Bridging Distance and Spectral Positional Encodings via Anchor-Based Diffusion Geometry Approximation

Zimo Yan and Zheng Xie* and Runfan Duan and Chang Liu and Wumei Du
National University of Defense Technology, Changsha, China
{yanzimo20,xiezheng81,duanrunfan24,liuchang_,wumeidu}@nudt.edu.cn

Abstract

Molecular graph learning benefits from positional signals that capture both local neighborhoods and global topology. Two widely used families are spectral encodings derived from Laplacian or diffusion operators and anchor-based distance encodings built from shortest-path information, but the relationship between them is still not well understood. In this paper, we study when anchor-based distance encodings can approximate diffusion geometry. Under a random r-regular graph model, we derive an explicit trilateration map that reconstructs truncated diffusion coordinates from transformed anchor distances and anchor spectral positions, together with pointwise and Frobenius-gap guarantees. On DrugBank with a shared GNP-based DDI backbone, anchor-distance Nyström accurately recovers diffusion geometry, and both DE and LapPE outperform models without positional encodings, with LapPE showing slightly more consistent performance.

1 Introduction

Molecular graph modeling includes tasks such as de novo generation, interaction prediction, and molecular reasoning. These tasks require representations that capture both local chemical neighborhoods and global topology (2017; 2020; 2020). A recurring challenge is how to inject structural positional information so that long-range dependencies and global organization are accessible to the model. This challenge is especially important in molecular settings, where useful representations must reflect not only local chemistry but also global graph geometry.

Many positional and structural signals have been explored, including random-walk statistics, structural-role descriptors, and relative-position biases in graph transformers (2022; 2023; 2021). We focus on two particularly common families. Spectral or diffusion encodings, such as Laplacian eigenmaps and diffusion-map features, provide coordinates aligned with diffusion geometry but often rely on global eigendecomposition or iterative spectral approximation (2007; 2023; 2022). Distance or anchor encodings instead represent each node by its shortest-path distances to a small set of anchors, optionally after a radial transformation, offering a simple and scalable alternative without explicit spectral computation (2025a; 2020).

Despite their widespread use, the relation between these two families remains largely heuristic. It is still unclear when anchor-based distance features preserve diffusion geometry, how reconstruction quality should be measured at both the coordinate and induced distance/kernel levels, and how such recovery diagnostics relate to downstream utility. This gap matters in molecular settings, where diffusion geometry is a natural inductive bias but full spectral computation may be undesirable. Our goal is therefore not to argue that distance encodings universally outperform spectral ones, but to characterize when they can act as principled surrogates for spectral geometry in tasks such as drug–drug interaction (DDI) prediction (2018; 2018).

Motivation. This leads to the following question: *When can anchor-based shortest-path encodings be converted into truncated diffusion coordinates with explicit pointwise and matrix-level error control, and when do the resulting recovery diagnostics indicate that a distance encoding is a reliable surrogate for spectral geometry?* Answering this question would connect discrete distance primitives to diffusion geometry and clarify the accuracy–efficiency trade-off behind replacing spectral computation with anchor-based features.

We address this question through theory and experiments. Theoretically, under a random r-regular graph model, we derive an explicit trilateration operator that reconstructs truncated diffusion coor-

*Corresponding author: Zheng Xie (xiezheng81@nudt.edu.cn).

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 13066–13085

dinates from transformed anchor distances and anchor spectral positions, with pointwise and matrix-level error guarantees. We use this model not as a realistic description of molecular graphs, but as a simple nontrivial setting in which the bridge can be analyzed: regularity removes easy local degree cues, and local tree-like structure supports a monotone relation between shortest-path and diffusion distances.

Empirically, we evaluate geometry recovery on DrugBank using a Nyström approximation based on anchor distances, and test downstream DDI prediction on DrugBank and ChCh-Miner with a shared GNP backbone as shown in Figure 1(2020; 2023; 2025b). Both DE and LapPE improve over NoPE, while LapPE is more consistent overall.

Our main contributions are:

1. We establish an explicit bridge between anchor-based distance encodings and truncated diffusion geometry via a trilateration operator, and show in an idealized random r-regular regime that the resulting reconstruction error can be controlled both pointwise and at the matrix level.

2. We translate this bridge into observable recovery diagnostics and show on DrugBank molecular graphs that a distance-driven Nyström scheme can closely recover diffusion geometry with a moderate anchor budget.

3. Under a shared GNP-based DDI backbone, we provide a controlled comparison of NoPE, DE, and LapPE, together with ablations on distance transforms and anchor counts, showing that both positional encodings help over NoPE while clarifying the trade-off between distance-based surrogacy and spectral stability.

2 Related Work

This section reviews structural and positional information for graph learning, emphasizing molecular graphs. We highlight limitations of purely local computation, summarize spectral and distance-based positional encodings, and situate drug-drug interaction (DDI) prediction as a downstream testbed to motivate our spectral-algebraic bridge.

2.1 Locality limitations in graph learning

Many graph learning methods rely on message passing, where node representations are iteratively updated using neighboring features (2017; 2017). While effective across molecular property prediction and other tasks, locality imposes structural constraints. Standard message passing is bounded by the Weisfeiler-Leman test, leaving some non-isomorphic nodes indistinguishable (2019; 2018). Long-range dependencies can also be difficult to capture due to over-smoothing and over-squashing (2021; 2022). These limitations motivate augmenting node features with explicit structural or positional signals.

2.2 Spectral positional encodings

Spectral encodings derive coordinates from Laplacian or diffusion operators, such as Laplacian eigenmaps or diffusion maps (2003; 2005; 2006). They align with diffusion geometry and can break symmetries in graph representations (2023). Variants include learned reweightings or propagation-based approximations (2022; 2023), but global spectral computation remains costly on large or evolving graphs. Prior work on approximate spectral methods highlights iterative eigensolvers or low-rank approximations, which reduce computation but still rely on global modes.

2.3 Distance-based positional encodings

Distance encodings represent each node by its shortest-path distances to a set of anchors, optionally after simple radial transformations (2020; 2019). These encodings scale efficiently and are easy to compute without global spectral routines. They enhance expressive power beyond Weisfeiler-Leman and have been used in graph transformers and other structural models (2020; 2021). However, their connection to spectral or diffusion geometry has been mostly qualitative, leaving open questions on when a finite anchor set can reliably approximate diffusion geometry and how reconstruction quality can be quantified at both the coordinate and kernel levels.

2.4 Drug-drug interaction prediction as application context

Graph-based approaches are widely used for DDI prediction. Early models combined molecular fingerprints or sequences with neural networks (2018), while later work leveraged knowledge graphs and GNNs to capture interactions in heterogeneous networks (2018; 2020). Recent multimodal or dual-graph frameworks explicitly model substructure interactions and co-attention (2026; 2025). These

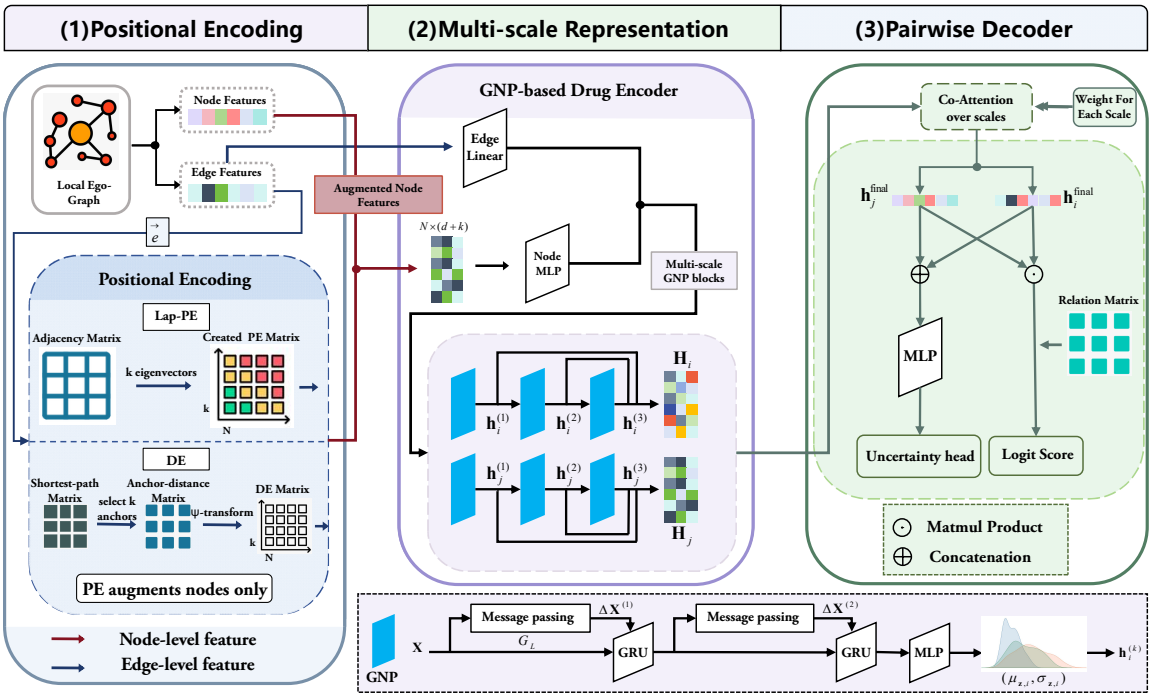


Figure 1: Architecture of the proposed framework for DDI prediction. LapPE or DE first generates a positional encoding matrix, which is concatenated only to node features. The resulting augmented node features, together with edge features, are processed by a multi-scale GNP-based drug encoder. A pairwise decoder with cross-scale co-attention then produces the interaction score and an uncertainty estimate.

studies show the importance of structural information, but most fix positional features and focus on architecture or data integration, leaving systematic analysis of distance versus spectral encodings under a unified theoretical lens largely unexplored.

2.5 Summary and remaining gap

Spectral encodings provide geometrically meaningful coordinates at a global cost, whereas distance encodings are scalable and simple but usually justified empirically. Existing work lacks a unified algebraic account of when distance-based encodings serve as low-rank surrogates for diffusion geometry, and how to evaluate recovery at both coordinate and kernel levels. We address this gap by defining reconstruction operators and error bounds that link distance and spectral encodings in Figure 4. This framing connects structural signals to measurable recovery diagnostics and provides practical guidance for choosing positional encodings.

3 Preliminaries

We introduce the notation and geometric objects used in the analysis, and formalize the two positional-encoding families considered in this paper: anchor-based shortest-path distance encodings and truncated Laplacian spectral coordinates. These definitions set up the algebraic comparison problem studied in Section 4.

3.1 Notations

Let $G = (V, E)$ be a finite, connected, undirected graph with $|V| = n$. For $u, v \in V$, let $\mathrm{SPD}(u, v)$ denote the shortest-path distance. For $u \in V$ and $R \in \mathbb{N}$, define the radius-R ball

$$B_G(u, R) = \{\, v \in V : \mathrm{SPD}(u, v) \le R \,\}. \quad (1)$$

Let $A \in \{0, 1\}^{n \times n}$ be the adjacency matrix and D the diagonal degree matrix. In the theoretical part we focus on r-regular graphs, so $D = rI$ and

$$L := I - \frac{1}{r} A. \quad (2)$$

Let $\{(\lambda_j, \varphi_j)\}_{j=1}^{n}$ be the eigenpairs of L, where $\{\varphi_j\}$ is an orthonormal basis of $\ell^2(V)$ and $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. For $t > 0$, define the heat semigroup

$$K_t = e^{-tL}, \quad (3)$$

with entries $k_t(u, v) = (K_t)_{uv}$. The diffusion dis-

tance at time t can be written equivalently as

$$d_t(u,v)^2 = \sum_{j=1}^{n} e^{-2t\lambda_j}\big(\varphi_j(u) - \varphi_j(v)\big)^2 = k_{2t}(u,u) + k_{2t}(v,v) - 2k_{2t}(u,v). \tag{4}$$

For an integer $m \geq 1$, define the truncated diffusion-map embedding

$$\Phi_t^{(m)}(v) = \big(e^{-t\lambda_{j+1}}\varphi_{j+1}(v)\big)_{j=1}^{m} \in \mathbb{R}^m, \tag{5}$$

and the corresponding truncated diffusion distance

$$d_t^{(m)}(u,v) = \|\Phi_t^{(m)}(u) - \Phi_t^{(m)}(v)\|_2. \tag{6}$$

Given anchors $a_1, \ldots, a_{m+1} \in V$, write

$$p_i = \Phi_t^{(m)}(a_i) \in \mathbb{R}^m, \quad i = 1, \ldots, m+1. \tag{7}$$

In the following, the coordinate-level target is $\Phi_t^{(m)}(v)$, while the matrix-level quantities of interest are the pairwise distances or kernels induced by these truncated coordinates.

3.2 Problem Statement

Fix a graph $G = (V, E)$ and an anchor set $S = \{a_1, \ldots, a_{m+1}\} \subset V$. For any node $v \in V$, define the transformed anchor-distance encoding

$$\tilde{\zeta}(v \mid S) = \big(\psi(\mathrm{SPD}(a_1, v)), \ldots, \psi(\mathrm{SPD}(a_{m+1}, v))\big)^\top, \tag{8}$$

where $\tilde{\zeta}(v \mid S) \in \mathbb{R}^{m+1}$ and $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a monotone scalar transform specified in Section 4; the raw distance encoding corresponds to $\psi = \mathrm{id}$. Our spectral target is the truncated coordinate $\Phi_t^{(m)}(v) \in \mathbb{R}^m$ (or an equivalent sign-/basis-invariant representation).

The central question is whether, under suitable structural conditions on G and a nondegenerate anchor configuration S, one can construct an explicit algebraic map from $\tilde{\zeta}(v \mid S)$ to $\Phi_t^{(m)}(v)$ with quantitatively controlled pointwise error, while also controlling the associated matrix-level discrepancy between the induced distance or kernel quantities.

Section 4 formalizes these conditions and shows that such a linkage exists. The resulting guarantees bound both coordinate-level reconstruction error and matrix-level discrepancy in terms of an explicit structural mismatch parameter δ, within a logarithmic-radius local regime.

4 Theoretical Analysis

This section analyzes the algebraic relation between shortest-path distance encodings (DE) and truncated Laplacian spectral coordinates (LapPE) on r-regular graphs. All proofs are deferred to Appendix A.4. Throughout this section we use the (symmetric normalized) Laplacian in the r-regular form $L = I - \frac{1}{r}A$, so that the diffusion operator is $K_t = e^{-tL}$ and the truncated diffusion-map embedding is $\Phi_t^{(m)}$ as defined in Section 3.

4.1 Assumptions

Assumption 1. *For the theoretical analysis, we consider graphs drawn from the random regular model $G \sim \mathcal{G}_{n,r}$, the uniform distribution on labeled r-regular graphs on $[n]$ with fixed degree $r \geq 3$.*

Assumption 2. *There exist $m \in \mathbb{N}$ and $\alpha > 0$ such that, with high probability over $G \sim \mathcal{G}_{n,r}$, the truncated embedding $\Phi_t^{(m)} : V \to \mathbb{R}^m$ is injective and satisfies*

$$\min_{u \neq v} \big\|\Phi_t^{(m)}(u) - \Phi_t^{(m)}(v)\big\|_2 \geq n^{-\alpha}. \tag{9}$$

Assumption 1 specifies the analytical regime used to derive an explicit bridge from anchor shortest-path distances to truncated diffusion coordinates. We adopt the random r-regular model not as a literal model of molecular graphs, but as a clean setting in which regularity and local tree-likeness make the relation between shortest-path distance and diffusion geometry tractable. Assumption 2 is a nondegeneracy condition ensuring stable recovery. Outside this regime, deviations are expected to manifest as larger linkage mismatch, truncation residuals, or poorer conditioning, which we later examine empirically.

4.2 Main Theoretical Results

We first define an explicit trilateration operator that maps transformed anchor-distance features to an m-dimensional spectral coordinate. Then we introduce the quantities controlling the approximation error and state the resulting pointwise and matrix-level guarantees.

Let $\psi_* : \mathbb{R}^{m+1} \to \mathbb{R}^{m+1}$ denote the elementwise application of a strictly increasing scalar function $\psi : [0, R] \to \mathbb{R}_+$, i.e.,

$$\big(\psi_*(y)\big)_i := \psi(y_i), \quad i = 1, \ldots, m+1. \tag{10}$$

Recall the raw anchor-distance encoding

$$\zeta(v \mid S) = \big(\mathrm{SPD}(a_1, v), \dots, \mathrm{SPD}(a_{m+1}, v)\big)^\top \in \mathbb{R}^{m+1}. \tag{11}$$

Definition 1. *Fix $t > 0$ and $m \in \mathbb{N}$, and let $a_1, \dots, a_{m+1} \in V$ be anchors with*

$$p_i = \Phi_t^{(m)}(a_i) \in \mathbb{R}^m, \quad i = 1, \dots, m+1. \tag{12}$$

Define

$$A = 2 \begin{pmatrix} (p_1 - p_{m+1})^\top \\ \vdots \\ (p_m - p_{m+1})^\top \end{pmatrix} \in \mathbb{R}^{m \times m}. \tag{13}$$

For any $q = (q_1, \dots, q_{m+1})^\top \in \mathbb{R}^{m+1}$, define

$$b(q) = \begin{pmatrix} \|p_1\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_1^2 \\ \vdots \\ \|p_m\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_m^2 \end{pmatrix} \in \mathbb{R}^m. \tag{14}$$

Whenever A is invertible, the trilateration operator is defined by

$$T(v) := A^{-1} b\big(\psi_*(\zeta(v \mid S))\big) \in \mathbb{R}^m. \tag{15}$$

We next introduce the quantities that enter the reconstruction error bounds.

Let $\{(\lambda_j, \varphi_j)\}_{j=1}^n$ be an orthonormal eigendecomposition of L. For $m \in \mathbb{N}$ and $t > 0$, define the truncation tail

$$\mathrm{Tail}_t^{(m)}(u, v)^2 := \sum_{j=m+2}^{n} e^{-2t\lambda_j} \big(\varphi_j(u) - \varphi_j(v)\big)^2. \tag{16}$$

Let $p : V \to \mathbb{R}^m$ and $V_1, \dots, V_{m+1} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(V)$. Define the degeneracy probability

$$\eta_n := \mathbb{P}\big(\det\big[p(V_1) - p(V_{m+1}) \;\; \cdots \;\; p(V_m) - p(V_{m+1})\big] = 0\big). \tag{17}$$

The next result shows that, within a local radius-R regime, the discrepancy between transformed shortest-path distance and truncated diffusion distance decomposes into a diffusion-geometry linkage term and a truncation term.

Theorem 2. *Fix $t > 0$ and $m \in \mathbb{N}$. Let $G = (V, E)$ be a finite connected graph and $R \geq 1$, $\psi : [0, R] \to \mathbb{R}_+$ be strictly increasing. Define*

$$\Delta_{\mathrm{geom}}(G; R, t, \psi) := \sup_{\mathrm{SPD}(u,v) \leq R} \big| d_t(u, v) - \psi(\mathrm{SPD}(u, v)) \big| \tag{18}$$

and

$$\Delta_{\mathrm{tail}}(G; R, t, m) := \sup_{\mathrm{SPD}(u,v) \leq R} \mathrm{Tail}_t^{(m)}(u, v). \tag{19}$$

Then, for all $u, v \in V$ with $\mathrm{SPD}(u, v) \leq R$,

$$\big| d_t^{(m)}(u, v) - \psi(\mathrm{SPD}(u, v)) \big| \leq \Delta_{\mathrm{geom}}(G; R, t, \psi) + \Delta_{\mathrm{tail}}(G; R, t, m). \tag{20}$$

For later use, define the local linkage/truncation error level

$$\delta_n^L := \Delta_{\mathrm{geom}}(G; R, t, \psi) + \Delta_{\mathrm{tail}}(G; R, t, m). \tag{21}$$

Thus, δ_n^L summarizes the local mismatch between transformed anchor distances and truncated diffusion geometry.

Proposition 3. *Fix $m \geq 1$ and let $p : V \to \mathbb{R}^m$ be the truncated diffusion embedding $p(v) := \Phi_t^{(m)}(v)$. Draw anchors $a_1, \dots, a_{m+1} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(V)$ and set $p_i := p(a_i)$. Define*

$$M := \big[p_1 - p_{m+1} \;\; \cdots \;\; p_m - p_{m+1}\big] \in \mathbb{R}^{m \times m}. \tag{22}$$

Then $p_1, \dots, p_{m+1}$ are affinely independent in $\mathbb{R}^m$ if and only if $\det(M) \neq 0$. Moreover, the trilateration matrix in Definition 1 satisfies $A = 2M^\top$.

Assume the degeneracy probability η_n in (17) satisfies $\eta_n = o(1)$. Then $\mathbb{P}\big(\det(M) \neq 0\big) = 1 - \eta_n = 1 - o(1)$.

Remark 4. *Fix $m \geq 1$ and let $p_1, \dots, p_{m+1} \in \mathbb{R}^m$ be arbitrary points. Let $\varepsilon > 0$ and let $\xi_1, \dots, \xi_{m+1} \in \mathbb{R}^m$ be random vectors whose joint law is absolutely continuous with respect to Lebesgue measure. Setting $\tilde{p}_i := p_i + \varepsilon \xi_i$ and $\tilde{M} := \big[\tilde{p}_1 - \tilde{p}_{m+1} \;\; \cdots \;\; \tilde{p}_m - \tilde{p}_{m+1}\big]$, we have $\mathbb{P}\big(\det(\tilde{M}) \neq 0\big) = 1$.*

The next two theorems bound the pointwise reconstruction error $\|\Phi_t^{(m)}(v) - T(v)\|_2$ and the induced matrix-level discrepancy.

Theorem 5. *Let $G \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$, and fix $t > 0$ and $m \in \mathbb{N}$. Let $a_1, \dots, a_{m+1}$ be anchors satisfying Proposition 3 (and, if needed, the jittered-genericity remark above), and let $\Phi_t^{(m)}$ be the truncated spectral embedding. Suppose Assumptions 1-2 hold, and Theorem 2 holds for $d_t^{(m)}$ with error δ_n^L.*

For a node $v \in V$ such that $\mathrm{SPD}(a_i, v) \leq R$ for all i, define

$$r_i := \psi\big(\mathrm{SPD}(a_i, v)\big), \quad r_i^\star := d_t^{(m)}(a_i, v). \tag{23}$$

Then, with probability $1 - o(1)$ *over* G*, we have*

$$\left\|\Phi_t^{(m)}(v) - T(v)\right\|_2 \;\le\; \|A^{-1}\|_{\mathrm{op}}\,\sqrt{m}\,\Big(4\,\rho_R\,\delta_n^L + 2(\delta_n^L)^2\Big), \tag{24}$$

where

$$\rho_R := \max_{0\le d\le R} \psi(d) \;+\; \delta_n^L. \tag{25}$$

In particular, for $\delta_n^L \le 1$*,*

$$\left\|\Phi_t^{(m)}(v) - T(v)\right\|_2 \;\le\; \|A^{-1}\|_{\mathrm{op}}\,\sqrt{m}\,\left(4\rho_R + 2\right)\,\delta_n^L. \tag{26}$$

Theorem 6. *Under the same setting and assumptions as in Theorem 5, define*

$$(D_{\mathrm{SPD}})_{v,i} = \mathrm{SPD}(v, a_i), (D_{\mathrm{diff}}^{(m)})_{v,i} = d_t^{(m)}(v, a_i), \tag{27}$$

and let $\psi_*(D_{\mathrm{SPD}})$ *denote the elementwise application of* ψ*. Then, with probability* $1 - o(1)$*,*

$$\left\|D_{\mathrm{diff}}^{(m)} - \psi_*(D_{\mathrm{SPD}})\right\|_F \;\le\; \delta_n^L\,\sqrt{n(m+1)}. \tag{28}$$

In particular, the average per-entry discrepancy is at most δ_n^L*.*

Proofs are given in Appendix A.4. Theorem 5 shows that the pointwise reconstruction error is controlled jointly by the local linkage/truncation error δ_n^L and the conditioning of the trilateration system through $\|A^{-1}\|_{\mathrm{op}}$. Theorem 6 aggregates the same local mismatch over all node-anchor pairs, yielding a Frobenius bound for the node-anchor distance matrix.

4.3 Implications for expressivity relative to NoPE

As a supporting implication, we recall a standard consequence of distance encodings for message passing GNNs on random regular graphs.

Theorem 7. *Let* $G \sim \mathcal{G}_{n,r}$ *be drawn from the random* r*-regular graph model with fixed* $r \ge 3$*, and let* $\mathcal{F}_{\mathrm{MP}}$ *denote the class of* T*-layer message passing GNNs without positional encodings (NoPE), whose distinguishing power is upper bounded by the 1-WL test (2019; 2018). Let* $\mathcal{F}_{\mathrm{DE}}$ *denote the class of* T*-layer message passing GNNs augmented with a distance encoding based on shortest-path distances to* $k = \Theta(\log n)$ *anchors, as in (2020). Then there exists a family of node-level classification tasks on* $\mathcal{G}_{n,r}$ *such that, with high probability over* G*,*

1. *no NoPE message passing GNN in* $\mathcal{F}_{\mathrm{MP}}$ *can realize the target labels (because 1-WL cannot distinguish the relevant nodes); but*
2. *some DE-augmented GNN in* $\mathcal{F}_{\mathrm{DE}}$ *separates all label classes exactly.*

In particular, on random regular graphs the function class $\mathcal{F}_{\mathrm{DE}}$ *is strictly more expressive than the NoPE class* $\mathcal{F}_{\mathrm{MP}}$*.*

Corollary 1. *Under Assumptions 1 and 2 (and the same random-regular/diffusion-geometry regime used above), consider the NoPE and DE variants of our backbone architecture: (i)* $\mathcal{F}_{\mathrm{NoPE}}$*, the backbone without positional encodings; and (ii)* $\mathcal{F}_{\mathrm{DE}}$*, the same backbone augmented with* $\zeta(\cdot \mid S)$ *using* $k = \Theta(\log n)$ *anchors. Then, with high probability over* $G \sim \mathcal{G}_{n,r}$ *and random anchors* S*,* $\mathcal{F}_{\mathrm{DE}}$ *is strictly more expressive than* $\mathcal{F}_{\mathrm{NoPE}}$*; in particular, there exist node classification tasks realizable by some DE-augmented instance but by no NoPE instance.*

Corollary 1 follows from Theorem 7 and the equivalence between NoPE message passing GNNs and the 1-WL test (2019; 2018). Together with Theorems 5-6, this supports the following informal expressivity chain on random regular graphs

$$\text{NoPE} \;\subsetneq\; \text{DE-augmented} \;\approx\; \text{LapPE-augmented}, \tag{29}$$

where $\approx$ is in the sense of the reconstruction and Frobenius-gap bounds above.

5 Experimental Setup

We conduct a controlled comparison of positional encodings under a unified GNP-based DDI prediction backbone. Unless stated otherwise, all implementation, preprocessing, anchor-selection, and hyperparameter details are provided in Appendix B. The same experimental setup is also used for the geometry-recovery and efficiency analyses reported later, where applicable.

Datasets and protocol. We evaluate on DrugBank, Decagon and ChCh-Miner, where nodes are drugs and edges are known interactions. Molecular graphs are built from SMILES using RDKit. We adopt the inductive link prediction protocol used in prior baselines, with train/val/test splits over drug pairs. Unless otherwise noted, all methods use the same molecular preprocessing, data splits, and training protocol.

Baselines. To isolate positional effects, we fix the backbone and compare three primary variants: NoPE, distance encoding (DE), and Laplacian positional encoding (LapPE). We additionally report

RWSE and heat-kernel signatures (HKS) as reference positional/structural baselines under the same backbone and training protocol. For the geometry-recovery and efficiency analyses, we also include an iterative spectral approximation baseline to provide a fair comparison with distance-driven approximation.

Model configuration. All models are implemented in PyTorch and PyTorch Geometric with standard atom/bond features. LapPE uses the first m nontrivial eigenvectors of the normalized Laplacian; DE uses shortest-path distances to k anchors with a radial transform $\psi(\cdot)$. RWSE uses return probabilities at steps $\{1, 2, 4, 8, 16\}$ and HKS uses diffusion times $\{0.1, 0.5, 1, 2, 5\}$ from a truncated Laplacian spectrum. Anchor selection for DE, the implementation of the spectral approximation baseline, and the remaining architecture, normalization, and optimization settings are given in Appendix B.

Evaluation. DDI prediction is treated as binary classification; we report AUROC and F1 on the held-out test set, with the F1 threshold selected on the validation set. Primary results (NoPE/DE/LapPE) are averaged over three random seeds; RWSE/HKS are reported using the available runs under the same protocol. In addition to downstream performance, we evaluate representation quality through diffusion-geometry recovery against the full spectral reference, using coordinate- and geometry-level discrepancy measures reported in Section 6. When comparing approximation routes, we also report wall-clock runtime under the same hardware/software environment.

Ablations. We probe the DE design by varying $\psi(\cdot)$ on DrugBank with the number of anchors fixed, and varying the number of anchors k on ChCh-Miner with $\psi(\cdot)$ fixed. These ablations test the effect of distance shaping and anchor budget under the same preprocessing and training settings; detailed results are in Section 6.5.

6 Experimental Results and Analysis

We evaluate the proposed DE-LapPE bridge from three perspectives: (i) theory-aligned validation on random r-regular graphs (Section 4); (ii) controlled spectral-approximation on real DrugBank molecular graphs; and (iii) downstream DDI prediction under a fixed GNP backbone (NoPE/DE/LapPE), together with targeted DE ablations on $\psi(\cdot)$ and the number of anchors k.

Table 1: Validation on random r-regular graphs (r=6, t=1.0, m=8; three seeds).

(a) Linkage and matrix-level discrepancy.

n	R	$\hat{\delta}_L^n$ (mean±std)	Frob-gap (mean±std)
256	6	0.1580 ± 0.0035	0.0350 ± 0.0009
512	7	0.1099 ± 0.0051	0.0290 ± 0.0028
1024	7	0.0855 ± 0.0040	0.0217 ± 0.0014
2048	8	0.0664 ± 0.0024	0.0156 ± 0.0002

(b) Pointwise reconstruction and conditioning.

n	R	$\|\Phi - T\|_2$ (median)	cond(A) (median)
256	6	0.449	46.89
512	7	0.757	112.41
1024	7	0.181	63.05
2048	8	0.0845	30.00

6.1 Theory-aligned validation on random r-regular graphs

We run a controlled study on $G \sim \mathcal{G}_{n,r}$ with $r = 6$ and $n \in \{256, 512, 1024, 2048\}$ (three seeds per n). For each graph, we form $L = I - \frac{1}{r}A$ and compute truncated diffusion coordinates $\Phi_t^{(m)}$ with $t = 1.0$ and $m = 8$. We set the locality radius $R = \lceil \log n \rceil$ and fit a monotone map ψ from SPD to $d_t^{(m)}$ using isotonic regression on pairs with $\mathrm{SPD}(u, v) \le R$.

We report (i) the empirical linkage error $\hat{\delta}_L^n := \max_{\mathrm{SPD}(u,v)\le R} |d_t^{(m)}(u, v) - \psi(\mathrm{SPD}(u, v))|$; (ii) the normalized Frobenius gap $\|D_{\mathrm{diff}}^{(m)} - \psi_*(D_{\mathrm{SPD}})\|_F / \sqrt{n(m+1)}$ on node-anchor matrices; and (iii) trilateration reconstruction with $m+1$ anchors via $T(v)$, summarized by the median pointwise error $\|\Phi_t^{(m)}(v) - T(v)\|_2$ and the median $\mathrm{cond}(A)$.

Table 1 shows that both $\hat{\delta}_L^n$ and the normalized Frobenius gap decrease as n grows, and trilateration reconstruction errors remain small in typical cases.

6.2 Diffusion-geometry recovery on DrugBank molecular graphs

We next test whether the bridge suggested by the theory remains useful on real molecular graphs. We analyze 80 DrugBank molecules (15–200 nodes), compute the reference diffusion kernel ($t = 1$) and truncated diffusion map (top $m = 8$ components), and then recover the same geometry via an anchor-based DE Nyström scheme with $k = 32$ anchors (farthest-point sampling) and Tikhonov regularization.

Nyström-DE is accurate: relative kernel Frobenius error 0.024 ± 0.021 (median 0.020), coordinate MSE 3.9×10^{-4} after Procrustes alignment,

Table 2: Comparison of spectral and DE approximation routes on subsets from DrugBank and Decagon.

Dataset	Method	Coord. MSE ↓	$\lvert\rho\rvert$ ↑	Kernel err. ↓	Runtime (ms) ↓
DrugBank	Full	0.000000	1.000000	0.000000	0.120
	IterSpec	0.000168	0.998126	0.001277	0.507
	DE-Nys	0.000374	0.988853	0.021152	0.574
Decagon	Full	0.000000	1.000000	0.000000	0.123
	IterSpec	0.000049	0.999827	0.000423	0.506
	DE-Nys	0.000514	0.986768	0.021351	0.592

and mean absolute Pearson correlation 0.988 between approximate and exact diffusion distances. Figure 2 further reports the fitted local isotonic link ψ^* (with $R = \lceil \log n \rceil$) and conditioning diagnostics, showing small linkage/gap values and heavy-tailed condition numbers that motivate regularization. These results are the empirical counterpart of the theorem-side bridge: when the local linkage residual is small and the reconstruction system is properly regularized, anchor-distance information can recover the leading diffusion geometry with high fidelity on real molecular graphs.

6.3 Fair comparison with iterative spectral approximation and computational cost

To address the fairness concern, we compare DE-based Nyström with an iterative spectral baseline (eigsh) and the full spectral reference on the same molecular graphs. Table 2 shows a consistent pattern on both DrugBank and Decagon: iterative spectral is almost indistinguishable from the full reference, while DE-based Nyström also preserves the leading diffusion geometry but with larger approximation error. In runtime, however, full eigendecomposition is already inexpensive on these small graphs, so neither iterative spectral nor DE provides a speed advantage. Thus, DE should be viewed here not as a faster solver, but as a distance-driven surrogate route with explicit geometric diagnostics.

6.4 Controlled DDI prediction under a fixed GNP backbone

We now turn to the downstream task. Under the same GNP-DDI backbone, training schedule, and data splits, we compare NoPE, DE, LapPE, and two reference baselines (RWSE, HKS), changing only the positional encoding module. Although geometry recovery is not the downstream objective itself, it helps explain when a distance-driven encoding can serve as a principled surrogate for spectral coordinates. Table 3 reports test AUROC and F1 (mean±std over three runs).

LapPE is best overall. On DrugBank, positional information is critical: NoPE attains $0.890/0.820$ (AUROC/F1), DE improves to $0.976/0.927$, and LapPE further to $0.980/0.934$. On ChCh-Miner, performance is already strong, but LapPE still yields consistent gains ($0.946/0.879$) over NoPE ($0.938/0.870$) and DE ($0.938/0.869$). RWSE is competitive on ChCh-Miner, while HKS is weaker on DrugBank under this plug-in setting.[1]

Table 3: Test AUROC and F1 (mean ± std over three runs) on DrugBank and ChCh-Miner under the same GNP-DDI backbone and training protocol. Only the positional encoding is varied.

Dataset	Method	Test AUROC	Test F1
DrugBank	NoPE	0.890 ± 0.002	0.820 ± 0.003
	DE	0.976 ± 0.002	0.927 ± 0.004
	LapPE	0.980 ± 0.003	0.934 ± 0.006
	RWSE	0.892 ± 0.008	0.812 ± 0.003
	HKS	0.863 ± 0.013	0.787 ± 0.007
ChCh-Miner	NoPE	0.938 ± 0.003	0.870 ± 0.002
	DE	0.938 ± 0.006	0.869 ± 0.004
	LapPE	0.946 ± 0.002	0.879 ± 0.003
	RWSE	0.944 ± 0.003	0.876 ± 0.002
	HKS	0.941 ± 0.002	0.870 ± 0.004

This pattern matches the recovery results above. On DrugBank, DE nearly matches LapPE while clearly outperforming NoPE, consistent with its strong geometry recovery. LapPE still performs best in this controlled setting, as exact spectral coordinates avoid the approximation errors of the DE route. Overall, the results support DE as a competitive surrogate rather than a replacement for exact spectral encodings.

6.5 Distance Encoding Ablation Results

On DrugBank (Table 4a), with $k = 16$ anchors, the choice of $\psi(\cdot)$ is decisive: identity mapping underperforms (test $0.9451/0.8835$ AUROC/F1), while $\psi(d) = \exp(-d)$ performs best (test $0.9760/0.9281$), and $\log(1+d)$ is slightly worse. Parameter counts are identical (1.168M), so the gains come from the distance shaping rather than capacity. The superiority of $\exp(-d)$ is consistent with the diffusion viewpoint: exponential shaping turns shortest-path distance into a more affinity-like quantity, which better matches the local decay pattern of diffusion geometry than the identity map.

On ChCh-Miner (Table 4b), fixing $\psi(d) =$

[1] RWSE uses return probabilities at steps $\{1, 2, 4, 8, 16\}$; HKS uses diffusion times $\{0.1, 0.5, 1, 2, 5\}$ with truncated spectrum dimension k=32 (normalized Laplacian).

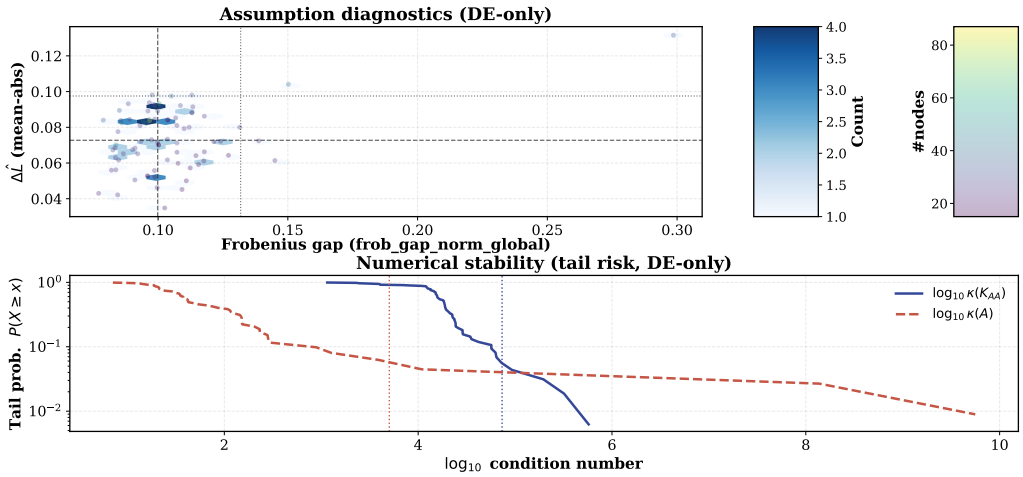


Figure 2: **DrugBank diffusion-geometry diagnostics (80 graphs). Top:** global Frobenius gap versus mean-abs linkage mismatch $\Delta\hat{L}$; median $= (0.0999, 0.0727)$, p95 $= (0.1318, 0.0975)$, Spearman $\rho = 0.16$, and $N = 80$. **Bottom:** CCDFs of $\log_{10}\kappa(K_{AA})$ and $\log_{10}\kappa(A)$; p95s are 4.866 for K_{AA} and 3.702 for A, and invalid/singular rates are 0.0% for K_{AA} and 30.0% for A.

Table 4: Ablation study of distance encoding (DE) on DrugBank and ChCh-Miner. The best configuration for each dataset is shaded.

(a) DrugBank ($k = 16$, varying $\psi(\cdot)$)

ψ	Par.	Val AUC	Val F1	Test AUC	Test F1
identity	1.168	0.9430	0.8795	0.9451	0.8835
exp	1.168	0.9762	0.9279	0.9760	0.9281
log1p	1.168	0.9723	0.9196	0.9721	0.9207

(b) ChCh-Miner ($\psi(d) = \exp(-d)$, varying k)

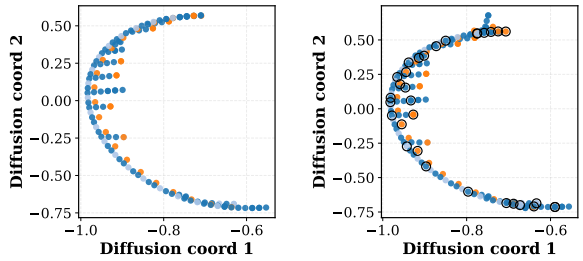
k	Par.	Val AUC	Val F1	Test AUC	Test F1
4	0.123	0.9415	0.8715	0.9410	0.8718
8	0.123	0.9411	0.8715	0.9411	0.8727
16	0.123	0.9406	0.8709	0.9412	0.8722
32	0.124	0.9455	0.8778	0.9453	0.8784

$\exp(-d)$, performance is stable for $k \in \{4, 8, 16\}$ (test $\approx 0.941/0.872$) and improves at $k = 32$ (test $0.9453/0.8784$) with negligible parameter increase (0.123M→0.124M). This suggests that a modest anchor budget already captures most of the benefit, while larger k can further improve coverage at minimal additional model cost.

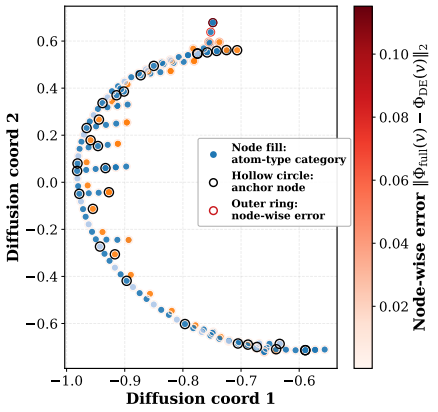
6.6 Qualitative case study on a DrugBank molecular graph

We visualize a single DrugBank molecule (DB00006, 155 atoms) to compare the reference diffusion-map embedding from the full Gaussian-kernel eigendecomposition with the DE-based Nyström embedding using $k = 32$ anchors (after Procrustes alignment). As shown in Fig. 3(a-b), DE Nyström closely matches the global diffusion geometry.

Fig. 3(c) reports the node-wise error $\|\Phi_{\text{full}}(v) - \Phi_{\text{DE}}(v)\|_2$: most nodes are at 10^{-2} scale or below, with mean 7.8×10^{-3} and maximum 1.15×10^{-1}. Together with the aggregate results in Subsection 6.2, this supports that a modest anchor set can recover leading diffusion coordinates with high fidelity in practice.



(a) Full Gaussian kernel eigen-decomposition (b) DE-based Nyström embedding



(c) Node-wise embedding error on DE Nyström

Figure 3: Qualitative comparison of diffusion-based embeddings for a single DrugBank molecular graph (DB00006).

Limitations

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

Theory scope. Our guarantees are established under a random regular graph model and a local linkage between diffusion and shortest-path distances. They do not directly cover graphs with strong degree heterogeneity, pronounced community structure, edge weights, or directionality that commonly arise in practice.

In practice, deviations from this idealized regime are expected to appear through larger linkage mismatch, larger truncation residuals, and poorer conditioning of the reconstruction operator, which are the same quantities we diagnose empirically. When a modest anchor budget already yields strong geometry recovery (e.g., low kernel error and high diffusion-distance correlation), DE is a reasonable scalable surrogate; when these diagnostics are poor, LapPE or iterative spectral approximations are preferable.

Approximation design choices. The dependence of the error on truncation level, anchor placement, and the radial transform is only partially characterized. We do not analyze data-driven anchor selection, learned distance metrics, or learned (potentially non-monotone) distance transforms beyond the targeted ablations.

Empirical generality. We evaluate on two DDI datasets and a single multi-scale Graph Neural Process backbone; results may not fully transfer to other domains, datasets with different labeling/curation 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. As a result, performance under standard splits may not fully reflect real-world pharmacovigilance settings.

Potential risks. Our DDI prediction experiments are intended for research and benchmarking rather than clinical decision making. Misuse or over-reliance on predictions (false positives/negatives) could lead to inappropriate conclusions without expert review and external validation.

Future directions. Positional encodings are treated as fixed precomputed features; we leave joint end-to-end learning of anchors, radial maps, and spectral regularizers 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

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

Mikhail Belkin and Partha Niyogi. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42.

Mengjie Chen, Ming Zhang, and Cunquan Qu. 2025. Towards interpretable drug-drug interaction prediction: A graph-based approach with molecular and network-level explanations. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 203–214.

Gautam Chinta, Jay Jorgenson, and Anders Karlsson. 2015. Heat kernels on regular graphs and generalized ihara zeta function formulas. *Monatshefte für Mathematik*, 178(2):171–190.

Fan Chung and Shing-Tung Yau. 1999. Coverings, heat kernels and spanning trees. *The Electronic Journal of Combinatorics*, 6(1):R12.

Ronald R Coifman and Stéphane Lafon. 2006. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30.

Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the national academy of sciences*, 102(21):7426–7431.

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.

Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai Maron. 2023. Graph positional encoding via random feature propagation. In *International conference on machine learning*, pages 9202–9223. PMLR.

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

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.

Changxiang He, Yuru Liu, Hao Li, Hui Zhang, Yaping Mao, Xiaofei Qin, Lele Liu, and Xuedian Zhang. 2022. Multi-type feature fusion based on graph neural network for drug–drug interaction prediction. *BMC Bioinformatics*, 23(1):224.

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

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

Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, and Xiangxiang Zeng. 2020. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In *IJCAI*, volume 380, pages 2739–2745.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *International Conference on Learning Representations*. ICLR.

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.

Sohir Maskey, Ali Parviz, Maximilian Thiessen, Hannes Stärk, Ylli Sadikaj, and Haggai Maron. 2022. Generalized laplacian positional encoding for graph representation learning. *arXiv preprint arXiv:2210.15956*.

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.

Dongjiang Niu, Lianwei Zhang, Beiyi Zhang, Qiang Zhang, and Zhen Li. 2024. Das-ddi: A dual-view framework with drug association and drug structure for drug–drug interaction prediction. *Journal of Biomedical Informatics*, 156:104672.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, and 2 others. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pages 8024–8035.

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*.

Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences*, 115(18):E4304–E4311.

Jian Sun, Maks Ovsjanikov, and Leonidas J. Guibas. 2009. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, volume 28, pages 1383–1392.

James Topping, Francesco Di Giovanni, Benjamin P Chamberlain, Xiaowen Dong, and Michael M Bronstein. 2022. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*.

Ulrike Von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.

Jihong Wang, Xiaodan Wang, and Yuyao Pang. 2024. Structnet-ddi: Molecular structure characterization-based resnet for prediction of drug–drug interactions. *Molecules*, 29(20):4829.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. 2020. 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. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Zimo Yan, Zheng Xie, Chang Liu, and Yuan Wang. 2025a. Resolving node identifiability in graph neural processes via laplacian spectral encodings. *arXiv preprint arXiv:2511.19037*.

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

Zimo Yan, Jie Zhang, Zheng Xie, Yiping Song, and Hao Li. 2026. A multi-scale graph neural process with cross-drug co-attention for drug-drug interactions prediction. *MATCH Communications in Mathematical and in Computer Chemistry*, 96(1):5–41.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do transformers really perform badly

for graph representation? In *Advances in Neural Information Processing Systems*, volume 34, pages 28877–28888.

Jiaxuan You, Rex Ying, and Jure Leskovec. 2019. Position-aware graph neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7134–7143. PMLR.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81.

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

A Algebraic Relation Between Distance Encoding and Laplacian Spectral Coordinates

In this appendix, we give detailed proofs of the results stated in Section 4, which connect distance encoding (DE) based on shortest-path distances and Laplacian spectral coordinates via an explicit algebraic map.

Throughout, we use the notation and standing assumptions introduced in Section 3 and Section 4. In particular, $G = (V, E)$ is a finite, simple, connected, undirected graph, L is the normalized Laplacian defined in Section 3, the truncated spectral embedding $\Phi_t^{(m)}$ and the truncated diffusion distance $d_t^{(m)}$ are given there, and the random-regular model, monotone linkage, anchor general position and spectral injectivity assumptions are Assumptions 1-2.

We also use the DE-to-LapPE trilateration operator $T(\cdot)$ introduced in Definition 1, which we briefly recall for convenience.

For ease of reference, Figure 4 provides a schematic summary of the spectral/diffusion view, the anchor-distance view, and the trilateration-based bridge that connects them before we proceed to the detailed proofs.

A.1 Reminder of the trilateration operator and distance matrices

Fix $t > 0$ and $m \in \mathbb{N}$, and let $a_1, \dots, a_{m+1} \in V$ be anchors with

$$p_i = \Phi_t^{(m)}(a_i) \in \mathbb{R}^m, \quad i = 1, \dots, m+1. \tag{30}$$

The $m \times m$ matrix A and the map $b : \mathbb{R}^{m+1} \to \mathbb{R}^m$ are

$$A = 2 \begin{pmatrix} (p_1 - p_{m+1})^\top \\ \vdots \\ (p_m - p_{m+1})^\top \end{pmatrix}, \tag{31}$$

$$b(q) = \begin{pmatrix} \|p_1\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_1^2 \\ \vdots \\ \|p_m\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_m^2 \end{pmatrix}, \tag{32}$$

for any $q = (q_1, \dots, q_{m+1})^\top \in \mathbb{R}^{m+1}$.

Let $\psi : [0, R] \to \mathbb{R}_+$ be the strictly increasing function from Theorem 2. For a vector $y = (y_1, \dots, y_{m+1})^\top \in \mathbb{R}^{m+1}$ we apply ψ elementwise and write

$$\big(\psi_*(y)\big)_i = \psi(y_i), \quad i = 1, \dots, m+1. \tag{33}$$

For a node $v \in V$ and anchor set $S = \{a_1, \dots, a_{m+1}\}$, recall the distance encoding

$$\zeta(v \mid S) = \big(\mathrm{SPD}(a_1, v), \dots, \mathrm{SPD}(a_{m+1}, v)\big)^\top. \tag{34}$$

Whenever A is invertible, Definition 1 sets

$$T(v) = A^{-1} b\big(\psi_*(\zeta(v \mid S))\big) \in \mathbb{R}^m. \tag{35}$$

In addition, the node-by-anchor shortest-path and truncated diffusion distance matrices are

$$\big(D_{\mathrm{SPD}}\big)_{v,i} = \mathrm{SPD}(v, a_i), \tag{36}$$

$$\big(D_{\mathrm{diff}}^{(m)}\big)_{v,i} = d_t^{(m)}(v, a_i) = \big\|\Phi_t^{(m)}(v) - \Phi_t^{(m)}(a_i)\big\|_2, \tag{37}$$

and we write $\psi_*(D_{\mathrm{SPD}})$ for the matrix obtained by applying ψ elementwise.

A.2 Proof of Theorem 2

We begin by recalling a standard logarithmic-depth exploration window on random regular graphs. This result is proved and used as a key input in the distance-encoding analysis of Li-Wang-Wang-Leskovec (2020), and we import it here as a black-box statement in order to fix a concrete $R = \Theta(\log n)$ regime in which local neighborhoods exhibit tree-like expansion.

Lemma 8 (Li et al. (2020))**.** *Fix an integer $r \geq 3$ and let $G \sim \mathcal{G}_{n,r}$. There exists a sufficiently small constant $\epsilon > 0$ such that, with probability $1 - o(n^{-3/2})$, the following holds. Choose a root $u \in V$ and define the BFS layer sets*

$$Q_k := \{v \in V : \mathrm{SPD}(u, v) = k\}. \tag{38}$$

Let p_k denote the number of frontier half-edges incident to Q_k that remain unmatched at the beginning of the $(k+1)$-th BFS step in the configuration-model exposure. Then for every integer

$$k \in \Big(\frac{\epsilon}{5} \cdot \frac{\log n}{\log(r-1)} + 1,\ \Big(\frac{2}{3} - \epsilon\Big) \cdot \frac{\log n}{\log(r-1)}\Big), \tag{39}$$

one has

$$|Q_k| \geq (r - 1 - \epsilon)^{k-1} \quad \text{and} \quad p_k \geq (r - 1 - \epsilon)\,|Q_k|. \tag{40}$$

Lemma 9. *Let T_r be the infinite r-regular tree and let $t > 0$. Let L_{T_r} denote the (normalized) graph Laplacian on T_r and $K_t := e^{-tL_{T_r}}$ its heat semigroup with heat kernel $k_t(x, y) := (K_t)_{xy}$.*

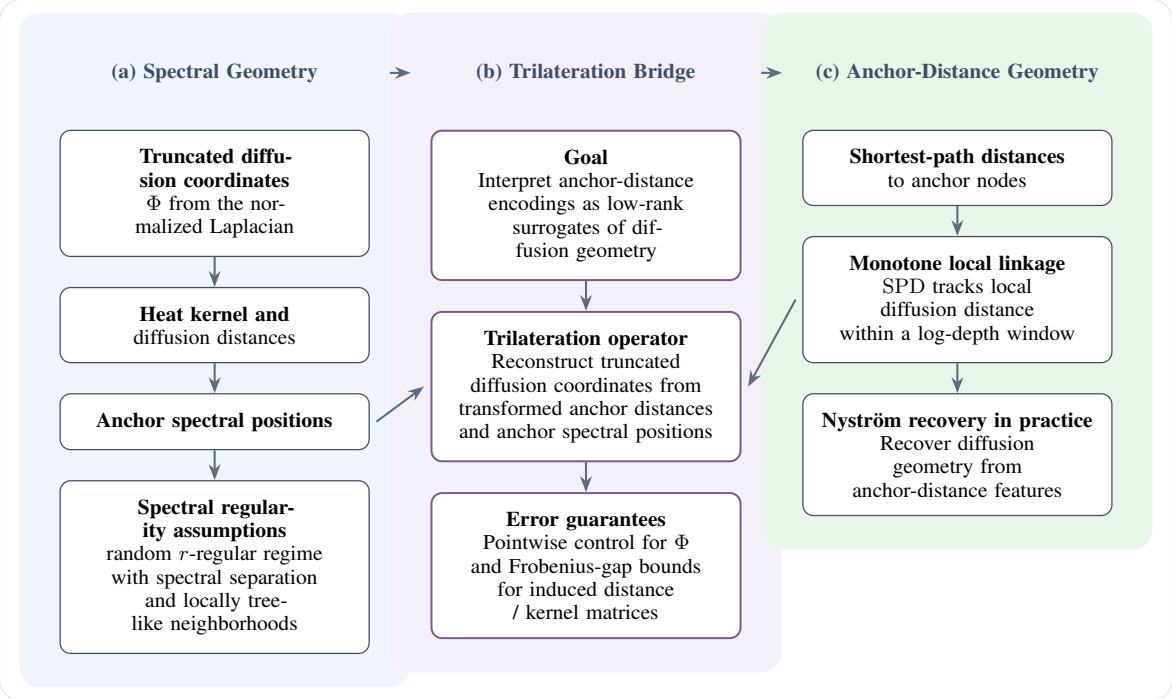


Figure 4: Overview of the theoretical bridge developed in this work. The left column summarizes the spectral view of diffusion geometry, the right column summarizes the anchor-distance view, and the middle column highlights the core bridge: transformed anchor distances and anchor spectral positions yield an explicit trilateration operator with pointwise and matrix-level error guarantees.

Then there exists a function $h_t : \mathbb{N}_0 \to (0, \infty)$ *such that*

$$k_t(x, y) = h_t(d(x, y)), \quad d(x, y) := \mathrm{SPD}_{T_r}(x, y). \tag{41}$$

If one defines the diffusion distance on T_r *by*

$$d_t(x, y)^2 := k_{2t}(x, x) + k_{2t}(y, y) - 2k_{2t}(x, y), \tag{42}$$

then $d_t(x, y) = \psi_{\mathrm{tree}}(d(x, y))$ *with*

$$\psi_{\mathrm{tree}}(d) := \sqrt{2\big(h_{2t}(0) - h_{2t}(d)\big)}. \tag{43}$$

Moreover, for fixed $t > 0$, $h_{2t}(d)$ *is strictly decreasing in* d, *hence* ψ_{tree} *is strictly increasing.*

Proof. Since T_r is distance-transitive, for any two pairs (x, y) and (x', y') with $d(x, y) = d(x', y')$ there exists an automorphism π of T_r such that $\pi(x) = x'$ and $\pi(y) = y'$. The Laplacian L_{T_r} is invariant under automorphisms, hence the heat semigroup $K_t = e^{-tL_{T_r}}$ commutes with π. Writing this invariance at the kernel level yields $k_t(x, y) = k_t(\pi(x), \pi(y)) = k_t(x', y')$, which implies the existence of a radial function h_t with $k_t(x, y) = h_t(d(x, y))$. Complete derivations and explicit formulas for h_t on regular trees are given in (2015); a covering-based treatment relating heat kernels on the infinite regular tree and finite regular graphs is given in (1999).

The diffusion-distance identity then follows algebraically. Since T_r is vertex-transitive, $k_{2t}(x, x) = k_{2t}(y, y) = h_{2t}(0)$, and radiality gives $k_{2t}(x, y) = h_{2t}(d(x, y))$. Substituting into the definition yields

$$\begin{aligned} d_t(x, y)^2 &= h_{2t}(0) + h_{2t}(0) - 2h_{2t}(d(x, y)) \\ &= 2\big(h_{2t}(0) - h_{2t}(d(x, y))\big), \end{aligned} \tag{44}$$

so $d_t(x, y) = \psi_{\mathrm{tree}}(d(x, y))$ with the stated ψ_{tree}. The strict monotonicity of $h_{2t}(d)$ in d for fixed $t > 0$ is established in the regular-tree heat-kernel analyses cited above, which implies that ψ_{tree} is strictly increasing. □

We now prove Theorem 2. In addition to the geometric comparison hypothesis stated in the theorem, the only nontrivial step is to quantify the effect of truncating the diffusion distance to its first m nontrivial eigenmodes. We incorporate that truncation calculation directly into the proof.

Proof of Theorem 2. Fix $t > 0$ and $m \in \mathbb{N}$. Let $G = (V, E)$ be a finite connected graph with symmetric normalized Laplacian L. Let $(\lambda_j, \phi_j)_{j=1}^n$ be an orthonormal eigenbasis with $0 = \lambda_1 \le \lambda_2 \le$

$\cdots \le \lambda_n$. For $t > 0$, define the full diffusion distance

$$d_t(u,v)^2 := \sum_{j=2}^{n} e^{-2t\lambda_j}\big(\phi_j(u) - \phi_j(v)\big)^2, \quad (45)$$

and the truncated diffusion distance

$$d_t^{(m)}(u,v)^2 := \sum_{j=2}^{m+1} e^{-2t\lambda_j}\big(\phi_j(u) - \phi_j(v)\big)^2. \quad (46)$$

Define the spectral tail energy

$$\mathrm{Tail}_t^{(m)}(u,v) := \left(\sum_{j=m+2}^{n} e^{-2t\lambda_j}\big(\phi_j(u)-\phi_j(v)\big)^2\right)^{1/2}. \quad (47)$$

Fix any pair $u, v \in V$ with $\mathrm{SPD}(u,v) \le R$. We first relate $d_t^{(m)}(u,v)$ to $d_t(u,v)$ by an exact decomposition. Splitting the defining sum of $d_t(u,v)^2$ at index $m+1$ yields

$$\begin{aligned} d_t(u,v)^2 &= \sum_{j=2}^{m+1} e^{-2t\lambda_j}\big(\phi_j(u) - \phi_j(v)\big)^2 \\ &\quad + \sum_{j=m+2}^{n} e^{-2t\lambda_j}\big(\phi_j(u) - \phi_j(v)\big)^2. \end{aligned} \quad (48)$$

By definition, the first sum equals $d_t^{(m)}(u,v)^2$ and the second sum equals $\mathrm{Tail}_t^{(m)}(u,v)^2$, hence

$$d_t(u,v)^2 = d_t^{(m)}(u,v)^2 + \mathrm{Tail}_t^{(m)}(u,v)^2. \quad (49)$$

Let $a := d_t^{(m)}(u,v)^2 \ge 0$ and $b := \mathrm{Tail}_t^{(m)}(u,v)^2 \ge 0$. Then $d_t(u,v) = \sqrt{a+b}$ and $d_t^{(m)}(u,v) = \sqrt{a}$, so

$$\begin{aligned} d_t(u,v) - d_t^{(m)}(u,v) &= \sqrt{a+b} - \sqrt{a} \\ &= \frac{(a+b)-a}{\sqrt{a+b}+\sqrt{a}} = \frac{b}{\sqrt{a+b}+\sqrt{a}}. \end{aligned} \quad (50)$$

Since $\sqrt{a+b} + \sqrt{a} \ge \sqrt{a+b} \ge \sqrt{b}$, we obtain

$$\begin{aligned} 0 \le d_t(u,v) - d_t^{(m)}(u,v) &= \frac{b}{\sqrt{a+b}+\sqrt{a}} \le \frac{b}{\sqrt{b}} \\ &= \sqrt{b} = \mathrm{Tail}_t^{(m)}(u,v), \end{aligned} \quad (51)$$

which implies

$$\big|d_t^{(m)}(u,v) - d_t(u,v)\big| \le \mathrm{Tail}_t^{(m)}(u,v). \quad (52)$$

We now compare $d_t^{(m)}(u,v)$ to $\psi(\mathrm{SPD}(u,v))$. By the triangle inequality,

$$\begin{aligned} \big|d_t^{(m)}(u,v) - \psi(\mathrm{SPD}(u,v))\big| &\le \big|d_t^{(m)}(u,v) - d_t(u,v)\big| \\ &\quad + \big|d_t(u,v) - \psi(\mathrm{SPD}(u,v))\big|. \end{aligned} \quad (53)$$

On the high-probability event from the truncation hypothesis in the theorem statement, the first term satisfies

$$\begin{aligned} &\big|d_t^{(m)}(u,v) - d_t(u,v)\big| \le \mathrm{Tail}_t^{(m)}(u,v) \\ &\le \sup_{\mathrm{SPD}(x,y)\le R} \mathrm{Tail}_t^{(m)}(x,y) \le \delta_n^{\mathrm{trunc}}. \end{aligned} \quad (54)$$

On the high-probability event from the geometric comparison hypothesis in the theorem statement, the second term satisfies

$$\big|d_t(u,v) - \psi(\mathrm{SPD}(u,v))\big| \le \delta_n^{\mathrm{geom}}. \quad (55)$$

Intersecting the two events and combining the two inequalities yields

$$\big|d_t^{(m)}(u,v) - \psi(\mathrm{SPD}(u,v))\big| \le \delta_n^{\mathrm{trunc}} + \delta_n^{\mathrm{geom}} = \delta_n^{L}. \quad (56)$$

Because the bounds are uniform over all pairs with $\mathrm{SPD}(u,v) \le R$ on the same intersection event, the conclusion holds with high probability over $G \sim \mathcal{G}_{n,r}$. □

We will use Theorem 2 in the proof of the next reconstruction theorem.

A.3 Proof of Proposition 3

Proof of Proposition 3. Let $a_1, \ldots, a_{m+1}$ be i.i.d. uniform on V and set $p_i := \Phi_t^{(m)}(a_i) \in \mathbb{R}^m$. Define $M = [p_1 - p_{m+1} \ \cdots \ p_m - p_{m+1}] \in \mathbb{R}^{m\times m}$.

Since $(a_1, \ldots, a_{m+1})$ and $(V_1, \ldots, V_{m+1})$ have the same law, then

$$\mathbb{P}\big(\det(M) = 0\big) = \eta_n. \quad (57)$$

Therefore $\mathbb{P}(\det(M) \ne 0) = 1 - \eta_n = 1 - o(1)$.

Assume next the jitter construction. Fix G and the anchors, hence fix $p_1, \ldots, p_{m+1}$. Let $\varepsilon > 0$ and let $\xi_1, \ldots, \xi_{m+1} \in \mathbb{R}^m$ be i.i.d. with a joint density, and define $\tilde{p}_i := p_i + \varepsilon\xi_i$. Let $\tilde{M} = [\tilde{p}_1 - \tilde{p}_{m+1} \ \cdots \ \tilde{p}_m - \tilde{p}_{m+1}]$. Then

$$\tilde{p}_i - \tilde{p}_{m+1} = (p_i - p_{m+1}) + \varepsilon(\xi_i - \xi_{m+1}), \quad (58)$$

$$\tilde{M} = M + \varepsilon X, \quad X := [\xi_1 - \xi_{m+1} \ \cdots \ \xi_m - \xi_{m+1}]. \quad (59)$$

Hence $\det(\tilde{M}) = \det(M + \varepsilon X)$ is a polynomial in the entries of $(\xi_1, \ldots, \xi_{m+1})$. This polynomial is

not identically zero: indeed, choose a deterministic realization with $\xi_{m+1} = 0$ and $\xi_i = e_i$ for $i = 1, \dots, m$, where $\{e_i\}$ is the standard basis of $\mathbb{R}^m$. Then $X = I_m$ and

$$\det(M + \varepsilon X) = \det(M + \varepsilon I_m). \tag{60}$$

As a polynomial in ε, $\det(M + \varepsilon I_m)$ has leading term ε^m and therefore is not the zero polynomial, so there exist noise values making $\det(M+\varepsilon X) \neq 0$. Consequently the zero set

$$Z := \{(\xi_1, \dots, \xi_{m+1}) \in (\mathbb{R}^m)^{m+1} : \det(M + \varepsilon X) = 0\} \tag{61}$$

is a proper algebraic variety and has Lebesgue measure zero. Because $(\xi_1, \dots, \xi_{m+1})$ has a joint density,

$$\mathbb{P}\big((\xi_1, \dots, \xi_{m+1}) \in Z \mid G, a_1, \dots, a_{m+1}\big) = 0, \tag{62}$$

which is equivalent to $\mathbb{P}(\det(\tilde{M}) \neq 0 \mid G, a_1, \dots, a_{m+1}) = 1$. This proves the affine independence under jitter. □

A.4 Proof of Theorem 5

Proof of Theorem 5. Fix $G \sim \mathcal{G}_{n,r}$, anchors $a_1, \dots, a_{m+1}$ and a node $v \in V$ with $\mathrm{SPD}(a_i, v) \leq R$ for all i. We denote

$$z = \Phi_t^{(m)}(v) \in \mathbb{R}^m, p_i = \Phi_t^{(m)}(a_i) \in \mathbb{R}^m, \tag{63}$$

and its distance encoding with respect to S by

$$y_i = \mathrm{SPD}(a_i, v), \quad y = \zeta(v \mid S) = (y_1, \dots, y_{m+1})^\top. \tag{64}$$

Step 1: From shortest-path distances to approximate radii. By Theorem 2, there exist a strictly increasing function ψ and a sequence $\delta_n^L = o(1)$ such that, with high probability, for all i with $y_i \leq R$,

$$\big|d_t^{(m)}(a_i, v) - \psi(y_i)\big| \leq \delta_n^L. \tag{65}$$

Define

$$q_i := \psi(y_i), \quad q := (q_1, \dots, q_{m+1})^\top. \tag{66}$$

For later use, define the exact truncated radius

$$r_i^\star := d_t^{(m)}(a_i, v), \quad i = 1, \dots, m+1, \tag{67}$$

and the corresponding errors

$$\zeta_{i,n} := r_i^\star - q_i, \tag{68}$$

so that

$$r_i^\star = q_i + \zeta_{i,n}, \quad |\zeta_{i,n}| \leq \delta_n^L. \tag{69}$$

Step 2: Exact spectral trilateration with $\{r_i^\star\}$ By definition of $d_t^{(m)}$, for each anchor a_i we have

$$d_t^{(m)}(a_i, v) = \big\|\Phi_t^{(m)}(a_i) - \Phi_t^{(m)}(v)\big\|_2 = \big\|p_i - z\big\|_2. \tag{70}$$

Squaring both sides yields

$$\big\|z - p_i\big\|_2^2 = \big(r_i^\star\big)^2. \tag{71}$$

Using the identity

$$\big\|z - p_i\big\|_2^2 = \|z\|_2^2 - 2\langle z, p_i\rangle + \|p_i\|_2^2, \tag{72}$$

this becomes

$$\|z\|_2^2 - 2\langle z, p_i\rangle + \|p_i\|_2^2 = \big(r_i^\star\big)^2. \tag{73}$$

These are $m+1$ equations in the unknown $z \in \mathbb{R}^m$ and the scalar $\|z\|_2^2$.

To eliminate $\|z\|_2^2$, we subtract the equation corresponding to index $m+1$ from that corresponding to a general index $i \in \{1, \dots, m\}$. For each such i, we consider

$$\begin{aligned}&\Big(\|z\|_2^2 - 2\langle z, p_i\rangle + \|p_i\|_2^2\Big) \quad - \Big(\|z\|_2^2 - 2\langle z, p_{m+1}\rangle + \|p_{m+1}\|_2^2\Big) \\ &= \big(r_i^\star\big)^2 - \big(r_{m+1}^\star\big)^2.\end{aligned} \tag{74}$$

On the left-hand side, we compute term by term

$$\begin{aligned}&\|z\|_2^2 - 2\langle z, p_i\rangle + \|p_i\|_2^2 - \|z\|_2^2 + 2\langle z, p_{m+1}\rangle - \|p_{m+1}\|_2^2 \\ &= -2\langle z, p_i\rangle + 2\langle z, p_{m+1}\rangle + \|p_i\|_2^2 - \|p_{m+1}\|_2^2.\end{aligned} \tag{75}$$

Factoring the inner products gives

$$-2\langle z, p_i\rangle + 2\langle z, p_{m+1}\rangle = 2\langle z, p_{m+1} - p_i\rangle = -2\langle z, p_i - p_{m+1}\rangle. \tag{76}$$

Therefore

$$-2\langle z, p_i - p_{m+1}\rangle + \|p_i\|_2^2 - \|p_{m+1}\|_2^2 = \big(r_i^\star\big)^2 - \big(r_{m+1}^\star\big)^2. \tag{77}$$

Multiplying both sides by -1 yields

$$2\langle z, p_i - p_{m+1}\rangle = \|p_i\|_2^2 - \|p_{m+1}\|_2^2 + \big(r_{m+1}^\star\big)^2 - \big(r_i^\star\big)^2. \tag{78}$$

For each $i = 1, \dots, m$, this is a linear equation in the components of z.

We now stack these equations for $i = 1, \dots, m$. On the left-hand side, the i-th entry is

$$2(p_i - p_{m+1})^\top z. \tag{79}$$

Define

$$A = 2\begin{pmatrix} (p_1 - p_{m+1})^\top \\ \vdots \\ (p_m - p_{m+1})^\top \end{pmatrix}. \tag{80}$$

Then the stacked left-hand side is exactly Az.

On the right-hand side, define

$$b^\star = \begin{pmatrix} \|p_1\|_2^2 - \|p_{m+1}\|_2^2 + (r_{m+1}^\star)^2 - (r_1^\star)^2 \\ \vdots \\ \|p_m\|_2^2 - \|p_{m+1}\|_2^2 + (r_{m+1}^\star)^2 - (r_m^\star)^2 \end{pmatrix}. \tag{81}$$

The m linear equations can be written compactly as

$$Az = b^\star. \tag{82}$$

By Proposition 3, the vectors $p_1 - p_{m+1}, \ldots, p_m - p_{m+1}$ are linearly independent in $\mathbb{R}^m$. Therefore the rows of $A/2$ are linearly independent, hence A has full rank m and is invertible. There is then a unique solution

$$z = A^{-1} b^\star. \tag{83}$$

Since $z = \Phi_t^{(m)}(v)$, this expresses the true spectral coordinate in terms of the exact truncated radii $r_i^\star$.

Step 3: Relating the exact and approximate right-hand sides. In practice we construct $T(v)$ using q_i instead of $r_i^\star$. The corresponding right-hand side is

$$b(q) = \begin{pmatrix} \|p_1\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_1^2 \\ \vdots \\ \|p_m\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_m^2 \end{pmatrix}. \tag{84}$$

Since $q = \psi_*(\zeta(v \mid S))$, Definition 1 gives

$$T(v) = A^{-1} b(q). \tag{85}$$

We now compare $b^\star$ and $b(q)$ entry by entry. For $i = 1, \ldots, m$, we have

$$\begin{aligned} b_i^\star - b(q)_i &= \Big(\|p_i\|_2^2 - \|p_{m+1}\|_2^2 + (r_{m+1}^\star)^2 - (r_i^\star)^2\Big) \\ &\quad - \Big(\|p_i\|_2^2 - \|p_{m+1}\|_2^2 + q_{m+1}^2 - q_i^2\Big) \\ &= (r_{m+1}^\star)^2 - (r_i^\star)^2 - q_{m+1}^2 + q_i^2. \end{aligned} \tag{86}$$

Using $r_i^\star = q_i + \zeta_{i,n}$, we compute

$$\begin{aligned} (r_{m+1}^\star)^2 - q_{m+1}^2 &= (q_{m+1} + \zeta_{m+1,n})^2 - q_{m+1}^2 \\ &= q_{m+1}^2 + 2q_{m+1}\zeta_{m+1,n} + \zeta_{m+1,n}^2 - q_{m+1}^2 \\ &= 2q_{m+1}\zeta_{m+1,n} + \zeta_{m+1,n}^2, \end{aligned} \tag{87}$$

and similarly

$$\begin{aligned} (r_i^\star)^2 - q_i^2 &= (q_i + \zeta_{i,n})^2 - q_i^2 \\ &= 2q_i\zeta_{i,n} + \zeta_{i,n}^2. \end{aligned} \tag{88}$$

Substituting these expressions, we obtain

$$\begin{aligned} &(r_{m+1}^\star)^2 - (r_i^\star)^2 - q_{m+1}^2 + q_i^2 \\ &= \big((r_{m+1}^\star)^2 - q_{m+1}^2\big) - \big((r_i^\star)^2 - q_i^2\big) \\ &= \big(2q_{m+1}\zeta_{m+1,n} + \zeta_{m+1,n}^2\big) - \big(2q_i\zeta_{i,n} + \zeta_{i,n}^2\big). \end{aligned} \tag{89}$$

Therefore

$$b_i^\star - b(q)_i = 2q_{m+1}\zeta_{m+1,n} + \zeta_{m+1,n}^2 - 2q_i\zeta_{i,n} - \zeta_{i,n}^2. \tag{90}$$

We now bound $|b_i^\star - b(q)_i|$ using the inequalities

$$|\zeta_{i,n}| \le \delta_n^L, \quad |\zeta_{m+1,n}| \le \delta_n^L. \tag{91}$$

Because m is fixed and the embedding dimension is finite, there exists a constant $B_t > 0$ such that

$$\|\Phi_t^{(m)}(u)\|_2 \le B_t \tag{92}$$

for all $u \in V$. Hence, for all i,

$$r_i^\star = d_t^{(m)}(a_i, v) = \|\Phi_t^{(m)}(a_i) - \Phi_t^{(m)}(v)\|_2 \le 2B_t, \tag{93}$$

and by $q_i = r_i^\star - \zeta_{i,n}$ we have

$$|q_i| \le |r_i^\star| + |\zeta_{i,n}| \le 2B_t + \delta_n^L. \tag{94}$$

Therefore, for all sufficiently large n (so that $\delta_n^L \le 1$), there exists a constant $\widetilde{B}_t > 0$ (depending only on t, m, r) such that

$$|q_i| \le \widetilde{B}_t, \quad |q_{m+1}| \le \widetilde{B}_t \quad \text{for all } i = 1, \ldots, m. \tag{95}$$

Using the triangle inequality, we obtain

$$\begin{aligned} |b_i^\star - b(q)_i| &\le 2|q_{m+1}|\,|\zeta_{m+1,n}| + \zeta_{m+1,n}^2 + 2|q_i|\,|\zeta_{i,n}| + \zeta_{i,n}^2 \\ &\le 2\widetilde{B}_t\delta_n^L + (\delta_n^L)^2 + 2\widetilde{B}_t\delta_n^L + (\delta_n^L)^2 \\ &= 4\widetilde{B}_t\delta_n^L + 2(\delta_n^L)^2. \end{aligned} \tag{96}$$

For sufficiently large n, $\delta_n^L \le 1$ implies $(\delta_n^L)^2 \le \delta_n^L$, hence there exists a constant $C_1 > 0$ (depending only on $\widetilde{B}_t$) such that

$$|b_i^\star - b(q)_i| \le C_1\,\delta_n^L, \quad i = 1, \ldots, m. \tag{97}$$

Collecting the m inequalities yields

$$\|b^\star - b(q)\|_2 \le \sqrt{m}\,C_1\,\delta_n^L. \tag{98}$$

Step 4: Stability of the linear system and conclusion. The true spectral coordinate and the DE-based coordinate satisfy

$$Az = b^\star, \quad AT(v) = b(q). \tag{99}$$

Subtracting yields

$$A\big(z - T(v)\big) = b^\star - b(q). \tag{100}$$

Since A is invertible, we have

$$z - T(v) = A^{-1}\big(b^\star - b(q)\big). \tag{101}$$

Taking norms gives

$$\big\|z - T(v)\big\|_2 \le \|A^{-1}\|_{\mathrm{op}} \big\|b^\star - b(q)\big\|_2. \tag{102}$$

Combining this with the bound on $\|b^\star - b(q)\|_2$ from Step 3 proves the claim. Recalling that $z = \Phi_t^{(m)}(v)$ completes the proof.

□

A.5 Proof of Theorem 6

Proof of Theorem 6. We prove the Frobenius norm bound for the discrepancy between the truncated diffusion distance matrix $D_{\mathrm{diff}}^{(m)}$ and the transformed shortest-path matrix $\psi_*(D_{\mathrm{SPD}})$.

Recall that

$$\begin{aligned}(D_{\mathrm{SPD}})_{v,i} &= \mathrm{SPD}(v, a_i),\\ (D_{\mathrm{diff}}^{(m)})_{v,i} &= d_t^{(m)}(v, a_i),\\ (\psi_*(D_{\mathrm{SPD}}))_{v,i} &= \psi(\mathrm{SPD}(v, a_i)).\end{aligned} \tag{103}$$

By Theorem 2, with high probability, for all nodes $v \in V$ and anchors a_i with $\mathrm{SPD}(v, a_i) \le R$,

$$\Big|(D_{\mathrm{diff}}^{(m)})_{v,i} - (\psi_*(D_{\mathrm{SPD}}))_{v,i}\Big| \le \delta_n^L. \tag{104}$$

Therefore,

$$\begin{aligned}&\big\|D_{\mathrm{diff}}^{(m)} - \psi_*(D_{\mathrm{SPD}})\big\|_F^2\\ &= \sum_{v\in V}\sum_{i=1}^{m+1}\Big((D_{\mathrm{diff}}^{(m)})_{v,i} - (\psi_*(D_{\mathrm{SPD}}))_{v,i}\Big)^2\\ &\le \sum_{v\in V}\sum_{i=1}^{m+1}(\delta_n^L)^2 = n(m+1)(\delta_n^L)^2.\end{aligned} \tag{105}$$

Taking square roots yields

$$\big\|D_{\mathrm{diff}}^{(m)} - \psi_*(D_{\mathrm{SPD}})\big\|_F \le \delta_n^L\sqrt{n(m+1)}. \tag{106}$$

This proves Theorem 6. □

B Experimental Details

B.1 Datasets

We use two public benchmark datasets, DrugBank and ChCh-Miner, for the downstream drug–drug interaction (DDI) prediction experiments. In addition, for the fairness-oriented geometry-recovery and efficiency comparison in Section 6.3, we include an auxiliary molecular-graph subset derived from Decagon. For all datasets, molecular graphs are constructed from SMILES strings using RDKit, with atoms as nodes and chemical bonds as edges.

DrugBank. DrugBank is derived from the DrugBank 5.0 database and is widely used in recent DDI studies (2018; 2024; 2024). After standard preprocessing, it contains about 1,700 drugs and 190,000 labeled drug pairs. We use it in the binary interaction prediction setting.

ChCh-Miner. ChCh-Miner is a BioSNAP DDI benchmark constructed from approved drugs (2018). It contains 1,514 drugs and 48,514 known DDI links (2024; 2024). As with DrugBank, we treat it as a binary link prediction task.

Decagon (auxiliary geometry-recovery / efficiency dataset). We additionally use a preprocessed subset of Decagon (2018) only for the geometry-recovery and computational-cost comparison in Section 6.3, not for the downstream DDI prediction results in Table 3. After preprocessing, this subset contains 40 drugs, 59,700 labeled drug pairs, and 255 interaction types. Molecular graphs are constructed with the same RDKit pipeline as for DrugBank and ChCh-Miner.

Data splits for downstream DDI prediction. The following train/validation/test split protocol applies to the downstream DDI prediction datasets DrugBank and ChCh-Miner only. Following Yan et al. (2026), we treat both datasets as inductive binary link prediction tasks on drug–drug graphs. Known interactions are split into training, validation, and test sets at the level of drug pairs. Negative examples are sampled uniformly from unobserved drug pairs while maintaining a fixed positive-to-negative ratio in each split (2024; 2022). We adopt an inductive setting in which a subset of drugs appears only in validation or test, requiring generalization to unseen molecules.

Auxiliary protocol for Section 6.3. The Decagon subset used in Section 6.3 is not used for supervised link-prediction training in the main controlled comparison. It is used only for unsupervised geometry-recovery and efficiency comparisons among the full spectral reference, iterative spectral approximation, and DE-based

Nyström reconstruction under the same molecular preprocessing pipeline.

B.2 Baselines and Model Variants

Because our focus is on understanding the role of positional encodings within a fixed architecture, we keep the backbone model identical across settings and vary only the positional encoding module. Concretely, we consider three primary variants (NoPE/DE/LapPE) for controlled comparisons, and additionally report two widely used PE baselines (RWSE and HKS) as sanity-check references under the same backbone and training protocol.

- **NoPE.** The original MPNP-DDI model (2026) without any explicit positional encodings. Node features are given solely by atom-level descriptors derived from RDKit, and the model relies on its multi-scale message-passing scheme to infer structural information.

- **DE.** The backbone model augmented with distance encodings. For each molecular graph, we sample k anchor atoms and compute shortest-path distances from every node to these anchors. A radial transformation $\psi(\cdot)$ is applied to each distance, and the resulting vectors are concatenated to the original node features. The DE module is shared across all experiments, and its design is studied in detail in the ablation experiments.

- **LapPE.** The backbone model augmented with Laplacian positional encodings. For each molecular graph, we compute the first m non-trivial eigenvectors of the normalized Laplacian and concatenate them to the node features, following standard practice in spectral GNNs (2023). We choose m to match the dimensionality of the DE features so that all variants have comparable parameter counts.

- **RWSE (reference baseline).** The backbone model augmented with random-walk structural encodings (RWSE), where each node is assigned a vector of K-step random-walk return probabilities and the resulting encoding is concatenated to node features, following common practice in graph Transformers and PE benchmarks(2022). We use steps $\mathcal{T} = \{1, 2, 4, 8, 16\}$.

- **HKS (reference baseline).** The backbone model augmented with heat-kernel signatures (HKS)(2009), constructed from a truncated eigenspace of the normalized Laplacian. Specifically, for each node we compute a diffusion-time embedding with times $\mathcal{S} = \{0.1, 0.5, 1, 2, 5\}$ using the top-m eigenpairs and concatenate it to node features. We use a truncated eigenspace dimension of m=32.

For fair comparison, all PE variants are used as plug-in augmentations to node features and trained with the same backbone, optimizer, and training budget. When applicable, we choose PE dimensionalities to be of comparable scale to the DE/LapPE feature augmentation.

All models are implemented in PyTorch (2019) using PyTorch Geometric for efficient graph operations (2019). Unless otherwise specified, we follow the main architectural and optimization choices of MPNP-DDI (2026).

B.2.1 Feature Normalization.

For molecular graphs, we construct standard atom and bond features using RDKit, including atom type, degree, valence, aromaticity, hybridization, and bond type. Categorical attributes are encoded as one-hot vectors and concatenated with numerical descriptors. We apply feature-wise standardization (zero mean, unit variance) across the training set for all continuous features and reuse the same statistics at validation and test time.

For Laplacian positional encodings, raw eigenvectors are normalized within each molecular graph to have zero mean and unit variance per eigenvector. To reduce sensitivity to global sign flips, we optionally add small Gaussian noise during training. For distance encodings, shortest-path distances are first rescaled by the median non-zero distance within each graph and then passed through a radial map $\psi(\cdot)$. We consider three choices in our experiments: $\psi(d) = d$, $\psi(d) = \exp(-d)$, and $\psi(d) = \log(1 + d)$. The transformed distances are further standardized across the training set.

B.2.2 Architecture Design.

The backbone architecture follows the multi-scale Graph Neural Process design of MPNP-DDI (2026). Each drug is represented by both its original molecular graph and its corresponding line graph, which captures bond-level interactions. Node and edge features (including positional encodings when present) are projected to a hidden

dimension of 64. The model stacks three Graph Neural Process blocks; each block runs two iterations of message passing on both the molecular graph and the line graph, followed by aggregation and cross-scale fusion, yielding a hierarchy of representations from local substructures to global topology.

For a pair of drugs, the block-wise representations are fed into a cross-drug co-attention module, which computes context-aware embeddings for each drug conditioned on its partner. These pairwise embeddings are then passed through a multilayer perceptron to predict the probability of an interaction. DE and LapPE variants differ only in the additional node-level inputs; all subsequent processing, including the co-attention and decoder, is shared.

B.2.3 Optimization Settings.

We train all models with the AdamW optimizer (2019) using a cosine annealing learning rate scheduler over 50 epochs, following (2026). The initial learning rate and weight decay are selected by grid search on the validation set within a standard range (for example, learning rates between 10^{-4} and 10^{-3} and weight decay between 10^{-5} and 10^{-3}). We use binary cross-entropy loss on the predicted interaction probabilities.

Early stopping is applied based on validation AUROC: if no improvement is observed for 10 consecutive epochs, training is terminated and the best checkpoint is kept. All experiments are repeated with three random seeds, and we report the mean and standard deviation across runs.

B.2.4 Batch Size and Gradient Variance.

Due to the large number of drug pairs and the memory cost of the multi-scale backbone, we use a mini-batch size of 32 drug pairs and accumulate gradients over 4 steps, resulting in an effective batch size of 128, as in (2026). This reduces the variance of stochastic gradients without exceeding GPU memory constraints. We also apply gradient clipping at a fixed maximum norm to prevent rare exploding gradients caused by highly connected molecular graphs.

B.3 Evaluation Measures

We treat DDI prediction as a binary classification problem on drug pairs. For both DrugBank and ChCh-Miner, we evaluate models using the Area Under the Receiver Operating Characteristic Curve (AUROC) and the F1 score, which are standard metrics in DDI prediction (2024; 2022; 2024). AUROC captures the ranking quality over all thresholds, while F1 summarizes the trade-off between precision and recall at a specific threshold.

During training and model selection, we monitor AUROC on the validation set. For F1, we select a threshold that maximizes validation F1 and apply the same threshold to the held-out test set. All reported numbers are computed on the test split using the checkpoint with the best validation AUROC and are averaged over three independent runs.

B.4 Details of the Ablation Study

We conduct a focused ablation study to isolate the impact of key design choices in the distance encoding module. Specifically, we examine (i) the choice of radial transformation $\psi(\cdot)$ applied to shortest-path distances on DrugBank while fixing the number of anchors, and (ii) the number of anchors k on ChCh-Miner while fixing $\psi(\cdot)$.

On DrugBank, we fix the number of anchors and compare three radial functions: $\psi(d) = d$, $\psi(d) = \exp(-d)$, and $\psi(d) = \log(1 + d)$. On ChCh-Miner, we fix $\psi(d) = \exp(-d)$ and vary $k \in \{4, 8, 16, 32\}$ to study the trade-off between approximation quality and model complexity. All ablation runs use the same data preprocessing, model architecture, optimizer, batch size, and early stopping criteria as the main experiments. Detailed numerical results and further analysis are reported in Section 6.5. For additional implementation details (datasets, baselines, and hyperparameters), please refer to our anonymous code repository.