

Information-Theoretic Limits of Positional Identifiability under Hybrid Graph Encodings

Zimo Yan¹, Zheng Xie¹, Chang Liu¹, Yiqin Lv¹
Runfan Duan¹, Yuan Wang², Bingyao Zeng¹

¹National University of Defense Technology, Changsha, Hunan, China

²School of Urban Design, Wuhan University, Wuhan, China

{yanzimo20, xiezheng81, liuchang_, lvyiqin98, duanrunfan24, zengbingyao}@nudt.edu.cn

2024282090042@whu.edu.cn Correspondence: xiezheng81@nudt.edu.cn

Abstract

Graph-based NLP and symbolic reasoning models often rely on positional encodings to distinguish tokens, entities, or reasoning states, but the identifiability provided by these encodings remains poorly understood. We analyze a hybrid graph encoding that combines anchor distances with quantized low-frequency spectral energy. Treating the encoding as a deterministic observation map gives an exact image-preimage identity for localization ambiguity and a one-sided information-theoretic converse: if the joint anchor, spectral, and quantization budget is too small, distinct nodes must share the same code, which also lower-bounds task-label ambiguity observable from that code. Experiments on random 3-regular graphs, CLUTRR relation graphs, and Universal Dependencies trees support the predicted scaling. Under matched normalized budgets, Hybrid is best or tied at 74 of 87 attained thresholds, while PE-only UD probes show improved recovery of syntactic-tree geometry. The results therefore position positional encoding as a pre-model, encoding-level necessary-condition diagnostic rather than an achievability guarantee or an end-to-end model claim.

1 Introduction

Graph neural networks and graph Transformers are widely used on relational data, including graph-based NLP and symbolic reasoning settings where nodes represent tokens, entities, or reasoning states. Their effectiveness often depends on whether such nodes can be distinguished by positional or structural information in the underlying graph. This has made graph positional encoding a central topic in modern graph learning (Dwivedi et al., 2023; Ying et al., 2021; Rampásek et al., 2022; Cantürk et al., 2024), with direct relevance to dependency-based

NLP and text-based relational reasoning (Marcheggiani and Titov, 2017; Sinha et al., 2019; de Marnette et al., 2021).

Existing positional encodings include spectral methods based on Laplacian eigenvectors or invariant spectral features (Lim et al., 2023; Huang et al., 2024), distance- and structure-based methods using anchor distances, random walks, or related descriptors (Li et al., 2020; Brüel-Gabrielsson et al., 2023), and learned or hybrid encoders that combine multiple positional signals (Ying et al., 2021; Dwivedi et al., 2022; Cantürk et al., 2024).

Despite this progress, the identifiability supplied by the encoding itself remains less understood. Downstream accuracy entangles encoding information with model architecture and optimization, while spectral encodings involve sign, basis, and stability issues (Lim et al., 2023; Huang et al., 2024). More broadly, positional resolution is graph-dependent and only partially characterized theoretically (Keriven and Vaiter, 2023; Li et al., 2024).

Motivation: This leads to the following question: *Can we characterize when a graph positional encoding contains enough information to localize nodes, and how such identifiability depends jointly on distance features, spectral features, and quantization?*

To answer this question, we analyze a hybrid observation map combining anchor-distance profiles with quantized low-frequency spectral energy. Its image-preimage structure determines node distinguishability and the task-label ambiguity observable from positional information alone. Figure 1 summarizes this diagnostic pipeline.

Outline & Primary Contributions: The primary contributions of this work are as follows:

1. We derive an information-theoretic converse for hybrid anchor-distance and quantized spec-

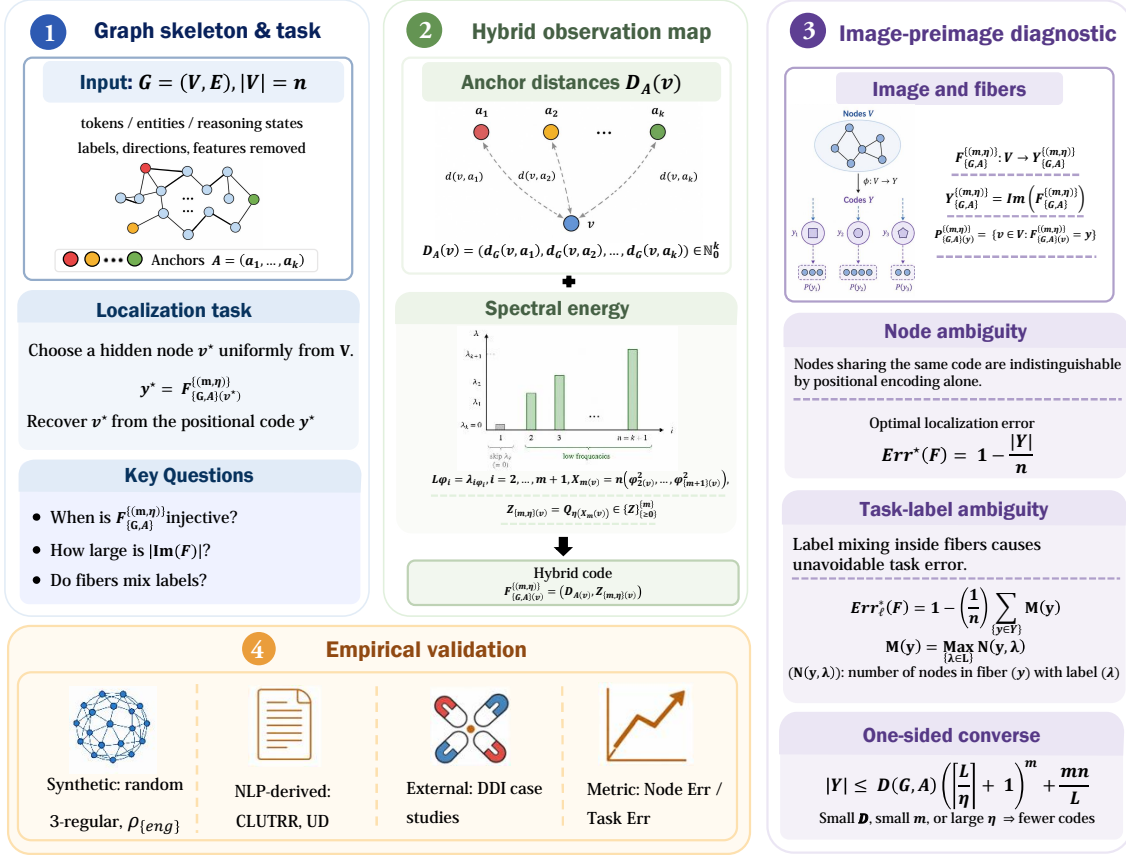


Figure 1: Overview of the proposed encoding-level diagnostic for positional identifiability.

tral encodings, identifying an impossibility regime governed by anchor number, spectral dimension, and quantization resolution.

2. Node localization is formulated as an observation-map problem, with an exact image-size identity relating optimal localization error to collisions among induced positional codes.
3. We extend the image-preimage argument to task-label ambiguity and evaluate it on random 3-regular graphs, CLUTRR relation graphs, and Universal Dependencies trees. We further include budget-controlled comparisons and PE-only UD structural probes to distinguish our diagnostic claim about the structural information preserved by the positional channel from claims about end-to-end model performance.

The paper is organized as follows. Section 2 reviews related work. Section 3 introduces the problem formulation. Section 4 develops the theoretical analysis. Section 5 presents the empirical evaluation.

2 Literature Review

Graph positional encoding provides graph Transformers and related architectures with structural signals beyond local message passing (Zhang et al., 2020; Mialon et al., 2021; Yuan et al., 2025). It is also relevant to graph-based NLP, where nodes may represent tokens, entities, dependencies, or reasoning states (Marcheggiani and Titov, 2017; Sinha et al., 2019; de Marneffe et al., 2021). Existing methods can be broadly grouped into spectral encodings, distance- and structure-aware encodings, and learned or hybrid positional/structural encoders.

2.1 Spectral Positional Encodings

Spectral positional encodings use Laplacian eigenvalues, eigenvectors, or invariant spectral coordinates to inject global graph geometry (Kreuzer et al., 2021; Dwivedi et al., 2022; Wang et al., 2022). This line of work includes spectral attention, learnable positional/structural representations, equivariant or stable encodings, and sign- or basis-invariant models such as SignNet/BasisNet (Lim et al., 2023; Huang et al., 2024). Recent extensions

study expressiveness, random-feature propagation, sheaf-based encodings, and directed-graph positional encodings (Zhang et al., 2024; Eliasof et al., 2023; He et al., 2024; Huang et al., 2025). These methods provide global information, but also face sign, basis, stability, and eigendecomposition challenges.

2.2 Distance- and Structure-aware Encodings

Distance- and structure-aware encodings define node position through relative topology, including shortest-path distances, anchor distances, random-walk or resistance statistics, and rewiring patterns (You et al., 2019; Li et al., 2020; Park et al., 2022; Black et al., 2024). Examples include P-GNN and Distance Encoding for anchor- or distance-based representations (You et al., 2019; Li et al., 2020), as well as Graphormer and GRPE, which incorporate relative structural biases into attention (Ying et al., 2021; Park et al., 2022). Related work studies rewiring, absolute versus relative positional schemes, topology-derived similarity, and diffusion links between anchor distances and spectral geometry (Brüel-Gabrielsson et al., 2023; Black et al., 2024; Zhang et al., 2025; Yan et al., 2026). These encodings are interpretable, but their resolution depends on anchor choice, distance definition, and graph symmetry.

2.3 Learned and Hybrid Positional/Structural Encoders

Learned and hybrid encoders integrate multiple positional cues within trainable graph architectures (Zhang et al., 2020; Yeom et al., 2024). GraphBert, GraphiT, GPS, GRIT, and GPSE incorporate structural information through Transformer biases, message passing, global attention, or transferable positional/structural representations (Zhang et al., 2020; Mialon et al., 2021; Rampásek et al., 2022; Ma et al., 2023; Cantürk et al., 2024). Recent work further studies foundation-style generalization, efficient learnable encodings, benchmarking, and alternative encoding families (Franks et al., 2025; Kanatsoulis et al., 2025; Grötschla et al., 2024; Thabet et al., 2024). These methods are flexible, but their architectural and training complexity makes it difficult to isolate the information supplied by the encoding itself.

2.4 Relation to Our Work

Our work differs from learned positional encoders, position-aware GNNs, and WL-style expressivity

analyses in the object being studied. Those lines ask what a trained architecture can represent or separate; we fix a finite-resolution observation map $F : V \rightarrow \mathcal{Y}$ and analyze the vertex partition induced by its code collisions (You et al., 2019; Li et al., 2024). This is also not a graph canonicalization objective: we do not construct a canonical ordering or a graph-level identifier.

For spectral information, SignNet/BasisNet-style methods provide architecture-level sign/basis invariance (Lim et al., 2023; Huang et al., 2024). Our coordinate-energy map is only sign-invariant when repeated eigenspaces are present, so below we also state a blockwise projector-energy observation map that is invariant to rotations within each eigenspace. Learned and hybrid PEs remain complementary to our framework: once a PE is viewed at finite resolution as a deterministic observation map, its induced collision partition can be diagnosed independently of downstream optimization. Complementary fixed-backbone DDI experiments in Yan et al. (2026) illustrate downstream utility when only the PE module is varied; our contribution instead isolates structural resolution before training.

3 Problem Formulation

We study positional identifiability on undirected graph skeletons underlying graph-based NLP and symbolic reasoning. Removing edge directions, labels, types, lexical attributes, and task-specific features isolates the information carried by the positional code. Our analysis is one-sided: nodes sharing the same hybrid code are indistinguishable to any model using positional information alone.

Let $G = (V, E)$ be a finite connected simple undirected graph with $|V| = n$. For $u, v \in V$, let

$$d_G(u, v) := \text{dist}_G(u, v) \quad (1)$$

denote the shortest-path distance. Given an ordered anchor list

$$\mathcal{A} = (a_1, \dots, a_k) \quad (2)$$

of distinct vertices in V , define the anchor-distance profile of $v \in V$ by

$$D_{\mathcal{A}}(v) := (d_G(v, a_1), \dots, d_G(v, a_k)) \in \mathbb{Z}_{\geq 0}^k. \quad (3)$$

The set of attained anchor-distance profiles and its cardinality are

$$T_{G, \mathcal{A}} := \{D_{\mathcal{A}}(v) : v \in V\}, \quad D(G, \mathcal{A}) := |T_{G, \mathcal{A}}|. \quad (4)$$

For each $t \in T_{G,\mathcal{A}}$, the corresponding distance bucket is

$$B_t(G, \mathcal{A}) := \{v \in V : D_{\mathcal{A}}(v) = t\}. \quad (5)$$

Let $\mathcal{L}$ be the normalized graph Laplacian of G , with eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \quad (6)$$

and an orthonormal eigenbasis

$$\phi_1, \phi_2, \dots, \phi_n \in \mathbb{R}^V \quad (7)$$

ordered by nondecreasing eigenvalue. For $m \in \{1, \dots, n-1\}$, define the normalized spectral energy coordinate

$$X_m(v) := n(\phi_2(v)^2, \phi_3(v)^2, \dots, \phi_{m+1}(v)^2) \in \mathbb{R}_{\geq 0}^m. \quad (8)$$

The factor n places each coordinate on an average $O(1)$ scale, since

$$\sum_{v \in V} n\phi_j(v)^2 = n \quad (9)$$

for every eigenvector ϕ_j . Equivalently,

$$\sum_{v \in V} \sum_{j=1}^m X_m(v)_j = mn. \quad (10)$$

This total-mass identity is the only spectral normalization property used in the trimming-based converse below.

The coordinate-energy signature in (8) is invariant to eigenvector sign flips but not to arbitrary rotations inside a repeated eigenspace. For an eigenspace block E_b with orthogonal projector P_b , a basis-invariant alternative is

$$X_b^{\text{proj}}(v) = n(P_b)_{vv} = n \sum_{j \in E_b} \phi_j(v)^2. \quad (11)$$

This quantity is unchanged by orthonormal basis rotations within E_b and satisfies $\sum_{v \in V} X_b^{\text{proj}}(v) = n|E_b|$. Hence the same trimming/counting argument applies to quantized projector-energy blocks, with one code coordinate per selected block and total spectral mass determined by the sum of their ranks. We use the coordinate form for the main notation and explicitly test the projector form under spectral degeneracy in Section 5.4.

For a quantization level $\eta > 0$, define the coordinatewise quantizer

$$Q_\eta(x_1, \dots, x_m) := (\lfloor x_1/\eta \rfloor, \dots, \lfloor x_m/\eta \rfloor) \in \mathbb{Z}_{\geq 0}^m. \quad (12)$$

We write

$$Z_{m,\eta}(v) := Q_\eta(X_m(v)). \quad (13)$$

The normalized hybrid observation map is

$$F_{G,\mathcal{A}}^{(m,\eta)}(v) := (D_{\mathcal{A}}(v), Z_{m,\eta}(v)), \quad (14)$$

with observable code set

$$\mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)} := \text{Im}(F_{G,\mathcal{A}}^{(m,\eta)}). \quad (15)$$

Let $v_\star$ be a uniformly random source vertex in V . Conditional on G , the anchor list $\mathcal{A}$ may be fixed or sampled uniformly without replacement from V , independently of $v_\star$. The available observation is

$$F_{G,\mathcal{A}}^{(m,\eta)}(v_\star) = (D_{\mathcal{A}}(v_\star), Z_{m,\eta}(v_\star)). \quad (16)$$

The localization problem is to recover $v_\star$ from this observation, given $(G, \mathcal{A})$. The uniform source distribution gives a task-agnostic identifiability criterion; non-uniform task priors would change the weighting of ambiguous codes, but not the basic collision mechanism.

For fixed $(G, \mathcal{A})$, a reconstruction rule is any map

$$s_{G,\mathcal{A}} : \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)} \rightarrow V. \quad (17)$$

Its conditional error probability is

$$\text{Err}(s_{G,\mathcal{A}}) := \mathbb{P}\left(s_{G,\mathcal{A}}(F_{G,\mathcal{A}}^{(m,\eta)}(v_\star)) \neq v_\star \mid G, \mathcal{A}\right). \quad (18)$$

The optimal conditional localization error is

$$\text{Err}^*(F_{G,\mathcal{A}}^{(m,\eta)}) := \inf_{s_{G,\mathcal{A}}} \text{Err}(s_{G,\mathcal{A}}). \quad (19)$$

The key combinatorial objects are the preimages of the observation map. For any $y \in \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)}$, define

$$\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(y) := \{v \in V : F_{G,\mathcal{A}}^{(m,\eta)}(v) = y\}. \quad (20)$$

When $y = F_{G,\mathcal{A}}^{(m,\eta)}(v)$, we also write $\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v)$. Large preimages correspond to observational ambiguity: if many nodes induce the same hybrid positional code, then no reconstruction rule observing only that code can reliably distinguish them. This image-preimage viewpoint underlies the one-sided converse analysis below.

4 Theoretical Analysis

This section develops a one-sided converse for positional identifiability. The argument reduces localization to observation-map image size, bounds the quantized spectral-energy code space by normalization and trimming, and yields a graph-dependent impossibility result. We then specialize it to random regular graphs with explicit subcritical budget conditions. Proofs are deferred to Appendix A.

4.1 Observation-map identity

We begin with a counting identity for deterministic observation maps. This reduction is independent of the particular choice of distance or spectral features.

Lemma 1. *Fix a graph $G = (V, E)$ and an anchor set $\mathcal{A} \subset V$. Let $v_\star \sim \text{Unif}(V)$ and define*

$$Y := F_{G,\mathcal{A}}^{(m,\eta)}(v_\star). \quad (21)$$

Then for any reconstruction rule

$$s_{G,\mathcal{A}} : \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)} \rightarrow V, \quad (22)$$

one has

$$\mathbb{P}(s_{G,\mathcal{A}}(Y) = v_\star \mid G, \mathcal{A}) \leq \mathbb{E} \left[\frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)|} \mid G, \mathcal{A} \right]. \quad (23)$$

Moreover,

$$\mathbb{E} \left[\frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)|} \mid G, \mathcal{A} \right] = \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \quad (24)$$

The upper bound is attained by any measurable section that selects one representative from each nonempty preimage. Consequently,

$$\text{Err}^\star(F_{G,\mathcal{A}}^{(m,\eta)}) = 1 - \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \quad (25)$$

The same fiber-counting argument also gives a task-level version, which will be used to evaluate ambiguity in downstream node-label tasks.

Corollary 2 (Task-level ambiguity). *Fix $(G, \mathcal{A})$ and write*

$$F = F_{G,\mathcal{A}}^{(m,\eta)}. \quad (26)$$

Let $\ell : V \rightarrow \mathcal{L}$ be any finite task-label map. For predictors

$$h : \text{Im}(F) \rightarrow \mathcal{L} \quad (27)$$

that observe only the positional code $F(v_\star)$, define

$$\text{Err}_\ell^\star(F) := \inf_h \mathbb{P}(h(F(v_\star)) \neq \ell(v_\star) \mid G, \mathcal{A}, \ell). \quad (28)$$

Then

$$\begin{aligned} & \text{Err}_\ell^\star(F) \\ &= 1 - \frac{1}{n} \sum_{y \in \text{Im}(F)} \max_{\lambda \in \mathcal{L}} |\{v \in V : F(v) = y, \ell(v) = \lambda\}|. \end{aligned} \quad (29)$$

In particular, taking $\ell(v) = v$ recovers

$$\text{Err}^\star(F) = 1 - \frac{|\text{Im}(F)|}{n}. \quad (30)$$

Lemma 1 shows that every converse bound below can be obtained by upper bounding the number of distinct hybrid positional codes. Corollary 2 gives the corresponding task-level ambiguity: if a code fiber mixes multiple task labels, then no predictor using only the positional code can separate those labels perfectly. In the experiments, this provides a direct diagnostic for ambiguity in NLP and symbolic-reasoning node-label tasks.

4.2 Spectral trimming and normalized converse

We next control the spectral part of the observation map. The argument uses only the normalized energy coordinates in (8) and the orthonormality of the Laplacian eigenvectors.

Lemma 3. *For a trimming threshold $L > 0$, define*

$$V_L := \{v \in V : \max_{1 \leq j \leq m} X_m(v)_j \leq L\}. \quad (31)$$

Then

$$|V \setminus V_L| \leq \frac{mn}{L}. \quad (32)$$

Moreover,

$$|Q_\eta(X_m(V_L))| \leq \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m. \quad (33)$$

Combining the distance partition induced by $\mathcal{A}$ with Lemma 3 gives the following one-sided converse.

Theorem 4. *For any finite connected graph $G = (V, E)$, anchor set $\mathcal{A} \subset V$, spectral dimension $m \in \{1, \dots, n-1\}$, quantization level $\eta > 0$, and trimming threshold $L > 0$, the normalized hybrid observation map $F_{G,\mathcal{A}}^{(m,\eta)}$ satisfies*

$$|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})| \leq D(G, \mathcal{A}) \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m + \frac{mn}{L}. \quad (34)$$

Consequently,

$$\text{Err}^*(F_{G,\mathcal{A}}^{(m,\eta)}) \geq 1 - \frac{D(G,\mathcal{A})}{n} \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m - \frac{m}{L}. \quad (35)$$

Define the converse code budget

$$B_{\text{conv}}(G, \mathcal{A}, m, \eta, L) := \log D(G, \mathcal{A}) + m \log \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right), \quad (36)$$

and the corresponding normalized converse complexity

$$\Gamma_{\text{conv}}(G, \mathcal{A}, m, \eta, L) := \exp(B_{\text{conv}}(G, \mathcal{A}, m, \eta, L) - \log n) + \frac{m}{L}. \quad (37)$$

Then (35) can be written compactly as

$$\text{Err}^*(F_{G,\mathcal{A}}^{(m,\eta)}) \geq 1 - \Gamma_{\text{conv}}(G, \mathcal{A}, m, \eta, L). \quad (38)$$

Thus, if the effective code budget is below $\log n$ while the trimming penalty m/L vanishes, the hybrid positional encoding cannot localize a uniformly random vertex with vanishing error.

4.3 Random-regular subcritical regime

We now specialize the normalized converse to random regular graphs. The only additional input is a bound on the number of attainable anchor-distance profiles.

Lemma 5. *If $\text{diam}(G) \leq R$, then for any anchor set $\mathcal{A}$ with $|\mathcal{A}| = k$,*

$$D(G, \mathcal{A}) \leq (R+1)^k. \quad (39)$$

In particular, if $G_n \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$, then there exists a constant $C_r > 0$, depending only on r , such that

$$D(G_n, \mathcal{A}_n) \leq (C_r \log n + 1)^{k_n} \quad (40)$$

with probability tending to one.

Combining Lemma 5 with Theorem 4 yields the following subcritical regime.

Corollary 6 (Random-regular subcritical regime). *Fix $r \geq 3$ and let $G_n \sim \mathcal{G}_{n,r}$. Let $\mathcal{A}_n \subset V(G_n)$ be sampled uniformly without replacement with $|\mathcal{A}_n| = k_n$. For any sequences*

$$m_n \in \{1, \dots, n-1\}, \quad \eta_n > 0, \quad L_n > 0, \quad (41)$$

the normalized hybrid observation map satisfies, with probability tending to one,

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \geq 1 - \frac{(C_r \log n + 1)^{k_n}}{n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} - \frac{m_n}{L_n}. \quad (42)$$

In particular, if

$$\frac{m_n}{L_n} \rightarrow 0 \quad (43)$$

and

$$(C_r \log n + 1)^{k_n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} = o(n), \quad (44)$$

then

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (45)$$

The next corollary gives a logarithmic sufficient condition.

Corollary 7 (Logarithmic budget form). *In the setting of Corollary 6, take $L_n = \log \log n$. If*

$$m_n = o(\log \log n), \quad (46)$$

and, for some fixed $\varepsilon > 0$ and all sufficiently large n ,

$$k_n \log(C_r \log n + 1) + m_n \log \left(\left\lceil \frac{\log \log n}{\eta_n} \right\rceil + 1 \right) \leq (1 - \varepsilon) \log n, \quad (47)$$

then

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (48)$$

If $0 < \eta_n \leq 1$, the condition in (47) is implied, up to lower-order terms absorbed by the slack $\varepsilon \log n$, by

$$k_n \log \log n + m_n \log \frac{1}{\eta_n} + m_n \log \log \log n \leq (1 - \varepsilon) \log n. \quad (49)$$

Corollaries 6 and 7 are one-sided impossibility results rather than sharp thresholds. Importantly, Eq. (25) is exact for any deterministic observation map: once the induced code partition is fixed, localization ambiguity is exactly determined by its image size. One-sidedness enters only when Theorem 4 upper-bounds $|\text{Im}(F)|$ through distance-profile counting and spectral-energy trimming. Hence the resulting budget conditions are sufficient for failure, not necessary-and-sufficient conditions for success.

For graph-based NLP and symbolic reasoning, the result should therefore be interpreted as an encoding-level necessary-condition diagnostic: if linguistic, entity, token, or reasoning-state nodes collide under $F_{G,A}^{(m,\eta)}$, no downstream predictor whose positional information factors through this code can resolve those collisions from positional information alone.

5 Empirical Evaluation

5.1 Experimental setup

We evaluate encoding-level identifiability on synthetic graphs, NLP-derived graph skeletons, and external real-graph case studies. The object is the deterministic observation map

$$F(v) = (d_A(v), Q_\eta(S_m(v))), \quad (50)$$

where $d_A(v)$ is the anchor-distance profile and $Q_\eta(S_m(v))$ is the relatively quantized low-frequency spectral code. Unless otherwise stated, anchors are sampled uniformly without replacement, the spectral component uses the first m non-trivial eigenvectors of the normalized Laplacian, and quantization is relative.

For the synthetic study, we use random 3-regular graphs with

$$\begin{aligned} n &\in \{500, 1000, 2000, 4000\}, \\ k &\in \{1, 2, 3, 4, 6, 8\}, \\ m &\in \{0, 1, 2, 5\}, \\ \eta &\in \{0.9, 0.7, 0.5, 0.3, 0.1\}, \end{aligned} \quad (51)$$

averaged over 20 graph trials. The case $m = 0$ is the distance-only baseline. We organize the results by

$$\rho_{\text{eng}} = \frac{k \log \log n + m \log(2/\eta)}{\log n}, \quad (52)$$

and report

$$k_{\text{emp}}(n, m, \eta) = \min \{k : \mathbb{E}[\text{err}(F)] \leq 0.1\}. \quad (53)$$

For the NLP-derived study, we use CLUTRR family-relation reasoning graphs and Universal Dependencies dependency trees. In CLUTRR, nodes are entities and edges are explicit family-relation facts; we evaluate both clean and distractor graphs and group results by relation length. In UD, nodes are tokens and edges are undirected head-dependent arcs; we use Chinese-GSD, English-EWT, and French-GSD. Edge directions, edge labels, lexical features, and task-specific node attributes are removed before computing the positional code.

On these graphs, we compare NoPE, Distance-only, Spectral-only, and Hybrid encodings. The main setting uses

$$k = 1, \quad m = 8, \quad \eta = 0.25, \quad (54)$$

and a stronger setting $k = 4, m = 8, \eta = 0.25$ is reported in Appendix C.2. Task-level ambiguity is evaluated using query role, source distance, and target distance for CLUTRR, and UPOS, DEPREL, and root depth for UD.

We also retain two DDI graphs as external case studies:

$$\begin{aligned} (N_{\text{DB}}, E_{\text{DB}}) &= (1684, 189774), \\ (N_{\text{DEC}}, E_{\text{DEC}}) &= (40, 466). \end{aligned} \quad (55)$$

For these fixed graphs, we evaluate

$$\begin{aligned} k &\in \{1, 2, 3, 4, 6, 8\}, \\ m &\in \{0, 2, 5, 10\}, \\ \eta &\in \{0.9, 0.5, 0.25, 0.1\}, \end{aligned} \quad (56)$$

averaged over 30 random anchor samplings.

The primary node-level metric is

$$\text{succ}(F) = \frac{|\text{Im}(F)|}{n}, \quad \text{err}(F) = 1 - \text{succ}(F). \quad (57)$$

Smaller error indicates stronger node identifiability. For NLP-derived graphs, we also report the task-level ambiguity from Corollary 2.

Full construction details and additional diagnostics are deferred to Appendix B. The primary ambiguity experiments are deliberately training-free and PE-only: they evaluate information supplied by the positional channel without lexical features, labeled edges, or a downstream task model. The lightweight UD structural probes in Section 5.3 are reported separately as task-relevant calibration and use only positional codes as inputs. Code, preprocessing scripts, baselines, and hyperparameter configurations are available at https://anonymous.4open.science/r/Graph_Node_Identifiability_via_Observation_Maps-5862.

5.2 Synthetic validation and mechanism diagnostics

Figure 2 validates the theory-guided budget on random 3-regular graphs. Localization error drops sharply in the transition region, and the mean preimage size collapses toward one when plotted against

ρ_{eng} , showing that the empirical crossover corresponds to shrinkage of observation-map fibers.

Across $m \in \{0, 1, 2, 5\}$, ρ_{eng} organizes the error curves more consistently than the raw anchor count. The crossover remains near $\rho_{\text{eng}} \approx 1$, suggesting an empirical scaling law consistent with the one-sided converse rather than a sharp threshold theorem.

As a lightweight positive-side calibration, let U, V be distinct uniformly sampled nodes and define $\kappa_H = \Pr(F(U) = F(V))$, $I_H = -\log \kappa_H$, $h = \frac{I_H}{\log n}$. Here h measures collision information relative to the $\log n$ scale needed to distinguish n nodes. Across 25,256 Hybrid configurations over five graph families and seven settings, the empirical rule $h \geq 1$ predicts $\text{Err}^*(F) \leq 0.1$ with precision 0.930 and recall 0.9998. In graph-clustered validation over 129,360 configurations from 210 graph-instance clusters, h achieves AUC 0.9995; at the fixed threshold $h \geq 1$, precision, recall, and F1 are 0.9393, 0.9999, and 0.9685, respectively. We use this only as empirical positive-side calibration; it is not an achievability theorem or a claim of a sharp threshold.

The empirical threshold k_{emp} follows the same trend. Distance-only features give $k_{\text{emp}} = 6$ for all tested n and η , while adding spectral codes lowers the threshold. At $\eta = 0.1$, $m = 5$ gives $k_{\text{emp}} = 1$ for all tested graph sizes. Full diagnostics are reported in Appendix C.

Because Hybrid concatenates the two component codes, its comparison at a fixed (k, m, η) is a refinement diagnostic rather than a budget-controlled dominance claim. We therefore compare, at each attained threshold of ρ_{eng} , the best Distance, Spectral, and Hybrid configuration available under the same or smaller normalized budget. Across the 87 thresholds where all three classes are feasible, Hybrid has the lowest mean error at 63 thresholds and ties at 11 more (74/87, 85.1%). Spectral-only is better at some low/intermediate thresholds before the next Hybrid allocation becomes feasible. This supports complementary refinement under matched budget, not strict dominance at every discrete budget.

Table 1 confirms the bucketwise refinement mechanism. From the collision-heavy to the near-injective regime, within-bucket collision drops from 0.727 to 4.45×10^{-4} , while the median normalized code count rises from 0.185 to 0.9999. Thus, anchor distances provide a coarse partition, and spectral codes resolve residual collisions

within distance buckets.

5.3 Positional Identifiability on NLP-Derived Graphs

We next test whether the observation-map diagnostic extends to NLP-derived graph skeletons. We use CLUTRR family-relation graphs and UD dependency trees after removing edge labels, directions, lexical features, and task-specific node attributes. The main low-budget setting is $(k, m, \eta) = (1, 8, 0.25)$.

Table 2 reports node-level ambiguity and macro task-level ambiguity. Node Err is $\text{err}(F)$. Task Err averages $\text{err}_\ell(F)$ over query role and source/target distance for CLUTRR, and over UPOS, DEPREL, and root depth for UD.

Hybrid gives the lowest ambiguity on all datasets. Relative to Distance-only, it reduces node error by 64.7% on CLUTRR-clean, 73.9% on CLUTRR-distractor, and 72.2%-75.9% on UD, with corresponding task-error reductions of 65.9%, 76.2%, and 73.9%-77.8%. This instantiates Corollary 2: positional collisions can mix reasoning and syntactic labels.

To connect ambiguity with task-relevant structure without reintroducing lexical features or edge labels, we additionally fit PE-only UD structural probes. Hybrid improves root-depth NMAE to 0.173, versus 0.190 for Distance and 0.196 for Spectral, and improves pairwise tree-distance Macro-F1 to 0.812, versus 0.642 and 0.532. A PE-row derangement control preserves marginal PE statistics while breaking node-code alignment; the larger alignment-specific margin for Hybrid indicates that the gain comes from syntactic-tree alignment rather than marginal feature distributions. Per-label ambiguity and stronger-budget results are reported in Appendices C.1 and C.2.

5.4 Robustness and external graph case studies

We first stress-test the spectral representation under repeated or near-repeated eigenspaces by comparing coordinate energy with blockwise projector energy and random basis rotations. On the main random 3-regular graphs, coordinate and projector energies give identical ambiguity at the reported precision. On symmetric stress-test graphs, projector-versus-coordinate gaps can be large (up to 0.915 for Spectral and 0.938 for Hybrid), confirming that coordinate energy can retain basis-dependent distinctions and that projector energy is the appropriate

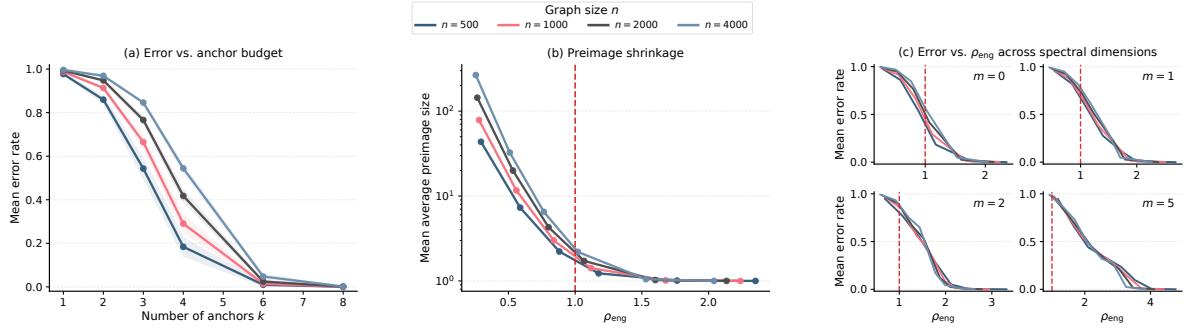


Figure 2: Synthetic validation on random 3-regular graphs. (a) Mean localization error versus the number of anchors k . (b) Mean average preimage size versus the budget ratio ρ_{eng} , showing the collapse of observation-map fibers in the low-error regime. (c) Error curves plotted against ρ_{eng} for spectral dimensions $m \in \{0, 1, 2, 5\}$. Across spectral dimensions, the empirical crossover is more consistently organized by the theory-guided budget ratio than by the anchor number alone.

Table 1: Bucketwise refinement diagnostics on random 3-regular graphs with $n = 2000$. Values are averaged over 20 graph trials and 10 anchor resamples per trial.

Regime	Overall Error	Singleton Frac.	Weighted Collision	Median $M(B_k)/ B_k $	$q_{0.9}$ Balance	Large Coll. (≥ 3)	Large Coll. (≥ 10)	Large Median $M/ B $ ($\geq 3, \geq 10$)
Collision-heavy	0.894	0.202	0.727	0.185	3.23	0.727	0.727	0.154, 0.093
Transition	0.703	0.202	0.253	0.465	6.97	0.253	0.253	0.413, 0.311
Near-injective	0.012	0.202	4.45×10^{-4}	0.9999	1.89	4.45×10^{-4}	4.46×10^{-4}	0.9999, 0.9999

Table 2: Main results on NLP-derived graph skeletons under $(k, m, \eta) = (1, 8, 0.25)$. Values are means over graph instances and random anchor draws. Lower is better.

Dataset	Node Err			Task Err		
	Dist.	Spec.	Hybrid	Dist.	Spec.	Hybrid
CLUTRR-clean	0.167	0.402	0.059	0.144	0.333	0.049
CLUTRR-distractor	0.261	0.233	0.068	0.192	0.188	0.046
UD-English	0.617	0.183	0.148	0.401	0.112	0.089
UD-Chinese	0.683	0.193	0.170	0.412	0.104	0.091
UD-French	0.678	0.215	0.189	0.432	0.129	0.113

Table 3: Representation ablation and anchor strategy on random 3-regular graphs.

Setting	Description	Error	Preimg.	Single.	Coll.
Representation ablation					
NoPE	no positional encoding	0.999	1166.7	0.000	1.000
Spectral	spectral only	0.933	25.00	0.261	0.198
Distance	anchor distances only	0.291	1.439	0.743	0.0009
Hybrid	distance + spectral	0.079	1.091	0.931	0.0002
Anchor strategy (Hybrid)					
Random	uniform anchors	0.094	1.112	0.920	0.0002
Farthest	farthest-point heuristic	0.083	1.097	0.926	0.0002
Degree	degree-based heuristic	0.076	1.087	0.934	0.0002

invariant formulation when degeneracy matters.

We also retain the DDI graphs only as external graph-dependence case studies. Under $(k, m, \eta) = (8, 10, 0.1)$, DrugBank remains highly colliding, whereas the Decagon-derived graph becomes nearly injective. Full DDI statistics are reported in Appendix C.5.

Table 3 isolates the contribution of each encoding component on random 3-regular graphs. NoPE is uninformative and Spectral-only remains weak in this setting. Distance-only gives a coarse partition, while the Hybrid map sharply reduces ambiguity. Anchor choice has a smaller effect than the representation choice.

Figure 3 shows that the improvement is not tied to a single parameter choice. At fixed $k = 4$, finer quantization reduces error only when spectral dimensions are available. At fixed $m = 5$, increas-

ing the number of anchors produces a sharp drop in ambiguity, with finer quantization reaching the low-error regime earlier.

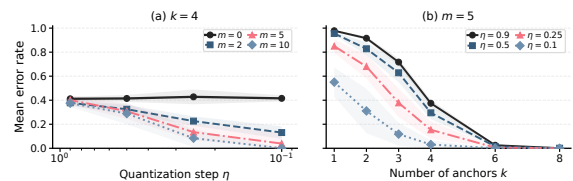


Figure 3: Hyperparameter interaction on random 3-regular graphs. Left: mean error versus quantization step η at fixed $k = 4$. Right: mean error versus number of anchors k at fixed $m = 5$.

Limitations

We note several limitations and practical considerations.

Theory scope. The exact image–preimage identities apply to any deterministic finite-resolution observation map, but the trimming/counting converse and random-regular scaling are specific to the anchor-distance plus quantized spectral family (including the projector-energy variant). They provide sufficient conditions for failure, not matching achievability guarantees or a sharp transition theorem. The present analysis also does not model directed, weighted, temporal, or training-dependent positional channels.

Graph abstraction. For NLP-derived graphs, the diagnostic is applied to unlabeled undirected graph skeletons. Edge directions, edge labels, lexical features, and task-specific node attributes are removed to isolate the positional code. Thus, a code collision indicates a limitation of positional information alone, not necessarily a failure of a full NLP model using lexical, semantic, or labeled-edge features.

Quantization. The theory uses an absolute-resolution quantizer on normalized spectral energy coordinates, while the main experiments use calibrated relative quantization for comparability across graph scales. The empirical resolution parameter should therefore be interpreted as a controlled analogue of the theoretical one. A tighter analysis of scale-calibrated quantizers remains future work.

Empirical scope. The evaluation covers representative regimes: random regular graphs for theory-guided scaling, CLUTRR relation graphs for entity-level reasoning, Universal Dependencies trees for token-level syntax, and DDI graphs as external case studies. These settings support the diagnostic across synthetic and NLP-derived graphs, but do not exhaust all graph-based NLP structures, such as AMR, discourse, knowledge, or retrieval graphs.

Downstream modeling. The experiments measure identifiability and PE-only structural recoverability rather than end-to-end task accuracy. Complementary fixed-backbone evidence in Yan et al. (2026) shows that positional information can affect downstream DDI prediction when the PE module is varied, but our diagnostic does not predict full-model accuracy, which also depends on features, architecture, supervision, and optimization.

References

- Mitchell Black, Zhengchao Wan, Gal Mishne, Amir Nayyeri, and Yusu Wang. 2024. [Comparing graph transformers via positional encodings](#). In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 4103–4139. PMLR.
- Rickard Brüel-Gabrielsson, Mikhail Yurochkin, and Justin Solomon. 2023. [Rewiring with positional encodings for graph neural networks](#). *Transactions on Machine Learning Research*.
- Semih Cantürk, Renming Liu, Olivier Lapointe-Gagné, Vincent Létourneau, Guy Wolf, Dominique Beaini, and Ladislav Rampásek. 2024. [Graph positional and structural encoder](#). In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 5533–5566. PMLR.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. [Universal Dependencies](#). *Computational Linguistics*, 47(2):255–308.
- 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.
- Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. 2022. [Graph neural networks with learnable structural and positional representations](#). In *International Conference on Learning Representations*.
- Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai Maron. 2023. [Graph positional encoding via random feature propagation](#). In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 9202–9223. PMLR.
- Billy Joe Franks, Moshe Eliasof, Semih Cantürk, Guy Wolf, Carola-Bibiane Schönlieb, Sophie Fellenz, and Marius Kloft. 2025. [Towards graph foundation models: A study on the generalization of positional and structural encodings](#). *Transactions on Machine Learning Research*.
- Florian Grötschla, Jiaqing Xie, and Roger Wattenhofer. 2024. [Benchmarking positional encodings for GNNs and graph transformers](#). *CoRR*, abs/2411.12732. Preprint.
- Yu He, Cristian Bodnar, and Pietro Liò. 2024. [Sheaf-based positional encodings for graph neural networks](#). In *Proceedings of the 2nd NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, volume 228 of *Proceedings of Machine Learning Research*, pages 1–18. PMLR.

- Yinan Huang, William Lu, Joshua Robinson, Yu Yang, Muhan Zhang, Stefanie Jegelka, and Pan Li. 2024. [On the stability of expressive positional encodings for graphs](#). In *International Conference on Learning Representations*.
- Yinan Huang, Haoyu Peter Wang, and Pan Li. 2025. [What are good positional encodings for directed graphs?](#) In *International Conference on Learning Representations*.
- Charilaos I. Kanatsoulis, Evelyn Choi, Stefanie Jegelka, Jure Leskovec, and Alejandro Ribeiro. 2025. [Learning efficient positional encodings with graph neural networks](#). In *International Conference on Learning Representations*.
- Nicolas Keriven and Samuel Vaiter. 2023. [What functions can graph neural networks compute on random graphs? the role of positional encoding](#). In *Advances in Neural Information Processing Systems*, volume 36.
- Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. 2021. [Rethinking graph transformers with spectral attention](#). In *Advances in Neural Information Processing Systems*, volume 34, pages 21618–21629.
- Hongkang Li, Meng Wang, Tengfei Ma, Sijia Liu, Zaixi Zhang, and Pin-Yu Chen. 2024. [What improves the generalization of graph transformers? A theoretical dive into the self-attention and positional encoding](#). In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 28784–28829. PMLR.
- 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.
- Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess E. Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. 2023. [Sign and basis invariant networks for spectral graph representation learning](#). In *International Conference on Learning Representations*.
- Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H. S. Torr, and Ser-Nam Lim. 2023. [Graph inductive biases in transformers without message passing](#). In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 23321–23337. PMLR.
- Diego Marcheggiani and Ivan Titov. 2017. [Encoding sentences with graph convolutional networks for semantic role labeling](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, Copenhagen, Denmark. Association for Computational Linguistics.
- Grégoire Mialon, Dexiong Chen, Margot Selosse, and Julien Mairal. 2021. [GraphiT: Encoding graph structure in transformers](#). *CoRR*, abs/2106.05667. Preprint.
- Wonpyo Park, Woonggi Chang, Donggeon Lee, Juntae Kim, and Seung won Hwang. 2022. [GRPE: Relative positional encoding for graph transformer](#). *CoRR*, abs/2201.12787. Preprint.
- Ladislav Rampášek, Michael 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*, volume 35.
- Koustuv Sinha, Shagun Sodhani, Jin Dong, Joelle Pineau, and William L. Hamilton. 2019. [CLUTRR: A diagnostic benchmark for inductive reasoning from text](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4506–4515, Hong Kong, China. Association for Computational Linguistics.
- Slimane Thabet, Mehdi Djellabi, Igor Olegovich Sokolov, Sachin Kasture, Louis-Paul Henry, and Loïc Henriët. 2024. [Quantum positional encodings for graph neural networks](#). In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 47965–47996. PMLR.
- Haorui Wang, Haoteng Yin, Muhan Zhang, and Pan Li. 2022. [Equivariant and stable positional encoding for more powerful graph neural networks](#). In *International Conference on Learning Representations*.
- Zimo Yan, Zheng Xie, Runfan Duan, Chang Liu, and Wumei Du. 2026. [Bridging distance and spectral positional encodings via anchor-based diffusion geometry approximation](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13066–13085, San Diego, California, United States. Association for Computational Linguistics.
- Jeyoon Yeom, Taero Kim, Rakwoo Chang, and Kyungwoo Song. 2024. [Structural and positional ensembled encoding for graph transformer](#). *Pattern Recognition Letters*, 183:104–110.
- 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.

- Chaozhao Yuan, Kangfei Zhao, Ercan Engin Kuruoglu, Liang Wang, Tingyang Xu, Wenbing Huang, Deli Zhao, Hong Cheng, and Yu Rong. 2025. [A survey of graph transformers: Architectures, theories and applications](#). *CoRR*, abs/2502.16533. Preprint.
- Bohang Zhang, Lingxiao Zhao, and Haggai Maron. 2024. [On the expressive power of spectral invariant graph neural networks](#). In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 60496–60526. PMLR.
- Jiawei Zhang, Haopeng Zhang, Congying Xia, and Li Sun. 2020. [Graph-Bert: Only attention is needed for learning graph representations](#). *CoRR*, abs/2001.05140. Preprint.
- Jie Zhang, Mingxuan Li, Yitai Xu, Hua He, Qun Li, and Tao Wang. 2025. [StrucGCN: Structural enhanced graph convolutional networks for graph embedding](#). *Information Fusion*, 117:102893.

A Proofs for Section 4

This appendix provides the full derivations for the results stated in Section 4. Each result is restated before its proof. All notation follows Sections 3 and 4. Throughout, $G = (V, E)$ is a finite connected graph with $n = |V|$, $\mathcal{A} \subseteq V$ is an anchor set, and $F_{G,\mathcal{A}}^{(m,\eta)}$ denotes the normalized hybrid observation map in (14).

A.1 Proof of Lemma 1

Lemma 1 (Lemma 1). *Fix a graph $G = (V, E)$ and an anchor set $\mathcal{A} \subset V$. Let $v_\star \sim \text{Unif}(V)$ and define*

$$Y := F_{G,\mathcal{A}}^{(m,\eta)}(v_\star). \quad (58)$$

Then for any reconstruction rule

$$s_{G,\mathcal{A}} : \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)} \rightarrow V, \quad (59)$$

one has

$$\begin{aligned} & \mathbb{P}(s_{G,\mathcal{A}}(Y) = v_\star \mid G, \mathcal{A}) \\ & \leq \mathbb{E} \left[\frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)|} \mid G, \mathcal{A} \right]. \end{aligned} \quad (60)$$

Moreover,

$$\mathbb{E} \left[\frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)|} \mid G, \mathcal{A} \right] = \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \quad (61)$$

The upper bound is attained by any measurable section that selects one representative from each nonempty preimage. Consequently,

$$\text{Err}^\star(F_{G,\mathcal{A}}^{(m,\eta)}) = 1 - \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \quad (62)$$

Proof. Fix $(G, \mathcal{A})$. For each $y \in \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)}$, define the fiber

$$C_y := \mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(y) = \{v \in V : F_{G,\mathcal{A}}^{(m,\eta)}(v) = y\}. \quad (63)$$

The nonempty fibers $\{C_y\}_{y \in \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)}}$ form a partition of V .

Conditional on $Y = y$, the source vertex $v_\star$ is uniform on C_y . Therefore, for any reconstruction rule $s_{G,\mathcal{A}}$,

$$\mathbb{P}(s_{G,\mathcal{A}}(Y) = v_\star \mid Y = y, G, \mathcal{A}) = \frac{\mathbf{1}_{\{s_{G,\mathcal{A}}(y) \in C_y\}}}{|C_y|}. \quad (64)$$

Averaging over Y gives

$$\begin{aligned} \mathbb{P}(s_{G,\mathcal{A}}(Y) = v_\star \mid G, \mathcal{A}) \\ = \mathbb{E} \left[\frac{\mathbf{1}_{\{s_{G,\mathcal{A}}(Y) \in \mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)\}}}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(Y)|} \mid G, \mathcal{A} \right]. \end{aligned} \quad (65)$$

Dropping the indicator yields the first inequality.

It remains to compute the expectation. Since $v_\star$ is uniform on V ,

$$\begin{aligned} \mathbb{E} \left[\frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v_\star)|} \mid G, \mathcal{A} \right] &= \frac{1}{n} \sum_{v \in V} \frac{1}{|\mathcal{P}_{G,\mathcal{A}}^{(m,\eta)}(v)|} \\ &= \frac{1}{n} \sum_{y \in \mathcal{Y}_{G,\mathcal{A}}^{(m,\eta)}} \sum_{v \in C_y} \frac{1}{|C_y|} \\ &= \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \end{aligned} \quad (66)$$

If $\tilde{s}_{G,\mathcal{A}}$ selects one representative from each nonempty fiber, then $\tilde{s}_{G,\mathcal{A}}(y) \in C_y$ for every y , and the upper bound is attained. Thus the optimal success probability is $|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|/n$, which proves the error identity. $\square$

A.2 Proof of Corollary 2

Corollary 1 (Task-level ambiguity; Corollary 2). *Fix $(G, \mathcal{A})$ and write*

$$F = F_{G,\mathcal{A}}^{(m,\eta)}. \quad (67)$$

Let $\ell : V \rightarrow \mathcal{L}$ be any finite task-label map. For predictors

$$h : \text{Im}(F) \rightarrow \mathcal{L} \quad (68)$$

that observe only the positional code $F(v_\star)$, define

$$\text{Err}_\ell^\star(F) := \inf_h \mathbb{P}(h(F(v_\star)) \neq \ell(v_\star) \mid G, \mathcal{A}, \ell). \quad (69)$$

Then

$$\begin{aligned} & \text{Err}_\ell^\star(F) \\ &= 1 - \frac{1}{n} \sum_{y \in \text{Im}(F)} \max_{\lambda \in \mathcal{L}} |\{v \in V : F(v) = y, \ell(v) = \lambda\}|. \end{aligned} \quad (70)$$

In particular, taking $\ell(v) = v$ recovers

$$\text{Err}^\star(F) = 1 - \frac{|\text{Im}(F)|}{n}. \quad (71)$$

Proof. Fix $(G, \mathcal{A}, \ell)$. For each $y \in \text{Im}(F)$, define the fiber

$$C_y := \{v \in V : F(v) = y\}. \quad (72)$$

The nonempty fibers $\{C_y\}_{y \in \text{Im}(F)}$ form a partition of V .

Conditional on $F(v_\star) = y$, the source vertex $v_\star$ is uniform on C_y . Hence, for any predictor h ,

$$\begin{aligned} \mathbb{P}(h(F(v_\star)) = \ell(v_\star) \mid F(v_\star) = y, G, \mathcal{A}, \ell) \\ = \frac{|\{v \in C_y : \ell(v) = h(y)\}|}{|C_y|}. \end{aligned} \quad (73)$$

The best choice of $h(y)$ is a majority label in C_y , so

$$\begin{aligned} \sup_h \mathbb{P}(h(F(v_\star)) = \ell(v_\star) \mid F(v_\star) = y, G, \mathcal{A}, \ell) \\ = \frac{\max_{\lambda \in \mathcal{L}} |\{v \in C_y : \ell(v) = \lambda\}|}{|C_y|}. \end{aligned} \quad (74)$$

Averaging over $y \in \text{Im}(F)$ gives

$$\begin{aligned} \sup_h \mathbb{P}(h(F(v_\star)) = \ell(v_\star) \mid G, \mathcal{A}, \ell) \\ = \sum_{y \in \text{Im}(F)} \frac{|C_y|}{n} \cdot \frac{\max_{\lambda \in \mathcal{L}} |\{v \in C_y : \ell(v) = \lambda\}|}{|C_y|} \\ = \frac{1}{n} \sum_{y \in \text{Im}(F)} \max_{\lambda \in \mathcal{L}} |\{v \in V : F(v) = y, \ell(v) = \lambda\}|. \end{aligned} \quad (75)$$

Taking one minus the optimal success probability proves the identity.

If $\ell(v) = v$, then each nonempty fiber contributes exactly one correct identity label under the majority rule. Therefore the sum equals $|\text{Im}(F)|$, and the formula reduces to

$$\text{Err}^\star(F) = 1 - \frac{|\text{Im}(F)|}{n}. \quad (76)$$

□

A.3 Proof of Lemma 3

Lemma 2 (Lemma 3). *For a trimming threshold $L > 0$, define*

$$V_L := \{v \in V : \max_{1 \leq j \leq m} X_m(v)_j \leq L\}. \quad (77)$$

Then

$$|V \setminus V_L| \leq \frac{mn}{L}. \quad (78)$$

Moreover,

$$|Q_\eta(X_m(V_L))| \leq \left(\left\lceil \frac{L}{\eta} \right\rceil + 1\right)^m. \quad (79)$$

Proof. Recall that

$$X_m(v) = n(\phi_2(v)^2, \phi_3(v)^2, \dots, \phi_{m+1}(v)^2). \quad (80)$$

For each $j \in \{1, \dots, m\}$, define

$$A_j := \{v \in V : n\phi_{j+1}(v)^2 > L\}. \quad (81)$$

Then

$$V \setminus V_L \subseteq \bigcup_{j=1}^m A_j. \quad (82)$$

For fixed j , if $v \in A_j$, then $n\phi_{j+1}(v)^2 > L$, and hence

$$L|A_j| < \sum_{v \in A_j} n\phi_{j+1}(v)^2 \leq \sum_{v \in V} n\phi_{j+1}(v)^2 = n. \quad (83)$$

Thus $|A_j| \leq n/L$. By the union bound,

$$|V \setminus V_L| \leq \sum_{j=1}^m |A_j| \leq \frac{mn}{L}. \quad (84)$$

For the code-count bound, if $v \in V_L$, then each coordinate of $X_m(v)$ lies in $[0, L]$. Hence each coordinate of $Q_\eta(X_m(v))$ belongs to

$$\left\{0, 1, \dots, \left\lceil \frac{L}{\eta} \right\rceil\right\}. \quad (85)$$

Therefore,

$$Q_\eta(X_m(V_L)) \subseteq \left\{0, 1, \dots, \left\lceil \frac{L}{\eta} \right\rceil\right\}^m, \quad (86)$$

which gives

$$|Q_\eta(X_m(V_L))| \leq \left(\left\lceil \frac{L}{\eta} \right\rceil + 1\right)^m. \quad (87)$$

The proof is complete. □

A.4 Proof of Theorem 4

Theorem 1 (Theorem 4). *For any finite connected graph $G = (V, E)$, anchor set $\mathcal{A} \subset V$, spectral dimension $m \in \{1, \dots, n-1\}$, quantization level $\eta > 0$, and trimming threshold $L > 0$, the normalized hybrid observation map $F_{G, \mathcal{A}}^{(m, \eta)}$ satisfies*

$$\begin{aligned} |\text{Im}(F_{G, \mathcal{A}}^{(m, \eta)})| &\leq D(G, \mathcal{A}) \left(\left\lceil \frac{L}{\eta} \right\rceil + 1\right)^m \\ &\quad + \frac{mn}{L}. \end{aligned} \quad (88)$$

Consequently,

$$\begin{aligned} \text{Err}^\star(F_{G, \mathcal{A}}^{(m, \eta)}) &\geq 1 - \frac{D(G, \mathcal{A})}{n} \left(\left\lceil \frac{L}{\eta} \right\rceil + 1\right)^m \\ &\quad - \frac{m}{L}. \end{aligned} \quad (89)$$

Proof. Let V_L be as in Lemma 3. Decompose

$$V = V_L \cup (V \setminus V_L). \quad (90)$$

Then

$$\begin{aligned} |\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})| &\leq |F_{G,\mathcal{A}}^{(m,\eta)}(V_L)| + |F_{G,\mathcal{A}}^{(m,\eta)}(V \setminus V_L)| \\ &\leq |F_{G,\mathcal{A}}^{(m,\eta)}(V_L)| + |V \setminus V_L|. \end{aligned} \quad (91)$$

For $v \in V_L$,

$$F_{G,\mathcal{A}}^{(m,\eta)}(v) = (D_{\mathcal{A}}(v), Q_{\eta}(X_m(v))). \quad (92)$$

The first component belongs to $T_{G,\mathcal{A}}$, and the second belongs to $Q_{\eta}(X_m(V_L))$. Hence

$$F_{G,\mathcal{A}}^{(m,\eta)}(V_L) \subseteq T_{G,\mathcal{A}} \times Q_{\eta}(X_m(V_L)). \quad (93)$$

Therefore, by Lemma 3,

$$\begin{aligned} |F_{G,\mathcal{A}}^{(m,\eta)}(V_L)| &\leq |T_{G,\mathcal{A}}| \cdot |Q_{\eta}(X_m(V_L))| \\ &\leq D(G, \mathcal{A}) \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m. \end{aligned} \quad (94)$$

Again by Lemma 3,

$$|V \setminus V_L| \leq \frac{mn}{L}. \quad (95)$$

Combining (91), (94), and (95) gives

$$|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})| \leq D(G, \mathcal{A}) \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m + \frac{mn}{L}. \quad (96)$$

The error bound follows from Lemma 1:

$$\text{Err}^*(F_{G,\mathcal{A}}^{(m,\eta)}) = 1 - \frac{|\text{Im}(F_{G,\mathcal{A}}^{(m,\eta)})|}{n}. \quad (97)$$

Substituting the image-size bound yields

$$\begin{aligned} \text{Err}^*(F_{G,\mathcal{A}}^{(m,\eta)}) &\geq 1 - \frac{D(G, \mathcal{A})}{n} \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m \\ &\quad - \frac{m}{L}. \end{aligned} \quad (98)$$

The proof is complete. $\square$

A.5 Proof of Lemma 5

Lemma 3 (Lemma 5). *If $\text{diam}(G) \leq R$, then for any anchor set $\mathcal{A}$ with $|\mathcal{A}| = k$,*

$$D(G, \mathcal{A}) \leq (R+1)^k. \quad (99)$$

In particular, if $G_n \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$, then there exists a constant $C_r > 0$, depending only on r , such that

$$D(G_n, \mathcal{A}_n) \leq (C_r \log n + 1)^{k_n} \quad (100)$$

with probability tending to one.

Proof. If $\text{diam}(G) \leq R$, then for every anchor $a_i \in \mathcal{A}$ and every $v \in V$,

$$d_G(v, a_i) \in \{0, 1, \dots, R\}. \quad (101)$$

Thus each coordinate of $D_{\mathcal{A}}(v)$ has at most $R+1$ possible values, and therefore

$$D(G, \mathcal{A}) = |\{D_{\mathcal{A}}(v) : v \in V\}| \leq (R+1)^k. \quad (102)$$

For $G_n \sim \mathcal{G}_{n,r}$ with fixed $r \geq 3$, the standard diameter bound for fixed-degree random regular graphs gives a constant $C_r > 0$ such that

$$\begin{aligned} \mathbb{P}(G_n \text{ is connected and } \text{diam}(G_n) \leq C_r \log n) \\ \rightarrow 1. \end{aligned} \quad (103)$$

Applying the deterministic bound on this event with $R = C_r \log n$ gives

$$D(G_n, \mathcal{A}_n) \leq (C_r \log n + 1)^{k_n} \quad (104)$$

with probability tending to one. $\square$

A.6 Proof of Corollary 6

Corollary 2 (Random-regular subcritical regime; Corollary 6). *Fix $r \geq 3$ and let $G_n \sim \mathcal{G}_{n,r}$. Let $\mathcal{A}_n \subset V(G_n)$ be sampled uniformly without replacement with $|\mathcal{A}_n| = k_n$. For any sequences*

$$m_n \in \{1, \dots, n-1\}, \quad \eta_n > 0, \quad L_n > 0, \quad (105)$$

the normalized hybrid observation map satisfies, with probability tending to one,

$$\begin{aligned} \text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \\ \geq 1 - \frac{(C_r \log n + 1)^{k_n}}{n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} - \frac{m_n}{L_n}. \end{aligned} \quad (106)$$

In particular, if

$$\frac{m_n}{L_n} \rightarrow 0 \quad (107)$$

and

$$(C_r \log n + 1)^{k_n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} = o(n), \quad (108)$$

then

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (109)$$

Proof. By Lemma 5, with probability tending to one,

$$D(G_n, \mathcal{A}_n) \leq (C_r \log n + 1)^{k_n}. \quad (110)$$

On this event, Theorem 4 gives

$$\begin{aligned} & \text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \\ & \geq 1 - \frac{D(G_n, \mathcal{A}_n)}{n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} - \frac{m_n}{L_n} \\ & \geq 1 - \frac{(C_r \log n + 1)^{k_n}}{n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n} - \frac{m_n}{L_n}. \end{aligned} \quad (111)$$

This proves the stated high-probability lower bound.

With $b_n := \frac{m_n}{L_n}$, define

$$a_n := \frac{(C_r \log n + 1)^{k_n}}{n} \left(\left\lceil \frac{L_n}{\eta_n} \right\rceil + 1 \right)^{m_n}. \quad (112)$$

The assumptions imply $a_n \rightarrow 0$ and $b_n \rightarrow 0$. Hence, for every fixed $\delta > 0$, $a_n + b_n < \delta$ for all sufficiently large n . On the high-probability event in (110),

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \geq 1 - a_n - b_n > 1 - \delta. \quad (113)$$

Therefore,

$$\mathbb{P}(\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \leq 1 - \delta) \rightarrow 0, \quad (114)$$

which is equivalent to

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (115)$$

□

A.7 Proof of Corollary 7

Corollary 3 (Logarithmic budget form; Corollary 7). *In the setting of Corollary 6, take $L_n = \log \log n$. If*

$$m_n = o(\log \log n) \quad (116)$$

and, for some fixed $\varepsilon > 0$ and all sufficiently large n ,

$$\begin{aligned} k_n \log(C_r \log n + 1) + m_n \log \left(\left\lceil \frac{\log \log n}{\eta_n} \right\rceil + 1 \right) \\ \leq (1 - \varepsilon) \log n, \end{aligned} \quad (117)$$

then

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (118)$$

If $0 < \eta_n \leq 1$, the preceding logarithmic budget is implied, up to lower-order terms absorbed by the slack, by

$$\begin{aligned} k_n \log \log n + m_n \log \frac{1}{\eta_n} + m_n \log \log \log n \\ \leq (1 - \varepsilon) \log n. \end{aligned} \quad (119)$$

Proof. Set $L_n = \log \log n$. The condition

$$m_n = o(\log \log n) \quad (120)$$

implies

$$\frac{m_n}{L_n} \rightarrow 0. \quad (121)$$

Define

$$A_n := k_n \log(C_r \log n + 1) + m_n \log \left(\left\lceil \frac{\log \log n}{\eta_n} \right\rceil + 1 \right). \quad (122)$$

The assumed budget condition gives

$$A_n \leq (1 - \varepsilon) \log n \quad (123)$$

for all sufficiently large n , and hence

$$\begin{aligned} & \frac{(C_r \log n + 1)^{k_n} \left(\left\lceil \frac{\log \log n}{\eta_n} \right\rceil + 1 \right)^{m_n}}{n} \\ & = \frac{\exp(A_n)}{n} \leq n^{-\varepsilon} \rightarrow 0. \end{aligned} \quad (124)$$

Thus the two assumptions of Corollary 6 hold, and therefore

$$\text{Err}^*(F_{G_n, \mathcal{A}_n}^{(m_n, \eta_n)}) \xrightarrow{\mathbb{P}} 1. \quad (125)$$

It remains to verify the simplified sufficient condition. Assume $0 < \eta_n \leq 1$ for all sufficiently large n . Since

$$C_r \log n + 1 \leq (C_r + 1) \log n \quad (126)$$

for large n ,

$$\log(C_r \log n + 1) \leq \log \log n + \log(C_r + 1). \quad (127)$$

Let

$$x_n := \frac{\log \log n}{\eta_n}. \quad (128)$$

Then $x_n \geq 1$ for all sufficiently large n , and

$$\lceil x_n \rceil + 1 \leq x_n + 2 \leq 3x_n. \quad (129)$$

Therefore,

$$\begin{aligned} & \log \left(\left\lceil \frac{\log \log n}{\eta_n} \right\rceil + 1 \right) \\ & \leq \log 3 + \log \left(\frac{\log \log n}{\eta_n} \right) \\ & = \log 3 + \log \log \log n + \log \frac{1}{\eta_n}. \end{aligned} \quad (130)$$

Combining these bounds yields

$$\begin{aligned} A_n & \leq k_n \log \log n + m_n \log \frac{1}{\eta_n} + m_n \log \log \log n \\ & \quad + k_n \log(C_r + 1) + m_n \log 3. \end{aligned} \quad (131)$$

Under the simplified budget, the first three terms are at most $(1 - \varepsilon) \log n$. Moreover,

$$k_n \log \log n \leq (1 - \varepsilon) \log n \quad (132)$$

implies

$$k_n = O\left(\frac{\log n}{\log \log n}\right) = o(\log n), \quad (133)$$

and $m_n = o(\log \log n)$ implies $m_n = o(\log n)$. Hence

$$k_n \log(C_r + 1) + m_n \log 3 = o(\log n). \quad (134)$$

Thus

$$A_n \leq (1 - \varepsilon) \log n + o(\log n). \quad (135)$$

After reducing the slack, for example replacing ε by $\varepsilon/2$, we obtain

$$A_n \leq \left(1 - \frac{\varepsilon}{2}\right) \log n \quad (136)$$

for all sufficiently large n . Hence the simplified budget implies the exact logarithmic budget up to lower-order terms absorbed by the slack. The proof is complete. $\square$

A.8 Algebra for the Converse Complexity

Recall the definitions

$$B_{\text{conv}}(G, \mathcal{A}, m, \eta, L) := \log D(G, \mathcal{A}) + m \log \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right), \quad (137)$$

and

$$\Gamma_{\text{conv}}(G, \mathcal{A}, m, \eta, L) := \exp(B_{\text{conv}}(G, \mathcal{A}, m, \eta, L) - \log n) + \frac{m}{L}. \quad (138)$$

Then

$$\exp(B_{\text{conv}}(G, \mathcal{A}, m, \eta, L) - \log n) = \frac{D(G, \mathcal{A})}{n} \left(\left\lceil \frac{L}{\eta} \right\rceil + 1 \right)^m. \quad (139)$$

Therefore, Theorem 4 is equivalently written as

$$\text{Err}^*(F_{G, \mathcal{A}}^{(m, \eta)}) \geq 1 - \Gamma_{\text{conv}}(G, \mathcal{A}, m, \eta, L). \quad (140)$$

B Detailed Experimental Protocol

B.1 Datasets

We evaluate the proposed encoding on synthetic graphs, NLP-derived graphs, and external real-world graph case studies.

For the synthetic study, we use random 3-regular graphs to examine the phase-transition-like behavior predicted qualitatively by the converse theory. The case $m = 0$ is treated as a distance-only baseline. The parameter grid is

$$\begin{aligned} n &\in \{500, 1000, 2000, 4000\}, \\ k &\in \{1, 2, 3, 4, 6, 8\}, \\ m &\in \{0, 1, 2, 5\}, \\ \eta &\in \{0.9, 0.7, 0.5, 0.3, 0.1\}, \end{aligned} \quad (141)$$

with 20 independent graph trials for each configuration.

For the NLP-derived study, we use CLUTRR family-relation reasoning graphs and Universal Dependencies dependency trees. In CLUTRR, nodes are entities and edges are explicit family-relation facts; we evaluate both clean and distractor graphs, grouped by relation length. In UD, nodes are tokens and edges are undirected head-dependent arcs. We use Chinese-GSD, English-EWT, and French-GSD. Edge directions, edge labels, lexical features, and task-specific node attributes are removed before computing the positional code.

The main setting is

$$k = 1, \quad m = 8, \quad \eta = 0.25, \quad (142)$$

and the stronger robustness setting is

$$k = 4, \quad m = 8, \quad \eta = 0.25. \quad (143)$$

For the external real-graph study, we retain two DDI graph constructions. The DrugBank graph has

$$N_{\text{DB}} = 1684, \quad E_{\text{DB}} = 189774, \quad (144)$$

and the Decagon-derived graph has

$$N_{\text{DEC}} = 40, \quad E_{\text{DEC}} = 466. \quad (145)$$

For the DDI experiments, we evaluate

$$\begin{aligned} k &\in \{1, 2, 3, 4, 6, 8\}, \\ m &\in \{0, 2, 5, 10\}, \\ \eta &\in \{0.9, 0.5, 0.25, 0.1\}, \end{aligned} \quad (146)$$

averaged over 30 independent random anchor samplings.

B.2 Baselines

All baselines are defined at the feature level:

$$\begin{aligned} \text{NoPE} : \quad & F(v) = 0, \\ \text{Distance} : \quad & F(v) = d_A(v), \\ \text{Spectral} : \quad & F(v) = Q_\eta(S_m(v)), \\ \text{Hybrid} : \quad & F(v) = (d_A(v), Q_\eta(S_m(v))). \end{aligned} \quad (147)$$

For the synthetic ablations, we also compare random anchors, degree-based anchors, and farthest-point anchors under the hybrid feature.

B.3 Encoding Configuration

The experiments do not involve end-to-end learning. Instead, the object under evaluation is the deterministic observation map

$$F(v) = (d_A(v), Q_\eta(S_m(v))). \quad (148)$$

For each anchor set $A = \{a_1, \dots, a_k\}$, the distance component is

$$d_A(v) = (\text{SPD}(v, a_1), \dots, \text{SPD}(v, a_k)). \quad (149)$$

The spectral component is constructed from the normalized Laplacian

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2}, \quad (150)$$

using the first m non-trivial eigenvectors:

$$S_m(v) = (\Phi_m(v, 1)^2, \dots, \Phi_m(v, m)^2). \quad (151)$$

In all experiments, we use relative quantization. Given the spectral matrix S_m , we define

$$\Delta_\eta = \eta \cdot \max_{i,j} |(S_m)_{ij}|, \quad (152)$$

and quantize each coordinate by

$$(Q_\eta(x))_j = \Delta_\eta \cdot \text{round}\left(\frac{x_j}{\Delta_\eta}\right), \quad j \in [m]. \quad (153)$$

Thus, smaller η corresponds to finer relative quantization. Unless otherwise stated, all main-text experiments use the relative quantization rule in (152)-(153). Absolute quantization is discussed separately in this appendix as a calibration and robustness check.

For the synthetic phase-transition study, results are organized by

$$\rho_{\text{eng}} = \frac{k \log \log n + m \log(C_{\text{ent}}/\eta)}{\log n}, \quad C_{\text{ent}} = 2. \quad (154)$$

B.4 Evaluation Measures

For a fixed graph and parameter configuration, we define

$$\begin{aligned} \text{succ}(F) &= \frac{|\text{Im}(F)|}{n}, \\ \text{err}(F) &= 1 - \text{succ}(F). \end{aligned} \quad (155)$$

A smaller error indicates stronger node-level identifiability.

For NLP-derived graphs, we also report task-level ambiguity. Given a label map $\ell : V \rightarrow \mathcal{L}$, define

$$\text{err}_\ell(F) = 1 - \frac{1}{n} \sum_{y \in \text{Im}(F)} \max_{\lambda \in \mathcal{L}} |\{v : F(v) = y, \ell(v) = \lambda\}|. \quad (156)$$

For CLUTRR, ℓ is query role, source distance, or target distance. For UD, ℓ is UPOS, DEPREL, or root depth.

For the synthetic experiments, we further define

$$k_{\text{emp}}(n, m, \eta) = \min \{k : \mathbb{E}[\text{err}(F)] \leq 0.1\}. \quad (157)$$

Table 4: Robustness under the stronger setting $(k, m, \eta) = (4, 8, 0.25)$. Node Err is identity ambiguity; Task Err is macro-averaged over non-identity task labels. Values are means over graph instances and random anchor draws. Lower is better.

Benchmark	Dataset	Node Err			Task Err		
		Dist.	Spec.	Hybrid	Dist.	Spec.	Hybrid
CLUTRR	clean-task1	0.0007	0.4020	0.0003	0.0002	0.3335	0.0001
CLUTRR	distractor-task3	0.0073	0.2325	0.0038	0.0044	0.1880	0.0022
UD	Chinese-GSD	0.3294	0.1931	0.1233	0.1824	0.1039	0.0670
UD	English-EWT	0.2724	0.1834	0.0939	0.1644	0.1116	0.0566
UD	French-GSD	0.3354	0.2150	0.1392	0.2008	0.1291	0.0838

B.5 Ablation Protocol

To understand the role of each component, we conduct ablations on random 3-regular graphs with

$$\begin{aligned} n &\in \{500, 1000, 2000\}, \\ k &\in \{1, 2, 3, 4, 6, 8\}, \\ m &\in \{0, 2, 5, 10\}, \\ \eta &\in \{0.9, 0.5, 0.25, 0.1\}, \end{aligned} \quad (158)$$

using 20 independent graph trials.

The ablation study includes three parts: comparison of the four feature variants, sensitivity analysis over (k, m, η) under the full feature, and comparison of the three anchor selection strategies.

For details on datasets, baselines, and hyperparameter configurations, please refer to https://anonymous.4open.science/r/Graph_Node_Identifiability_via_Observation_Maps-5862.

Use of AI assistants. AI assistants were used to support code development, including debugging and boilerplate generation. All experimental results, analyses, and claims were produced and verified by the authors.

C Supplementary Analysis and Detailed Results

C.1 Detailed Results on NLP-Derived Graphs

Table 5 expands the macro Task Err in Table 2 into individual task labels under $(k, m, \eta) = (1, 8, 0.25)$. Hybrid is lowest for every CLUTRR reasoning label and every UD syntactic label.

Table 5: Per-label task ambiguity on NLP-derived graphs under $(k, m, \eta) = (1, 8, 0.25)$. Values are means over graph instances and random anchor draws. Lower is better.

Benchmark	Dataset	Task label	Distance	Spectral	Hybrid
CLUTRR	clean-task1	query role	0.135	0.220	0.042
CLUTRR	clean-task1	source distance	0.148	0.389	0.051
CLUTRR	clean-task1	target distance	0.148	0.391	0.053
CLUTRR	distractor-task3	query role	0.227	0.199	0.061
CLUTRR	distractor-task3	source distance	0.169	0.161	0.026
CLUTRR	distractor-task3	target distance	0.181	0.204	0.050
UD	English-EWT	DEPREL	0.549	0.164	0.133
UD	English-EWT	root depth	0.134	0.011	0.003
UD	English-EWT	UPOS	0.521	0.160	0.132
UD	Chinese-GSD	DEPREL	0.576	0.155	0.137
UD	Chinese-GSD	root depth	0.141	0.003	0.001
UD	Chinese-GSD	UPOS	0.519	0.153	0.137
UD	French-GSD	DEPREL	0.585	0.192	0.169
UD	French-GSD	root depth	0.153	0.005	0.001
UD	French-GSD	UPOS	0.558	0.190	0.168

C.2 Robustness under Stronger Anchor Budgets

Table 4 reports the stronger setting $(k, m, \eta) = (4, 8, 0.25)$. CLUTRR becomes nearly identifiable under Distance-only because the graphs are small. Hybrid remains non-degrading and continues to improve all UD treebanks.

C.3 Synthetic threshold diagnostics

Figure 4 reports the empirical anchor threshold k_{emp} . The distance-only baseline is stable with $k_{\text{emp}} = 6$ across all tested n and η . Adding spectral codes reduces the threshold, especially under finer quantization.

Table 6 reports the finest-quantization case $\eta = 0.1$. The full threshold, ρ_{emp} , and spectral-code tables are given in Tables 7, 8, and 9.

C.4 Complete Diagnostic Tables for the Observation Map

For each configuration (n, m, η) , we define k_{emp} as the smallest anchor number such that the mean error rate is at most 0.1. We also record the corresponding theory-guided ratio ρ_{emp} and the spectral

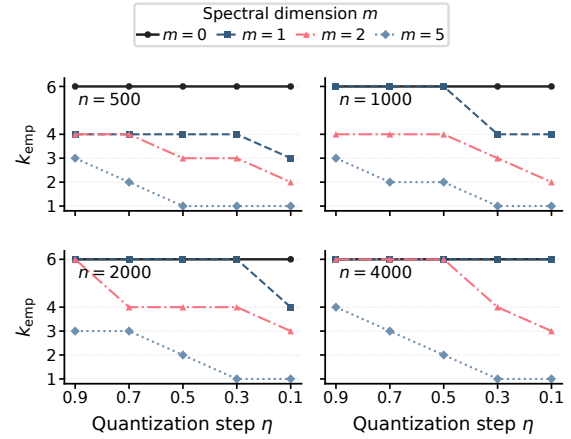


Figure 4: Empirical threshold k_{emp} versus quantization step η for random 3-regular graphs. Each panel fixes n , and each curve corresponds to a spectral dimension m .

Table 6: Threshold diagnostics at the finest quantization level $\eta = 0.1$.

n	m	k_{emp}	ρ_{emp}	$ \text{Im}(F) /n$	Avg. preimage	$ Q_\eta(\tilde{S}_m(V)) $
500	0	6	1.764	0.992	1.008	1
	1	3	1.364	0.906	1.105	51.75
	2	2	1.552	0.956	1.046	285.6
	5	1	2.704	1.000	1.000	499.45
1000	0	6	1.679	0.987	1.014	1
	1	4	1.553	0.963	1.038	66
	2	2	1.427	0.915	1.093	444.2
	5	1	2.448	0.999	1.001	997.8
2000	0	6	1.601	0.973	1.027	1
	1	4	1.462	0.937	1.067	76.3
	2	3	1.589	0.967	1.034	692.75
	5	1	2.237	0.999	1.001	1990.1
4000	0	6	1.530	0.948	1.055	1
	1	6	1.892	0.995	1.005	87.65
	2	3	1.488	0.940	1.064	1033.6
	5	1	2.061	0.998	1.002	3962

Note. Larger spectral codebooks are associated with smaller anchor thresholds. For $m = 5$, the codebook size is already close to n , and $k_{\text{emp}} = 1$ for all tested graph sizes.

Table 7: Full empirical threshold table k_{emp} .

n	m	$\eta = 0.9$	$\eta = 0.7$	$\eta = 0.5$	$\eta = 0.3$	$\eta = 0.1$
500	0	6	6	6	6	6
	1	4	4	4	4	3
	2	4	4	3	3	2
	5	3	2	1	1	1
1000	0	6	6	6	6	6
	1	6	6	6	4	4
	2	4	4	4	3	2
	5	3	2	2	1	1
2000	0	6	6	6	6	6
	1	6	6	6	6	4
	2	6	4	4	4	3
	5	3	3	2	1	1
4000	0	6	6	6	6	6
	1	6	6	6	6	6
	2	6	6	6	4	3
	5	4	3	2	1	1

Table 8: Theory-guided threshold ratio ρ_{emp} .

n	m	$\eta = 0.9$	$\eta = 0.7$	$\eta = 0.5$	$\eta = 0.3$	$\eta = 0.1$
500	0	1.764	1.764	1.764	1.764	1.764
	1	1.304	1.345	1.399	1.481	1.364
	2	1.433	1.514	1.328	1.492	1.552
	5	1.524	1.433	1.409	1.820	2.704
1000	0	1.679	1.679	1.679	1.679	1.679
	1	1.794	1.831	1.879	1.394	1.553
	2	1.350	1.423	1.520	1.389	1.427
	5	1.417	1.319	1.563	1.653	2.448
2000	0	1.601	1.601	1.601	1.601	1.601
	1	1.706	1.739	1.783	1.851	1.462
	2	1.811	1.344	1.432	1.567	1.589
	5	1.326	1.491	1.446	1.515	2.237
4000	0	1.530	1.530	1.530	1.530	1.530
	1	1.627	1.657	1.698	1.759	1.892
	2	1.723	1.784	1.865	1.478	1.488
	5	1.502	1.398	1.346	1.399	2.061

Table 9: Spectral code complexity $|Q_\eta(\tilde{S}_m(V))|$ at threshold.

n	m	$\eta = 0.9$	$\eta = 0.7$	$\eta = 0.5$	$\eta = 0.3$	$\eta = 0.1$
500	0	1.0	1.0	1.0	1.0	1.0
	1	9.1	11.7	15.7	23.8	51.8
	2	31.2	46.6	68.7	117.7	285.6
	5	213.2	290.6	395.6	469.8	499.5
1000	0	1.0	1.0	1.0	1.0	1.0
	1	10.5	14.4	17.5	27.1	66.0
	2	46.2	64.5	97.2	169.3	444.2
	5	345.2	491.6	693.9	900.9	997.8
2000	0	1.0	1.0	1.0	1.0	1.0
	1	12.4	16.0	22.0	32.5	76.3
	2	59.6	78.0	128.6	240.5	692.8
	5	561.4	820.3	1196.5	1677.7	1990.1
4000	0	1.0	1.0	1.0	1.0	1.0
	1	14.4	21.1	25.1	37.1	87.7
	2	74.2	109.8	163.6	315.2	1033.6
	5	898.6	1345.2	2091.1	3123.6	3962.0

code complexity $|Q_\eta(\tilde{S}_m(V))|$, where

$$F(v) = (d_A(v), Q_\eta(\tilde{S}_m(v))), \quad (159)$$

$$\tilde{S}_m(v) = n(\phi_2(v)^2, \dots, \phi_{m+1}(v)^2).$$

The main threshold patterns are simple: (i) the distance-only baseline is stable, with $k_{\text{emp}} = 6$ for all tested n and η ; (ii) increasing m and refining η generally reduce k_{emp} ; (iii) the reduction in k_{emp} is accompanied by a sharp growth of $|Q_\eta(\tilde{S}_m(V))|$.

C.5 External DDI graph case studies

Tables 12 and 10 give the DDI graph statistics and representative identifiability results. DrugBank remains highly colliding under all tested settings, whereas the smaller Decagon-derived graph becomes nearly injective when spectral information is added. These results are used as external graph case studies rather than as NLP-derived evidence.

D Additional ablation details

D.1 Quantization calibration

The main experiments use relative quantization. We additionally compare absolute quantization as a calibration check. Relative quantization is stable throughout the tested range, with mean error at most 6.7×10^{-5} and spectral-code ratio at least 0.9933. For absolute quantization, the result depends on the step size. As the absolute step decreases from 5×10^{-3} to 2×10^{-4} , the mean error drops from 0.1686 to 10^{-4} , while the occupied spectral-code ratio increases from 0.0593 to 0.9473. For $\eta \leq 10^{-4}$, absolute quantization also becomes essentially injective.

D.2 Bucketwise refinement diagnostics

We further examine spectral refinement inside non-singleton distance buckets. Figure 5 plots mean error against singleton-node fraction and mean within-bucket collision. In an intermediate regime such as $(n, k, m, \eta) = (2000, 4, 5, 0.25)$, the error is 0.1521, the singleton-node fraction is 0.3803, and the mean within-bucket collision is 0.2603. In a near-injective regime such as $(2000, 6, 10, 0.1)$, the error drops to 10^{-4} , the singleton-node fraction rises to 0.9422, and the mean within-bucket collision drops to 0.0039.

These diagnostics support the bucketwise interpretation: anchor distances form a coarse partition, and spectral codes reduce residual collisions inside non-singleton buckets.

Table 10: Representative identifiability results on the two real-world DDI graphs. Here $|\text{Im}(F)|/n$ is the normalized image size, and the remaining columns report preimage- and distance-bucket-based diagnostics.

Graph	k	m	η	Error	$ \text{Im}(F) /n$	Vertex-mean preimage	Singleton-preimage frac.	Singleton distance-bucket frac.	Median non-singleton collision	Median non-singleton balance
DrugBank DDI	1	0	0.5	0.9971	0.0029	959.65	0.0007	0.0007	1.0000	1.0000
DrugBank DDI	1	10	0.1	0.9918	0.0082	949.07	0.0037	0.0007	0.9637	3.4002
DrugBank DDI	4	0	0.1	0.9773	0.0227	264.77	0.0057	0.0057	1.0000	1.0000
DrugBank DDI	4	10	0.1	0.9711	0.0289	288.81	0.0104	0.0064	1.0000	1.0000
DrugBank DDI	8	0	0.1	0.9093	0.0907	92.46	0.0340	0.0340	1.0000	1.0000
DrugBank DDI	8	10	0.1	0.9023	0.0977	79.33	0.0384	0.0346	1.0000	1.0000
Decagon-derived DDI	1	0	0.5	0.9175	0.0825	24.14	0.0283	0.0283	1.0000	1.0000
Decagon-derived DDI	1	10	0.1	0.0233	0.9767	1.05	0.9533	0.0275	0.0015	1.3520
Decagon-derived DDI	4	0	0.1	0.6958	0.3042	9.47	0.1792	0.1792	1.0000	1.0000
Decagon-derived DDI	4	10	0.1	0.0200	0.9800	1.04	0.9600	0.1792	0.0000	1.0000
Decagon-derived DDI	8	0	0.1	0.4250	0.5750	3.83	0.4367	0.4367	1.0000	1.0000
Decagon-derived DDI	8	10	0.1	0.0158	0.9842	1.03	0.9683	0.4458	0.0000	1.0000

Table 11: Calibration of the absolute quantization step.

η	Mean error	Mean preimage	Spectral-code ratio
5×10^{-3}	0.1686	1.2464	0.0593
2×10^{-3}	0.0528	1.0604	0.2489
10^{-3}	0.0142	1.0149	0.4908
5×10^{-4}	0.0018	1.0018	0.7354
2×10^{-4}	0.0001	1.0001	0.9473
10^{-4}	0.000025	1.000025	0.9933
5×10^{-5}	0.0000	1.0000	0.9994
2×10^{-5}	0.0000	1.0000	1.0000

Table 12: Structural statistics of the two real-world DDI graphs.

Statistic	DrugBank DDI	Decagon-derived DDI
n	1684	40
$ E $	189774	466
Average degree	225.385	23.300
Density	0.1339	0.5974
Diameter	5	3
Avg. shortest-path length	2.086	1.414
Avg. clustering coefficient	0.550	0.874
Transitivity	0.496	0.801
Degree variance	38041.436	106.610
Degree Gini	0.478	0.246

D.3 Hyperparameter interaction and anchor robustness

Figure 3 reports the interaction among k , m , and η on random 3-regular graphs. Increasing the number of anchors, increasing the spectral dimension, and refining the quantization step all reduce ambiguity, but their effects are coupled. This is consistent with the bucketwise view: anchor distances provide the coarse partition, while spectral codes refine nodes that remain indistinguishable within distance buckets.

Random-anchor sampling is stable under repeated resampling. Under five random-anchor resamplings, the mean within-graph error range is 0.0848 for $n = 1000$ and 0.1189 for $n = 2000$, with maximum observed ranges 0.1870 and 0.2120, respectively. These fluctuations do not change the overall trends.

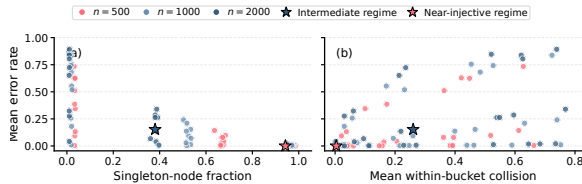


Figure 5: Bucketwise diagnostics on random 3-regular graphs. Left: mean error versus singleton-node fraction. Right: mean error versus mean within-bucket collision. Highlighted points show an intermediate regime and a near-injective regime.