniformly random labeled r-regular graph on $[n]$.

C.1 Treelike window on $\mathcal{G}_{n,r}$ and seed-restricted union bounds

We record the locally tree-like property in the exact form used in the main text: it holds for a single root, and simultaneously for any *prescribed* polylogarithmic set of seeds; we do not claim it uniformly over all vertices.

Lemma 10. *There exists $\epsilon_0 > 0$ such that for any radius*

$$R \;\leq\; \left(\tfrac{1}{2} - \epsilon_0\right) \frac{\log n}{\log(r-1)}, \tag{33}$$

the following holds. For every prescribed seed set $S \subseteq V$ *with* $|S| \le (\log n)^C$ *for a constant* $C > 0$*,*

$$\Pr\Big(\forall u \in S,\ B_G(u,R) \textit{ is a tree}\Big) \;=\; 1 - o(1). \tag{34}$$

Proof. Fix a vertex u. In the configuration model exposure process, the breadth-first exploration of $B_G(u,R)$ reveals at most $1 + r\sum_{\ell=0}^{R-1}(r-1)^{\ell} = O((r-1)^R)$ vertices and $O((r-1)^R)$ half-edges. A cycle occurs only if two revealed half-edges are paired within the revealed set, which has probability

$$\Pr\big(B_G(u,R) \text{ contains a cycle}\big) \;\le\; O\Big(\frac{(r-1)^{2R}}{n}\Big). \tag{35}$$

Under (33), $(r-1)^{2R} \le n^{1-2\epsilon_0}$, so the bound is $O(n^{-2\epsilon_0}) = o(1)$. For a prescribed S with $|S| \le (\log n)^C$, a union bound gives

$$\begin{aligned}&\Pr\Big(\exists u \in S:\ B_G(u,R) \text{ contains a cycle}\Big)\\ &\le |S|\cdot O(n^{-2\epsilon_0}) \;=\; (\log n)^C \cdot O(n^{-2\epsilon_0}) \;=\; o(1).\end{aligned} \tag{36}$$

□

Remark 11 (Quantifier discipline)**.** *At radii* $R \asymp \frac{1}{2}\log_{r-1} n$*, one cannot have* $B_G(u,R)$ *tree for all* u *w.h.p. because* $\Theta(\log n)$*-cycles occur. Lemma 10 is the seed-restricted statement used throughout.*

C.2 WL buckets and the lower bound

We prove Lemmas 3-4 and Proposition 5 from the main text.

Proof of Lemma 3. Let $D := \operatorname{diam}(G)$. It is standard that $D = \Theta(\log n)$ w.h.p. for $G \sim \mathcal{G}_{n,r}$. Condition on $(G, \{v_i\}_{i=1}^k)$. The distance-key vector

$$\mathbf{d}(u) = \big(\mathrm{SPD}(u,v_1),\dots,\mathrm{SPD}(u,v_k)\big) \tag{37}$$

takes values in $\{0,1,\dots,D\}^k$, hence the number of realized keys satisfies $|\Im(\mathbf{d})| \le (D+1)^k$. Therefore the number of singleton buckets is at most $(D+1)^k$. Since v_0 is uniform,

$$\Pr\big(|\{u : \mathbf{d}(u) = \mathbf{d}(v_0)\}| = 1 \,\big|\, G, \{v_i\}\big) \le \frac{(D+1)^k}{n}. \tag{38}$$

With $D \le C_r \log n$ w.h.p. and $k = o(\log n)$,

$$\begin{aligned}\frac{(D+1)^k}{n} \le \frac{(C_r \log n)^k}{n} &= \frac{\exp(k\log\log n + O(k))}{n}\\ &= n^{-1+o(1)} = o(1).\end{aligned} \tag{39}$$

Thus $\Pr(|\{u : \mathbf{d}(u) = \mathbf{d}(v_0)\}| \ge 2) = 1 - o(1)$. □

Proof of Lemma 4. On an r-regular graph with trivial initial node features, the initial 1-WL colors are identical. A 1-WL-bounded encoder is a deterministic function of the rooted computation tree and the multiset of neighbor colors at each depth, together with the marked locations of the context vertices. If $\mathbf{d}(u) = \mathbf{d}(u')$, then each context vertex v_i lies in the same distance shell around u and around u'. By regularity and identical initial colors, the rooted computation trees are isomorphic at every depth and the marking pattern is identical, hence 1-WL assigns identical color sequences to u and u'. Therefore any 1-WL-bounded encoder outputs identical embeddings for u and u' and cannot distinguish them given the context. □

Proof of Proposition 5. By Lemma 3, with probability $1 - o(1)$ the bucket containing v_0 has size at least 2. By Lemma 4, WLGNPs cannot distinguish nodes within this bucket, so the WLGNP indistinguishability class $[v_0]$ satisfies $|[v_0]| \ge 2$ on that event.

Using the Bayes-risk lower bound,

$$\Pr(\widehat{v} \ne v_0 \mid Z, R) \;\ge\; 1 - \mathbf{E}\big[|[v_0]|^{-1} \mid Z, R\big]. \tag{40}$$

Taking expectations gives $\mathrm{Risk}(\delta) \ge 1 - \mathbf{E}[|[v_0]|^{-1}]$. Since $|[v_0]|^{-1} \le \frac{1}{2}$ on $\{|[v_0]| \ge 2\}$,

$$\begin{aligned}\mathbf{E}[|[v_0]|^{-1}] &\le \Pr(|[v_0]| = 1) + \tfrac{1}{2}\Pr(|[v_0]| \ge 2)\\ &\le \tfrac{1}{2} + o(1),\end{aligned} \tag{41}$$

and therefore $\mathrm{Risk}(\delta) \ge \frac{1}{2} - o(1)$. □

C.3 Monotone diffusion-distance linkage: tree monotonicity and quantitative transfer

This subsection proves Theorem 2 from the main text, and thus implies Corollary 1. We rely on the heat-kernel identities stated in Section 3 (not repeated here).

Lemma 12. *Fix* $r \ge 3$ *and* $t > 0$*. Let* $(X_s)_{s\ge 0}$ *be the continuous-time simple random walk on the infinite* r*-regular tree* $\mathcal{T}_r$ *with generator* $P - I$ *(equivalently, heat semigroup* $K_t = e^{-t(I-P)}$*), and let* $h_t(d) := k_t(o,x)$ *for any* $x \in \mathcal{T}_r$ *with* $\mathrm{SPD}(o,x) = d$*. Then* $h_t(d)$ *is well-defined (radial), and for every* $t > 0$ *it is strictly decreasing in* d*:*

Consequently, writing $\kappa_{2t} := k_{2t}(o,o)$ *and* $p_{2t}(d) := h_{2t}(d)$*, the diffusion distance satisfies*

$$d_t(o,x)^2 \;=\; 2\big(\kappa_{2t} - p_{2t}(\mathrm{SPD}(o,x))\big), \tag{42}$$

and the map $d \mapsto d_t(o,x)$ is strictly increasing in $d = \mathrm{SPD}(o,x)$.

Proof. Radiality follows from vertex-transitivity: for any two vertices x, y at the same distance from o, there exists an automorphism of $\mathcal{T}_r$ fixing o and mapping x to y, and the walk (hence $k_t(o,\cdot)$) is invariant under such automorphisms. Thus $h_t(d)$ is well-defined.

We next derive the forward equations for $h_t(d)$. Since the chain has total jump rate 1 and, at each jump, moves to a uniformly random neighbor, the Kolmogorov forward equation for a vertex x at distance $d \geq 1$ reads

$$\partial_t k_t(o,x) = -k_t(o,x) + \frac{1}{r}\sum_{y\sim x} k_t(o,y). \tag{43}$$

Among the r neighbors of x, exactly one lies at distance $d-1$ from o and the remaining $r-1$ neighbors lie at distance $d+1$. By radiality, for $d \geq 1$,

$$\partial_t h_t(d) = -h_t(d) + \frac{1}{r}h_t(d-1) + \frac{r-1}{r}h_t(d+1). \tag{44}$$

At the root o (distance 0), all r neighbors lie at distance 1, so

$$\partial_t h_t(0) = -h_t(0) + h_t(1). \tag{45}$$

The initial condition is $h_0(0) = 1$ and $h_0(d) = 0$ for all $d \geq 1$.

Define the discrete gradient

$$s_t(d) \;:=\; h_t(d) - h_t(d+1), \quad d \geq 0. \tag{46}$$

We claim that $s_t(d) > 0$ for every $t > 0$ and every $d \geq 0$, which implies strict monotonicity of $h_t(\cdot)$.

Subtracting the forward equations at d and $d+1$ yields, for all $d \geq 1$,

$$\begin{aligned}\partial_t s_t(d) =& \partial_t h_t(d) - \partial_t h_t(d+1)\\ =& \Big(-h_t(d) + \tfrac{1}{r}h_t(d-1) + \tfrac{r-1}{r}h_t(d+1)\Big)\\ & - \Big(-h_t(d+1) + \tfrac{1}{r}h_t(d) + \tfrac{r-1}{r}h_t(d+2)\Big)\\ =& -\big(h_t(d) - h_t(d+1)\big) + \tfrac{1}{r}\big(h_t(d-1) - h_t(d)\big)\\ & + \tfrac{r-1}{r}\big(h_t(d+1) - h_t(d+2)\big)\\ =& -s_t(d) + \tfrac{1}{r}s_t(d-1) + \tfrac{r-1}{r}s_t(d+1).\end{aligned} \tag{47}$$

At the boundary $d = 0$, using $\partial_t h_t(0) = -h_t(0) + h_t(1)$ and the $d = 1$ equation gives

$$\begin{aligned}\partial_t s_t(0) =& \partial_t h_t(0) - \partial_t h_t(1)\\ =& \big(-h_t(0) + h_t(1)\big)\\ & - \Big(-h_t(1) + \tfrac{1}{r}h_t(0) + \tfrac{r-1}{r}h_t(2)\Big)\\ =& -\tfrac{r+1}{r}\,s_t(0) + \tfrac{r-1}{r}\,s_t(1).\end{aligned} \tag{48}$$

Thus $s_t(\cdot)$ satisfies a linear ODE system $\partial_t s_t = A s_t$ on $Z_{\geq 0}$ where the matrix A is tridiagonal with strictly positive off-diagonal entries:

$$\begin{aligned}&(As)(0) = -\tfrac{r+1}{r}s(0) + \tfrac{r-1}{r}s(1),\\ &(As)(d) = \tfrac{1}{r}s(d-1) - s(d) + \tfrac{r-1}{r}s(d+1) \;\; (d \geq 1).\end{aligned} \tag{49}$$

Moreover, the initial data are

$$s_0(0) = h_0(0) - h_0(1) = 1, \quad s_0(d) = 0 \;\; (d \geq 1). \tag{50}$$

We now show that for every $t > 0$ and every $d \geq 0$, one has $s_t(d) > 0$. Since $s_t = e^{tA}s_0$ and $s_0 = e_0$ (the unit mass at 0), it suffices to show that the matrix exponential has strictly positive entries in the first column: $(e^{tA})_{d0} > 0$ for all d and all $t > 0$.

By the power-series expansion,

$$(e^{tA})_{d0} = \sum_{m=0}^{\infty}\frac{t^m}{m!}(A^m)_{d0}. \tag{51}$$

For each $d \geq 1$, consider the term $m = d$. Because $A_{j,j-1} = \frac{1}{r} > 0$ for all $j \geq 1$, the product along the path $0 \to 1 \to \cdots \to d$ contributes

$$(A^d)_{d0} \;\geq\; A_{1,0}\,A_{2,1}\cdots A_{d,d-1} \;=\; \Big(\tfrac{1}{r}\Big)^d \;>\; 0. \tag{52}$$

Therefore $(e^{tA})_{d0} \geq \frac{t^d}{d!}(A^d)_{d0} > 0$ for all $t > 0$ and all $d \geq 1$. For $d = 0$, clearly $(e^{tA})_{00} > 0$. Hence $s_t(d) = (e^{tA})_{d0} > 0$ for every $t > 0$ and every $d \geq 0$, proving $h_t(d) > h_t(d+1)$ strictly.

Finally, on $\mathcal{T}_r$ the diffusion-distance kernel identity simplifies by vertex-transitivity:

$$\begin{aligned}d_t(o,x)^2 &= k_{2t}(o,o) + k_{2t}(x,x) - 2k_{2t}(o,x)\\ &= 2\big(\kappa_{2t} - h_{2t}(\mathrm{SPD}(o,x))\big),\end{aligned} \tag{53}$$

and since $h_{2t}(d)$ is strictly decreasing, $d \mapsto d_t(o,x)$ is strictly increasing. ☐

Lemma 13. *Fix $t > 0$ and $r \geq 3$. On the infinite r-regular tree $\mathcal{T}_r$, $k_t(x,y)$ is radial and strictly decreases in $d = \mathrm{SPD}(x,y)$; consequently $d_t(x,y)$ is strictly increasing in d. Moreover, let*

$G \sim \mathcal{G}_{n,r}$ and let S be a prescribed seed set with $|S| \le (\log n)^C$. For R satisfying (33), with probability $1 - o(1)$ the following holds simultaneously for all prescribed pairs (u, v) with $u \in S$ and $\mathrm{SPD}_G(u, v) \le R$:

$$\left|\, k_{2t}^{G}(u,v) - k_{2t}^{\mathcal{T}_r}(u',v') \,\right| \;\le\; \delta_n, \tag{54}$$

where $u', v' \in \mathcal{T}_r$ satisfy $\mathrm{SPD}_{\mathcal{T}_r}(u', v') = \mathrm{SPD}_G(u, v)$ and

$$\begin{aligned}\delta_n \le{} & \Pr\Big(\exists s \in S:\ B_G(s,R)\Big) \\ & + \Pr\big(\mathrm{Poisson}(2t) \ge R+1\big).\end{aligned} \tag{55}$$

In particular, $\delta_n = o(1)$, and thus $d_t^G(u,v) = \psi(\mathrm{SPD}_G(u,v)) \pm o(1)$ for $\psi(d) = \sqrt{2(\kappa_{2t} - p_{2t}(d))}$, which yields $d_t^G(u,v) = \psi(\mathrm{SPD}_G(u,v)) \pm o(1)$ for all such queried pairs. In particular, ψ is strictly increasing (by Lemma 12), and this establishes Theorem 2.

Proof. On $\mathcal{T}_r$, radiality follows from automorphism invariance. Strict monotonicity of $d \mapsto k_t(d)$ can be shown by analyzing the radial birth-death chain and Poissonization; we omit repetitions of standard forward-equation steps and only use the conclusion that $p_t(d)$ is strictly decreasing in d, hence d_t strictly increasing.

For the transfer bound, fix R as in (33) and a prescribed seed set S with $|S| \le (\log n)^C$. On the event $\mathcal{E}_{\text{tree}} := \{\forall s \in S,\ B_G(s,R) \text{ is a tree}\}$, each rooted ball $(B_G(s,R), s)$ is isomorphic to the rooted depth-R tree in $\mathcal{T}_r$. Fix a prescribed pair (u, v) with $u \in S$ and $\mathrm{SPD}_G(u,v) \le R$. Let $(X_s)_{s\ge 0}$ be the continuous-time walk on G started at u. Up to time $2t$, the number of jumps N satisfies $N \sim \mathrm{Poisson}(2t)$. If $N \le R$, then the walk stays within $B_G(u,R)$ and its entire path distribution matches the corresponding walk on $\mathcal{T}_r$ under the rooted isomorphism. Hence

$$\left|k_{2t}^{G}(u,v) - k_{2t}^{\mathcal{T}_r}(u',v')\right| \;\le\; \Pr(\mathcal{E}_{\text{tree}}^c) + \Pr(N \ge R+1), \tag{56}$$

which is exactly (54) with (55). Taking the union bound over the prescribed polylog-many queried pairs keeps the same upper bound because the right-hand side is already global in S and N.

Finally, Lemma 10 gives $\Pr(\mathcal{E}_{\text{tree}}^c) = o(1)$. For the Poisson tail we may use the standard bound (for $R \to \infty$)

$$\Pr(N \ge R+1) \;\le\; \left(\frac{e \cdot 2t}{R+1}\right)^{R+1}, \tag{57}$$

so $\Pr(N \ge R+1) = o(1)$ since $R = \Theta(\log n) \to \infty$. Thus $\delta_n = o(1)$.

Converting the kernel bound to diffusion distance uses the kernel identity $d_t^2(u,v) = k_{2t}(u,u) + k_{2t}(v,v) - 2k_{2t}(u,v)$ and the fact that (u,u) and (v,v) are also among the seed-restricted queried pairs; this yields

$$d_t^G(u,v) = \psi(\mathrm{SPD}_G(u,v)) \pm o(1), \tag{58}$$

where $\psi(d) = \sqrt{2(\kappa_{2t} - p_{2t}(d))}$ is strictly increasing by Lemma 12. This establishes Theorem 2.

□

Proof of Theorem 2. Let S be the prescribed seed set consisting of all anchor nodes queried by the procedure (and, if needed, the corresponding nodes for diagonal terms (u,u) and (v,v) in the kernel identity). Apply Lemma 13 with this seed set S and radius R satisfying (33). Then, with probability $1 - o(1)$, for every prescribed queried pair (u,v) with $\mathrm{SPD}_G(u,v) \le R$,

$$d_t^G(u,v) = \psi(\mathrm{SPD}_G(u,v)) \pm o(1), \tag{59}$$

where $\psi(d) = \sqrt{2(\kappa_{2t} - p_{2t}(d))}$ is strictly increasing by Lemma 12. This implies Theorem 2. □

C.4 Spectral trilateration

We give a continuous proof of Lemma 6. For clarity we state the linear system explicitly.

Lemma 14. *Let $p_i \in \mathbf{R}^m$ $(i = 1, \dots, m+1)$ be affinely independent and let $z \in \mathbf{R}^m$ satisfy $\|z - p_i\|_2 = r_i$ for $i = 1, \dots, m+1$. Then z is uniquely determined and is the unique solution to $Az = b$, where*

$$A := 2\begin{bmatrix}(p_1 - p_{m+1})^\top \\ \vdots \\ (p_m - p_{m+1})^\top\end{bmatrix},$$
$$b := \begin{bmatrix}\|p_1\|_2^2 - \|p_{m+1}\|_2^2 + r_{m+1}^2 - r_1^2 \\ \vdots \\ \|p_m\|_2^2 - \|p_{m+1}\|_2^2 + r_{m+1}^2 - r_m^2\end{bmatrix}. \tag{60}$$

Proof of Lemma 14. Write $\|z - p\|_2^2 = \|z\|_2^2 - 2\langle z, p\rangle + \|p\|_2^2$. Subtract the $(m+1)$-st equation from the i-th equation for $i \le m$:

$$\|z - p_i\|_2^2 - \|z - p_{m+1}\|_2^2 = r_i^2 - r_{m+1}^2. \tag{61}$$

The $\|z\|_2^2$ terms cancel and for $i = 1, \dots, m$ we obtain

$$2\langle z, p_i - p_{m+1}\rangle = \|p_i\|_2^2 - \|p_{m+1}\|_2^2 + r_{m+1}^2 - r_i^2, \tag{62}$$

which is exactly the linear system (60). Affine independence means $\{p_i - p_{m+1}\}_{i=1}^m$ is a basis of $\mathbf{R}^m$, hence A is invertible and the solution is unique. ☐

Proof of Lemma 6. Apply Lemma 14 with $p_i = \Phi_t^{(m)}(a_i)$ and $z = \Phi_t^{(m)}(v_0)$, where the distances are $r_i = d_t^{(m)}(a_i, v_0) = \|\Phi_t^{(m)}(a_i) - \Phi_t^{(m)}(v_0)\|_2$. Assumption 2 gives affine independence w.h.p., hence uniqueness follows. ☐

C.5 Quantitative injectivity under a random-wave surrogate

This subsection verifies Theorem 9 under a random-wave surrogate (as stated in your theory section).

Assumption 3. *Fix M. For each $v \in V$, the vector $g(v) = (\phi_1(v), \dots, \phi_M(v)) \in \mathbf{R}^M$ has an absolutely continuous density, is isotropic subgaussian with bounded parameter, and $\{g(v)\}_{v\in V}$ are independent. Coordinates across j are independent for each fixed v.*

Theorem 15. *Assume 3. There exist absolute constants $c_0, C_0 > 0$ such that for any distinct $u, v \in V$ and any $\varepsilon \in (0,1)$,*

$$\Pr\big(\|\chi(u) - \chi(v)\|_2 \le \varepsilon\big) \;\le\; (C_0\varepsilon)^{c_0 M}, \tag{63}$$

where χ is the node-dependent part of Ψ (squares and absolute products). Consequently, if $M \ge C \log n$ with $C > 2/c_0$, then for any fixed $\alpha \in (0, Cc_0/3)$,

$$\begin{aligned} &\Pr\big(\exists u \neq v:\ \chi(u) = \chi(v)\big) = o(1), \\ &\min_{u\neq v} \|\chi(u) - \chi(v)\|_2 \;\ge\; n^{-\alpha}, \textit{w.h.p.} \end{aligned} \tag{64}$$

Proof. Identical to the standard product small-ball argument in your draft: use the independent coordinates $Z_j = X_j^2 - Y_j^2$ to obtain a one-dimensional quadratic small-ball bound and take products across j, then union bound over $\le n^2$ pairs and set $\varepsilon = n^{-\alpha}$ with $M = \Theta(\log n)$. ☐

Corollary 2 (Verification of Theorem 9)**.** *Under Assumption 3, Theorem 15 implies Theorem 9.*

Proof. Ψ is obtained by concatenating the (global) eigenvalue list with $\chi(v)$, so injectivity/separation of χ implies the same for Ψ. ☐

C.6 A separation lemma for the truncated embedding

To avoid relying on informal "the same mechanism applies" arguments, we record a direct separation lemma for the truncated embedding used by trilateration.

Lemma 16. *Fix $m \ge 2$ and $t > 0$. Suppose $\{\Phi_t^{(m)}(v)\}_{v\in V}$ are i.i.d. in $\mathbf{R}^m$ with an absolutely continuous density bounded by a constant (this holds under standard Gaussian/subgaussian random-wave surrogates for the first m eigenvector coordinates). Then there exists $\beta = \beta(m) > 0$ such that with probability* $1 - o(1)$

$$\min_{u\neq v} \|\Phi_t^{(m)}(u) - \Phi_t^{(m)}(v)\|_2 \;\ge\; n^{-\beta}. \tag{65}$$

Proof. Let $X_1, \dots, X_n$ be i.i.d. in $\mathbf{R}^m$ with density bounded by K. For any fixed $\varepsilon > 0$ and any pair $i \neq j$,

$$\begin{aligned} \Pr(\|X_i - X_j\|_2 \le \varepsilon) &\le \sup_x \Pr(X_j \in B(x, \varepsilon)) \\ &\le K \cdot \mathrm{Vol}(B(0,\varepsilon)) = K c_m \varepsilon^m, \end{aligned} \tag{66}$$

where c_m is the volume of the unit m-ball. By a union bound over $\binom{n}{2}$ pairs,

$$\Pr\Big(\min_{i\neq j} \|X_i - X_j\|_2 \le \varepsilon\Big) \le n^2 \cdot K c_m \varepsilon^m. \tag{67}$$

Taking $\varepsilon = n^{-\beta}$ with any $\beta > 2/m$ makes the right-hand side $o(1)$, yielding the claim. ☐

Remark 17. *Lemma 16 is independent of Theorem 9; it directly guarantees the uniqueness of a nearest-neighbor decoder in $\Phi_t^{(m)}$ once trilateration returns an $o(1)$-accurate coordinate. If you prefer to keep everything tied to Ψ, you can alternatively decode by nearest neighbor in Ψ provided you reconstruct (or approximate) the corresponding $\Psi(v_0)$; in the present pipeline, trilateration naturally produces $\Phi_t^{(m)}(v_0)$, hence we state separation directly for $\Phi_t^{(m)}$.*

C.7 Proof of the main separation theorem

Proof of Theorem 7. Part (i) is Proposition 5 proved in Section C.2.

For part (ii), with Proposition 1, Theorem 2, and Assumptions 2. Let $k \ge k_0 = m + 1$ and write the context as $\mathcal{C} = \{(a_i, y_i)\}_{i=1}^k$ with $y_i = \mathrm{SPD}(a_i, v_0)$. Fix $m+1$ anchors among them and relabel as $a_1, \dots, a_{m+1}$. By Assumption 2, the points $p_i := \Phi_t^{(m)}(a_i)$ are affinely independent with probability $1 - o(1)$, hence the trilateration linear system (Lemma 14) is well-posed.

By Theorem 2, there exists a strictly increasing $\psi : [0, R] \to \mathbf{R}_+$ such that, for the prescribed anchor-source pairs with $\mathrm{SPD}(a_i, v_0) \leq R$,

$$d_t(a_i, v_0) = \psi(\mathrm{SPD}(a_i, v_0)) \pm o(1). \tag{68}$$

Define the proxy radii $\widehat{r}_i := \psi(\mathrm{SPD}(a_i, v_0)) = \psi(y_i)$, so that

$$\left|\widehat{r}_i - d_t(a_i, v_0)\right| = o(1),\ i = 1, \ldots, m+1. \tag{69}$$

Applying Lemma 14 with $p_i = \Phi_t^{(m)}(a_i)$ and radii $r_i = d_t^{(m)}(a_i, v_0)$ shows that $\Phi_t^{(m)}(v_0)$ is uniquely determined from the true radii. Moreover, since the system matrix A in (60) is fixed-size and invertible, replacing the true radii by the $o(1)$-perturbed radii $\widehat{r}$ changes the right-hand side b by $o(1)$, hence the recovered coordinate $\widehat{z}$ satisfies

$$\|\widehat{z} - \Phi_t^{(m)}(v_0)\|_2 = o(1). \tag{70}$$

To identify v_0, decode by nearest neighbor in the truncated embedding:

$$\widehat{v} \in \arg\min_{w \in V} \|\Phi_t^{(m)}(w) - \widehat{z}\|_2. \tag{71}$$

This step requires a minimum-separation condition for $\{\Phi_t^{(m)}(v)\}_{v \in V}$; under the same random-wave (continuous) surrogate, Lemma 16 gives

$$\Delta_m := \min_{u \neq v} \|\Phi_t^{(m)}(u) - \Phi_t^{(m)}(v)\|_2 \geq n^{-\beta} \quad \text{w.h.p.} \tag{72}$$

For n large, the $o(1)$ reconstruction error is $< \Delta_m/2$, which implies that v_0 is the unique minimizer and therefore $\widehat{v} = v_0$.

Finally, once v_0 is identified, the function is recovered everywhere by $\widehat{f}(v) = \mathrm{SPD}(v, \widehat{v}) = \mathrm{SPD}(v, v_0)$. $\square$

D Additional Robustness Results: Calibration Metrics

We report calibration robustness on DDI under the same corruption protocol as Table 1. Entries are mean±sd (reference values for formatting; replace with measured sd in the final version). Lower is better for both ECE and AURC.

E Large-scale Case Study on *web-Google*

We include this appendix experiment to demonstrate that Laplacian-based spectral coordinates remain practical at web scale, and that simple frequency-band and diffusion-time choices provide a controllable inductive bias. We use the SNAP *web-Google* graph and restrict to its largest connected component (LCC), containing 855,802 nodes.

Setup. We compare a baseline PLAIN model (same backbone without spectral positional information) against LAP, which augments node representations with sign-/basis-invariant Laplacian spectral encodings computed from a fixed number of eigenpairs. We use an m-dimensional truncated diffusion-style embedding (cf. the main text), and vary two hyperparameters that control scale: (i) a *spectral band* (a contiguous range of eigenindices), (ii) the diffusion time t (larger t emphasizes smoother, more global structure). We also test a lightweight standardization option (`Std.`) applied to the spectral coordinates (either no standardization, or mean-centering), and keep all other training details consistent with the main experimental protocol.

Evaluation. We report three complementary diagnostics that probe global-to-local geometry: (i) Spearman ρ, which measures rank consistency under a global ordering signal (higher indicates better monotone global organization); (ii) Triplet accuracy, which evaluates relative distance comparisons in the learned geometry (higher indicates better local consistency of the metric structure); (iii) ΔRecall@10, the improvement in local retrieval (Recall@10) relative to PLAIN (positive values indicate better neighborhood retrieval). All reported numbers are LAP minus PLAIN.

Results. Table 4 shows that LAP yields consistent gains over PLAIN across these diagnostics. Using a deeper spectral band together with a larger diffusion time improves local retrieval more strongly (Recall@10), while Spearman and Triplet accuracy also increase modestly. This pattern is consistent with the interpretation that higher-frequency bands and larger t provide an adjustable bias toward multiscale separability: the representation can emphasize either global ordering or finer local distinctions through the choice of spectral region and diffusion smoothing.

Table 3: Robustness on DDI (calibration metrics). Entries are mean±sd (reference values for formatting; replace with measured sd in the final version). Columns 0 and 5 denote severity 0 (clean) and 5 (worst). Δ is computed on the means (sev5−sev0). “Perturb.” matches Table 1. Gray cells indicate the best mean at sev$=5$ within each corruption.

Corruption	Perturb. (sev=5)	Method	ECE ↓			AURC ↓		
			0	5	Δ	0	5	Δ
EDGE_DROP	$p_{\text{drop}} = 0.50$	Baseline (No PE)	0.377±0.006	0.378±0.006	+0.001	0.327±0.012	0.349±0.014	+0.022
		Laplacian PE	0.382±0.005	0.378±0.006	-0.004	0.167±0.010	0.241±0.015	+0.074
		Unsigned Laplacian PE	0.371±0.005	0.370±0.005	-0.001	0.146±0.009	0.218±0.014	+0.072
LABEL_FLIP	$p_{\text{flip}} = 0.50$	Baseline (No PE)	0.321±0.010	0.320±0.010	-0.001	0.136±0.020	0.290±0.028	+0.154
		Laplacian PE	0.373±0.008	0.374±0.009	+0.001	0.091±0.012	0.252±0.022	+0.161
		Unsigned Laplacian PE	0.366±0.008	0.369±0.009	+0.003	0.096±0.012	0.254±0.022	+0.158
PE_SIGNFLIP	$p_{\text{sign}} = 1.0$	Laplacian PE	0.373±0.005	0.370±0.005	-0.003	0.092±0.006	0.093±0.006	+0.001
		Unsigned Laplacian PE	0.367±0.005	0.366±0.005	-0.001	0.096±0.006	0.097±0.006	+0.001
PE_SUBSPACE_ROT	$\gamma = 1.0$ (Haar)	Laplacian PE	0.344±0.004	0.344±0.004	+0.000	0.282±0.010	0.280±0.010	-0.002
		Unsigned Laplacian PE	0.341±0.004	0.342±0.004	+0.001	0.291±0.010	0.290±0.010	-0.001

Table 4: Plain vs. LAP on *web-Google*. LAP tunes multi-scale geometry via spectral band and diffusion time t. Values are reported as LAP minus PLAIN.

Dataset (LCC)	#Nodes	m	Band	t	Std.	Spearman ρ (%)	Triplet (%)	ΔRecall@10 (%)
web-Google	855,802	32	33-64	80	center	+2.12	+0.64	+2.86
web-Google	855,802	32	90-121	120	none	+4.37	+1.75	+6.36

Practical considerations. The spectral preprocessing is performed offline using a modest number of eigenpairs and then treated as fixed node features during training. Since our encoding is sign-/basis-invariant by construction (Proposition **??**), it is stable to eigenvector sign flips and orthogonal rotations within eigenspaces, and is compatible with routine preprocessing such as centering.