JFT-Exp: Quantitative Joint Feature–Topology Expressivity for GNNs and Graph Transformers

Liz Lemma   Future Detective

January 18, 2026

Abstract

Existing assessments of graph neural network (GNN) expressivity largely quantify topology-only separation power (e.g., via the Weisfeiler–Lehman hierarchy, polynomial expressivity, or homomorphism-based metrics) and therefore miss a central empirical fact emphasized by recent surveys: node features can dramatically amplify topology representation, and the practical performance of GNNs/graph transformers depends on feature–topology interaction. We introduce JFT-Exp, a feature-aware, quantitative expressivity measure based on a joint feature–topology signature $\Phi_s(G)$ built from feature-decorated homomorphism polynomials and neighborhood feature-distribution moments. We prove a representation theorem showing that all permutation-invariant polynomials of degree $\leq s$ factor through Φ_s, and we derive a tight Lipschitz-type approximation bound: if a model recovers the signature within η, it recovers any Lipschitz task functional of the signature within $O(\eta)$. To make the theory actionable, we propose an efficiently computable sketch $\widehat{\Phi}_s(G)$ with near-linear complexity on sparse graphs and provide probabilistic error guarantees. We also establish strict refinement over WL-only/topology-only metrics and give lower bounds showing WL-equivalent message passing cannot approximate certain signature coordinates (e.g., cycle-derived invariants) on explicit families. Finally, we outline an evaluation protocol correlating JFT-Exp gaps with failure modes across MPNNs and graph transformers, enabling model selection and debugging before expensive training in 2026-scale settings.

Table of Contents

3. 3. Problem Formulation: define the target function class (degree-s invariant polynomials and Lipschitz functionals thereof); define what it means for a model to be expressive w.r.t. this class; define JFT-Exp order and score.

4. 4. The Joint Feature–Topology Signature Φ_s: full basis (decorated homomorphism polynomials up to size s); restricted/computable basis $\widehat{\Phi}_s$ (paths/cycles + feature moments); invariances and stability properties.

5. 5. Efficient Computation: near-linear sketching algorithms for $\widehat{\Phi}_s$ on sparse graphs (stochastic trace, sparse matvec pipelines, moment diffusion); concentration/error bounds and parameter choices (R, ε, δ).

6. 6. Representation Theorem (Upper Bound): all invariant degree-s polynomials factor through Φ_s; specialization to $\widehat{\Phi}_s$ for restricted function families; implications for universal approximation under symmetry constraints.

7. 7. Approximation and Generalization Bounds: Lipschitz transfer bound from signature error to task error; dataset-level diagnostic variants; discussion of when the bound is tight.

8. 8. Strict Refinement and Lower Bounds: explicit graph families where WL-equivalent MPNNs fail to approximate specific signature coordinates; matching constructions with architectures that succeed (global mixing / stable positional channels).

9. 9. Empirical Protocol (Optional but Strengthening): probing experiments predicting signature coordinates from embeddings; correlation with downstream accuracy; ablations across MPNNs, subgraph GNNs, and graph transformers; synthetic WL-hard benchmarks + real molecular/citation datasets.

10. 10. Discussion, Limitations, and Extensions: feature privacy/inductive settings, heterogeneous graphs, dynamic graphs; open questions on optimal bases and compute–expressivity frontiers.

1 Introduction and Motivation

We consider the problem of assigning to a featured graph a numerical object that can serve simultaneously as (a) an expressivity witness and (b) a computationally tractable target for learning. Our guiding constraint is permutation symmetry: any reasonable quantity extracted from a graph should depend only on its isomorphism class, and any model intended to process graphs should respect relabeling of vertices. Within this constraint, we would like a quantitative notion of "how much" joint information about topology and features a representation carries, in a sense that is compatible with both classical invariant theory and the algorithmic primitives available in modern graph learning (sparse aggregation, sparse attention, and randomized linear-algebraic estimators).

A first obstruction is that topology-only metrics, even when they are sensitive to global structure (e.g., spectral summaries or short-cycle statistics), do not address the feature–topology interaction that drives many learning tasks. If we fix a topology and vary node features, then tasks such as graph-level regression or classification may change substantially while any purely topological signature remains constant. Conversely, if we fix features (or restrict to constant features), then topological summaries can be informative, but a metric that ignores features cannot certify whether a model has exploited feature information correctly rather than merely fitting spurious topological correlates. Hence, a meaningful expressivity measure for *featured* graphs should incorporate feature values into the same algebraic object that records topology, rather than treating feature statistics as an auxiliary add-on. In particular, we should avoid a construction that consists only of (i) a topological invariant and (ii) independent moments of the feature multiset, since such a separation cannot encode whether feature values occur in specific structural locations (e.g., on a cycle versus on a tree-like neighborhood) or whether they align along multi-hop dependencies.

A second obstruction arises from the known limitations of message-passing neural networks (MPNNs) that are equivalent to the one-dimensional Weisfeiler–Leman (1-WL) refinement. The 1-WL procedure aggregates information from rooted neighborhoods and, by construction, cannot distinguish certain non-isomorphic graphs that share the same 1-WL color refinement sequence. Consequently, any expressivity metric derived solely from the invariants accessible to 1-WL-equivalent architectures will necessarily identify these graphs, even when they differ by global properties such as short cycle counts, walk traces, or other homomorphism-type statistics. This suggests that, if we wish the metric to be able to witness a strict refinement over WL-limited expressivity, then it must include coordinates that are provably outside the 1-WL closure, while still being computable without resorting to expensive enumeration over tuples or subgraphs.

These two obstructions are resolved simultaneously by adopting a joint

feature–topology signature whose coordinates are *feature-decorated pattern functionals*. The relevant phenomenon is that many graph quantities of interest can be expressed as sums over walks or homomorphisms, possibly weighted by feature monomials evaluated at the image vertices. For example, quadratic forms of the type

$$x^\top A^\ell x$$

are invariant under relabeling and encode a coupling between length-ℓ connectivity and feature alignment: they are large precisely when feature mass is connected through many length-ℓ walks. More generally, if we allow multiple feature channels and higher-order feature monomials, then the resulting quantities represent a controlled family of invariant polynomials that jointly record (i) how patterns occur in the graph and (ii) where particular feature configurations appear relative to those patterns. Such quantities are therefore sensitive to "feature placement" within the topology, which is precisely what decoupled topology-only plus feature-only summaries fail to capture.

We emphasize that our objective is not to compute an intractably large complete invariant (which is impossible in general under standard hardness assumptions), but rather to specify a signature family with three concrete properties. First, it should admit a *theoretical* full-basis form that is expressive enough to span all permutation-invariant polynomials up to a chosen degree/order parameter s. This provides an algebraic characterization: if a target functional is of bounded polynomial degree, then it factors through the full signature. Second, it should admit a *computable* sketched form that is compatible with sparse graph access. In practice, this means that the dominant operations must be sparse matrix–vector multiplication and simple global reductions, and that the use of randomness (e.g., Hutchinson-type trace probes or sketching transforms) is permitted to trade exactness for controlled additive error. Third, it should yield a *predictive* notion of approximation: if a learned representation allows one to reconstruct the signature within a small norm error, then any Lipschitz functional of the signature is automatically approximated with proportionally small task error. This moves the metric from a purely descriptive statistic to a tool for certifying downstream performance relative to a chosen function class.

In this framework, the order parameter s plays two roles. On the algebraic side, it bounds the complexity of patterns and the total degree of feature decorations we are willing to consider. On the algorithmic side, it bounds the maximum walk length or diffusion depth used to probe topology and to propagate features before taking global moments. Treating s as a tunable constant is essential: we obtain a monotone refinement property (increasing s adds information) while preserving near-linear computation in the number of edges for sparse graphs. Likewise, a separate sketch budget parameter R controls the variance of randomized estimators used for trace-like quantities; increasing R reduces estimator noise without changing the

underlying invariance structure.

We summarize the design goals as follows.

1. *Permutation invariance by construction.* Each coordinate should be an invariant under vertex relabeling, so that the signature is a well-defined function on isomorphism classes.

2. *Joint feature–topology coupling.* Coordinates should include feature-decorated functionals that cannot be written as a sum of a purely topological term and a purely feature-multiset term, thereby encoding where features occur in the graph.

3. *Strict refinement over 1-WL-limited expressivity.* The signature should contain coordinates (e.g., walk traces or cycle-related quantities, optionally feature-weighted) that separate known 1-WL-indistinguishable graph families.

4. *Efficient, sparse computation with guarantees.* A sketched version should be computable using a small number of sparse passes and randomized probes, with explicit additive error and failure probability bounds.

5. *Transfer to learning via reconstruction error.* If a model representation can be decoded to recover the signature within error η, then any Lipschitz downstream functional of the signature is approximated within $O(\eta)$.

These requirements enforce a particular structure: the signature must be expressive enough to witness separations, but also regular enough to admit stable approximation and efficient estimation.

Finally, we note that the present viewpoint is deliberately modular. The signature is not tied to any specific neural architecture; rather, it is an external object against which architectures may be compared. WL-limited MPNNs are expected to fail to reconstruct certain global coordinates on appropriate graph families, while architectures endowed with global mixing channels (for instance, via positional information, spectral augmentation, or sparse global attention) can, in principle, approximate these coordinates within small error using near-linear computation. This separation is not merely qualitative: by measuring reconstruction error in a chosen norm on the signature space, we obtain a quantitative expressivity score that can be aggregated over datasets and linked directly to task performance through Lipschitz transfer.

In the subsequent preliminaries we formalize the symmetry model, the WL/MPNN baseline, and the pattern-based functionals underlying the signature coordinates, and we fix the analytic setup (norms, Lipschitz function classes, and error notions) required for the approximation guarantees.

2 Preliminaries

Featured graphs and basic notation. We work with featured graphs $G = (A, X)$ on n vertices, where the adjacency matrix satisfies $A \in \{0, 1\}^{n \times n}$, $A = A^\top$, and $A_{ii} = 0$ for all $i \in [n]$, and the node-feature matrix satisfies $X \in [-1, 1]^{n \times d}$. We write $x_v \in [-1, 1]^d$ for the v-th row of X, and $X_{\cdot, j} \in [-1, 1]^n$ for the j-th feature channel. Throughout we regard $s \geq 2$ as a fixed order parameter that bounds pattern size and/or polynomial degree, while d is the feature dimension. When we discuss algorithmic access we assume adjacency-list input (equivalently, sparse matrix access via sparse matrix–vector multiplication), but the algebraic definitions below are stated in terms of (A, X).

Permutation action; invariance and equivariance. Let π be a permutation of $[n]$ and $P \in \{0, 1\}^{n \times n}$ its permutation matrix. The relabeled graph is
$$\pi \cdot (A, X) := (PAP^\top, PX).$$
A graph-level map F is *permutation invariant* if $F(\pi \cdot G) = F(G)$ for all π, and a node-level map $H(G) \in \mathbb{R}^{n \times k}$ is *permutation equivariant* if $H(\pi \cdot G) = P H(G)$. In particular, any graph-level statistic intended to be well defined on isomorphism classes must be invariant, whereas the intermediate representations of a node-processing architecture are naturally equivariant. We shall repeatedly use the fact that sums over vertices and traces of matrix expressions are invariant under simultaneous conjugation by permutations, e.g.,

$$\operatorname{tr}\!\left((PAP^\top)^\ell\right) = \operatorname{tr}(A^\ell), \qquad (Px)^\top (PAP^\top)^\ell (Px) = x^\top A^\ell x.$$

Message passing and the 1-WL baseline. A (feature-only, edge-unlabeled) message-passing neural network (MPNN) computes node representations $\{h_v^{(t)}\}_{v \in [n]}$ by iterating an update of the form

$$h_v^{(t+1)} = \phi_t\!\left(h_v^{(t)}, \square_{u \in N(v)} \psi_t\!\left(h_v^{(t)}, h_u^{(t)}\right)\right), \qquad h_v^{(0)} = x_v, \tag{1}$$

where $N(v)$ is the neighborhood of v, $\square$ is a permutation-invariant multiset aggregator (typically sum/mean/max), and ϕ_t, ψ_t are learnable functions. A graph embedding is then obtained by an invariant readout, e.g. $\mathrm{Readout}(\{h_v^{(T)}\}_v) = \rho(\sum_v h_v^{(T)})$. Such architectures are permutation equivariant at the node level and invariant at the graph level by construction.

The standard expressivity reference point is the one-dimensional Weisfeiler–Leman (1-WL) color refinement. Informally, 1-WL assigns initial colors from node features and iteratively refines them by hashing a node's current color together with the multiset of neighbor colors. It is a classical fact that

MPNNs of the form (1) with sufficiently expressive ϕ_t, ψ_t and injective multiset aggregation are *no more expressive* than 1-WL at distinguishing non-isomorphic graphs: if two graphs are 1-WL-indistinguishable (starting from the given features), then any such MPNN produces identical multisets of node representations at each depth, hence identical graph-level embeddings after invariant readout. We therefore regard 1-WL equivalence as the natural limitation class for "local" message passing. The separations we later use exploit that 1-WL can fail to detect global properties such as certain short-cycle statistics or other walk-based invariants, even on regular graphs with constant features.

Graph homomorphisms and homomorphism polynomials. Let $H = (V(H), E(H))$ be a (small) pattern graph with $|V(H)| \leq s$. A graph homomorphism $\varphi : H \to G$ is a map $\varphi : V(H) \to [n]$ such that $(a,b) \in E(H)$ implies $(\varphi(a), \varphi(b)) \in E(G)$. We denote the set of homomorphisms by $\mathrm{Hom}(H \to G)$. The homomorphism count can be written as the polynomial

$$\mathrm{hom}(H, G) := |\mathrm{Hom}(H \to G)| = \sum_{\varphi : V(H) \to [n]} \prod_{(a,b) \in E(H)} A_{\varphi(a)\varphi(b)}. \quad (2)$$

Expression (2) is invariant under relabeling since it is a sum over all assignments φ and depends on A only through edge indicators composed with φ. Many classical walk- and cycle-statistics are special cases: for the cycle C_ℓ one has

$$\mathrm{tr}(A^\ell) = \sum_{i_1,\ldots,i_\ell} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_\ell i_1} = \mathrm{hom}(C_\ell, G),$$

noting that homomorphisms allow repeated vertices and thus correspond to closed walks.

Feature-decorated homomorphism polynomials. To couple topology and features within a single invariant, we decorate homomorphism sums by monomials in feature coordinates evaluated at the image vertices. Concretely, let $\alpha = \{\alpha_{a,j}\}_{a \in V(H), j \in [d]}$ be a collection of nonnegative integers (a feature-monomial decoration). We define the feature-decorated homomorphism polynomial

$$p_{H,\alpha}(G) := \sum_{\varphi : V(H) \to [n]} \left(\prod_{(a,b) \in E(H)} A_{\varphi(a)\varphi(b)} \right) \left(\prod_{a \in V(H)} \prod_{j=1}^{d} X_{\varphi(a),j}^{\alpha_{a,j}} \right). \quad (3)$$

The total degree of $p_{H,\alpha}$ in the entries of (A, X) is $|E(H)| + \sum_{a,j} \alpha_{a,j}$. By construction $p_{H,\alpha}$ is invariant under $\pi \cdot (A, X)$: conjugating A and permuting rows of X merely reindexes the summation in (3). We regard the collection

of all $p_{H,\alpha}$ with $|V(H)| \leq s$ and total degree bounded by s as the canonical coordinates of the full signature $\Phi_s(G)$.

Two basic examples, which will also serve as computable coordinates in the sketched signature, are the quadratic forms

$$X_{\cdot,j}^\top A^\ell X_{\cdot,j} = \sum_{u,v} X_{u,j} (A^\ell)_{uv} X_{v,j},$$

and mixed-channel variants $X_{\cdot,j}^\top A^\ell X_{\cdot,k}$. These are invariant and encode feature alignment along length-ℓ connectivity. Likewise, global feature moments of diffused features, such as

$$\mu_{k,r}(G) := \frac{1}{n} \sum_{v=1}^n \left((P^k X)_{v,\cdot} \right)^{\odot r},$$

for a chosen diffusion operator P (e.g. normalized adjacency), provide permutation-invariant summaries of k-hop feature distributions; we will treat these as additional low-cost coordinates capturing feature statistics after controlled propagation.

Norms, Lipschitz functionals, and reconstruction error. We consider downstream targets that are Lipschitz functions of a chosen signature vector. Fix a norm $\|\cdot\|$ on the relevant signature space (typically ℓ_2 or coordinatewise ℓ_∞ on $\mathbb{R}^p$). A function $g : \mathbb{R}^p \to \mathbb{R}$ is L-Lipschitz if

$$|g(u) - g(v)| \leq L\|u - v\| \qquad \text{for all } u, v \in \mathbb{R}^p.$$

Given a model representation $F(G)$ and a fixed decoder ψ intended to reconstruct $\widehat{\Phi}_s(G)$, we measure *signature reconstruction error* on a graph class $\mathcal{C} \subseteq \mathcal{G}_{n,d}$ by

$$\eta := \sup_{G \in \mathcal{C}} \left\| \widehat{\Phi}_s(G) - \psi(F(G)) \right\|.$$

This choice is tailored to an immediate transfer principle: once η is small, every L-Lipschitz functional of the signature is automatically approximated to error at most $L\eta$. We will exploit this observation in the problem formulation by defining expressivity relative to the class of targets that factor through $\widehat{\Phi}_s$ (or through the full Φ_s at bounded degree), and by using η as a quantitative surrogate for task error.

3 Problem formulation

Target function class at order s. Fix an order parameter $s \geq 2$ and feature dimension d. We are interested in graph-level targets $f : \mathcal{G}_{n,d} \to \mathbb{R}$ that are permutation invariant and whose dependence on the input (A, X) is controlled at "resolution" s. We formalize this in two nested ways.

First, we consider the algebraic class of bounded-degree invariants,

$$\mathcal{F}_s \; := \; \mathbb{R}[A, X]_{\leq s}^{S_n},$$

the set of S_n-invariant polynomials in the entries of (A, X) of total degree at most s. This class captures precisely the invariants accessible to degree-s polynomial functionals and will serve as our canonical "expressivity frontier" at order s. In particular, we regard two graphs as indistinguishable at order s if all functions in $\mathcal{F}_s$ agree on them.

Second, since many practical tasks are not polynomial but are well approximated by smooth or Lipschitz maps of suitable summary statistics, we also consider a Lipschitz envelope of the signature. Let $\widehat{\Phi}_s(G) \in \mathbb{R}^p$ denote a chosen computable signature vector at order s (potentially randomized through sketching). For $L \geq 0$ we define the induced target class

$$\mathcal{T}_{s,L} \; := \; \big\{\, g \circ \widehat{\Phi}_s \; : \; g : \mathbb{R}^p \to \mathbb{R} \text{ is } L\text{-Lipschitz under } \|\cdot\| \,\big\}.$$

The point of $\mathcal{T}_{s,L}$ is not that the task must literally be given as $g \circ \widehat{\Phi}_s$, but that, once a representation can stably reconstruct $\widehat{\Phi}_s$, it can approximate *all* such tasks uniformly, with constants depending only on L and the reconstruction error.

Expressivity as signature recoverability. Let F be a graph model producing a representation $F(G)$ (either graph-level or node-level, with an implicit invariant pooling). We evaluate expressivity relative to $\widehat{\Phi}_s$ through the existence of a fixed decoder (readout) ψ mapping model outputs to the signature space. Concretely, given a graph class $\mathcal{C} \subseteq \mathcal{G}_{n,d}$ we define the order-s reconstruction error of F by

$$\eta_s(F; \mathcal{C}) \; := \; \inf_{\psi} \; \sup_{G \in \mathcal{C}} \big\| \widehat{\Phi}_s(G) - \psi(F(G)) \big\|. \tag{4}$$

We interpret $\eta_s(F; \mathcal{C})$ as a quantitative proxy for expressivity: small η_s means that F retains (up to the decoder) essentially all information contained in $\widehat{\Phi}_s$, whereas a constant lower bound on η_s certifies that F provably discards some order-s joint feature–topology information.

This definition is tailored to the transfer principle stated later (cf. Thm. 3): for every L-Lipschitz g and every $G \in \mathcal{C}$ one has

$$\big| g(\widehat{\Phi}_s(G)) - g(\psi(F(G))) \big| \; \leq \; L \, \big\| \widehat{\Phi}_s(G) - \psi(F(G)) \big\| \; \leq \; L \, \eta_s(F; \mathcal{C}). \tag{5}$$

Thus, controlling η_s uniformly yields simultaneous approximation guarantees for the entire class $\mathcal{T}_{s,L}$.

From the full signature Φ_s to a computable sketch. At the level of invariant polynomials, the "ideal" coordinate system is the full signature $\Phi_s(G)$ consisting of all feature-decorated homomorphism polynomials of total degree at most s. The representation theorem (Thm. 1) can then be read as the statement that every $f \in \mathcal{F}_s$ factors as

$$f(G) = \tilde{g}\big(\Phi_s(G)\big) \qquad \text{for some linear functional } \tilde{g},$$

and hence Φ_s is universal for $\mathcal{F}_s$ at fixed s. In contrast, exact computation of the full basis is intractable in general, motivating the sketched signature $\widehat{\Phi}_s(G)$, which retains a structured subset of coordinates together with randomized estimators (e.g. trace probes) to achieve near-linear time. Since $\widehat{\Phi}_s$ is a function of the sketch budget R and of the chosen coordinate family, we implicitly treat η_s as depending on these algorithmic choices; increasing R and/or enriching the coordinate set can only improve the approximation quality of $\widehat{\Phi}_s$ to Φ_s and reduce estimator variance.

JFT-Exp order. We summarize a model's joint feature–topology expressivity by asking: up to what order s can the model stably reconstruct the corresponding signature? Fix a tolerance level $\tau > 0$ and a graph class $\mathcal{C}$. We define the *JFT-Exp order* of F at tolerance τ by

$$\text{ord}_{\text{JFT}}(F; \tau, \mathcal{C}) := \max\big\{ s \in \mathbb{N} : \eta_s(F; \mathcal{C}) \leq \tau \big\}, \qquad (6)$$

with the convention that the maximum is taken over the range of s under consideration (e.g. $2 \leq s \leq s_{\max}$ in experiments). This definition makes the monotonicity requirement explicit: as s increases, $\widehat{\Phi}_s$ contains strictly richer joint statistics, so reconstructing it is at least as hard. Consequently, ord_{JFT} captures the highest signature order for which the model behaves as an approximately sufficient statistic for the induced target class $\mathcal{T}_{s,L}$.

JFT-Exp score (continuous). For comparisons at a fixed s we also use a scalar *JFT-Exp score* derived from η_s. Any monotone transformation is acceptable; we adopt a bounded score in $[0, 1]$,

$$\text{score}_{\text{JFT}}(F; s, \mathcal{C}) := \frac{1}{1 + \eta_s(F; \mathcal{C})}. \qquad (7)$$

A score near 1 indicates near-perfect recoverability of $\widehat{\Phi}_s$ from $F(G)$ on $\mathcal{C}$, while scores bounded away from 1 certify an intrinsic information gap. When $\widehat{\Phi}_s$ is randomized (through sketches), we may replace the supremum in (4) by a high-probability or in-expectation counterpart; for instance, one may consider $\eta_s^{(\delta)}$ defined by requiring the inequality to hold with probability at least $1 - \delta$ over the signature randomness, uniformly over $G \in \mathcal{C}$.

Empirical instantiation over datasets. In practice $\mathcal{C}$ is represented by a finite dataset $\mathcal{D} = \{G_i\}_{i=1}^N$. We then replace the uniform reconstruction error by an empirical risk, e.g.

$$\widehat{\eta}_s(F; \mathcal{D}) := \inf_\psi \left(\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\Phi}_s(G_i) - \psi(F(G_i)) \right\|^2 \right)^{1/2},$$

and define the empirical score analogously. The conceptual content of the framework remains the same: we compare models by the degree to which their representations preserve the joint feature–topology statistics encoded by $\widehat{\Phi}_s$, and we translate this preservation into worst-case (or average-case) guarantees for all Lipschitz functionals of the signature via (5). This reduces expressivity evaluation to a concrete reconstruction problem whose difficulty can be calibrated by s (and algorithmically by R), and whose separations can be certified by explicit indistinguishable graph families for WL-limited architectures.

4 The joint feature–topology signature Φ_s

We now define the order-s joint feature–topology signature. Conceptually, $\Phi_s(G)$ is a coordinate system for permutation-invariant information in (A, X) up to "resolution" s, where s controls both the maximum pattern size and the total polynomial degree. The signature has a full (ideal) version Φ_s, which is universal for $\mathcal{F}_s = \mathbb{R}[A, X]_{\leq s}^{S_n}$, and a restricted (computable) version $\widehat{\Phi}_s$ obtained by selecting structured coordinates admitting near-linear sketching.

Decorated homomorphism polynomials (full basis). Let $H = (V(H), E(H))$ be a finite simple pattern graph with $k := |V(H)| \leq s$. A (graph) homomorphism $\varphi \in \mathrm{Hom}(H \to G)$ is a map $\varphi : V(H) \to [n]$ such that $(u, v) \in E(H)$ implies $A_{\varphi(u)\varphi(v)} = 1$. To incorporate node features, we consider monomial decorations on the vertices of H. Concretely, for each $u \in V(H)$ let $\alpha(u) \in \mathbb{N}^d$ be a multi-index with $|\alpha(u)|_1 = \sum_{j=1}^d \alpha(u)_j \geq 0$, and define the feature monomial at a node $v \in [n]$ by

$$X_v^{\alpha(u)} := \prod_{j=1}^d X_{vj}^{\alpha(u)_j}.$$

We then define the *feature-decorated homomorphism polynomial*

$$p_{H,\alpha}(G) := \sum_{\varphi:V(H)\to[n]} \left(\prod_{(u,v)\in E(H)} A_{\varphi(u)\varphi(v)} \right) \left(\prod_{u\in V(H)} X_{\varphi(u)}^{\alpha(u)} \right), \quad (8)$$

where the sum is over all maps φ (not necessarily injective). The total degree of $p_{H,\alpha}$ is $|E(H)| + \sum_{u \in V(H)} |\alpha(u)|_1$. In the full signature we include all such coordinates with total degree at most s:

$$\Phi_s(G) := \big(p_{H,\alpha}(G)\big)_{\substack{|V(H)| \leq s \\ |E(H)| + \sum_u |\alpha(u)|_1 \leq s}} \in \mathbb{R}^{p(s,d)}.$$

The choice of homomorphisms (rather than induced/injective embeddings) is deliberate: (8) is a polynomial in the entries of (A, X), and the family $\{p_{H,\alpha}\}$ is stable under algebraic operations, which is the correct level of generality for $\mathcal{F}_s$.

Permutation invariance and normalization. Each coordinate $p_{H,\alpha}$ is invariant under relabeling of G. Indeed, for any permutation matrix P and permuted graph $\pi \cdot (A, X) = (PAP^\top, PX)$, changing variables $\varphi \mapsto \pi^{-1} \circ \varphi$ yields

$$p_{H,\alpha}(\pi \cdot G) = p_{H,\alpha}(G).$$

Thus $\Phi_s(\pi \cdot G) = \Phi_s(G)$, and the same holds for any sub-collection of these coordinates. When comparing graphs across varying sizes, we may optionally apply simple normalizations (e.g., division by $n^{|V(H)|}$, or by appropriate powers of average degree) to control scaling; however, the core invariance statements are independent of such choices.

Stability and boundedness at fixed order. Since $A_{ij} \in \{0, 1\}$ and $X_{vj} \in [-1, 1]$, each monomial term in (8) is bounded in magnitude by 1, hence

$$|p_{H,\alpha}(G)| \leq \sum_{\varphi: V(H) \to [n]} 1 = n^{|V(H)|} \leq n^s.$$

In particular, at fixed s each coordinate is uniformly bounded by a polynomial in n. Moreover, $p_{H,\alpha}$ is multilinear in adjacency entries along the edges of H and polynomial in features with degree $\sum_u |\alpha(u)|_1$, implying a Lipschitz-type stability with respect to entrywise perturbations when s is fixed. This boundedness is sufficient for the transfer principle in (5): once a representation can approximately reconstruct $\widehat{\Phi}_s$, any Lipschitz functional of it is automatically stable.

Structured subfamilies: cycles, walks, and feature decoration. While the full basis is the correct universal object for $\mathcal{F}_s$, it is neither necessary nor feasible to compute all coordinates exactly. We therefore emphasize structured subfamilies that already capture salient interactions between topology and features.

For topology-only information, the cycle homomorphism counts correspond to traces of powers:

$$\mathrm{tr}(A^\ell) \;=\; \sum_{v_1,\ldots,v_\ell} A_{v_1 v_2} A_{v_2 v_3} \cdots A_{v_\ell v_1} \;=\; |\mathrm{Hom}(C_\ell \to G)|, \qquad 1 \le \ell \le s,$$

where C_ℓ is the ℓ-cycle. For joint feature–topology information, quadratic forms of the type $x^\top A^\ell x$ act as feature-decorated walk energies. Writing $x_j := X[:, j] \in \mathbb{R}^n$ for a feature channel,

$$x_j^\top A^\ell x_{j'} \;=\; \sum_{u,v} X_{uj}(A^\ell)_{uv} X_{vj'}$$

aggregates length-ℓ walks with endpoint decorations given by feature channels j, j'. These quantities can be viewed as specific linear combinations of the polynomials $p_{H,\alpha}$ with H a path (or a cycle with a marked pair of vertices) and α placing degree-one feature monomials on selected vertices.

Diffusion moments as distributional summaries. To retain information about the distribution of features over neighborhoods without enumerating patterns, we also include moment summaries of diffused features. Let P be a fixed linear diffusion operator derived from A (e.g., $P = D^{-1/2}AD^{-1/2}$ or $P = D^{-1}A$ where D is the degree matrix). For $k \le s$ we form $M_k := P^k X \in \mathbb{R}^{n \times d}$ and aggregate coordinate-wise moments across nodes, for moment order $1 \le r \le s$, via

$$\mu_{k,r} \;:=\; \frac{1}{n} \sum_{v=1}^{n} \big(M_k[v,:]\big)^{\odot r} \in \mathbb{R}^d,$$

where $(\cdot)^{\odot r}$ denotes elementwise powers. These statistics are permutation invariant (they are sums over nodes), and they couple topology and features through the repeated application of P. Although they are not, in general, a complete basis for $\mathcal{F}_s$, they provide a controlled and interpretable family of coordinates that is sensitive to both feature geometry and multi-hop connectivity.

Restricted/computable signature and monotonic refinement. We define the computable signature $\widehat{\Phi}_s(G)$ by selecting a finite list of such structured coordinates up to order s (e.g., $\{\mathrm{tr}(A^\ell)\}_{\ell \le s}$, a prescribed set of feature-decorated energies $\{x_j^\top A^\ell x_{j'}\}$, and diffusion moments $\{\mu_{k,r}\}_{k,r \le s}$), optionally followed by a linear sketching/compression map to a fixed width. By construction, $\widehat{\Phi}_s$ inherits permutation invariance from its coordinates. Furthermore, the family is *monotone in s*: increasing s only adds new coordinates and/or increases the maximal walk length and moment order, thereby refining the signature and making it strictly more informative. This monotonic

refinement is the structural reason that the reconstruction error $\eta_s(F; \mathcal{C})$ is a meaningful expressivity proxy as s varies: higher order forces the representation to preserve progressively more global joint statistics.

Finally, we stress the separation of roles: Φ_s is the algebraically universal object that characterizes $\mathcal{F}_s$, whereas $\widehat{\Phi}_s$ is the algorithmically accessible proxy used in practice. In the next section we specify sketching procedures for computing $\widehat{\Phi}_s$ on sparse graphs, together with probabilistic error guarantees.

Efficient computation via sparse linear algebra and sketching. We now specify how the restricted signature $\widehat{\Phi}_s(G)$ can be computed in near-linear time on sparse graphs. The guiding principle is that every chosen coordinate is either (i) a trace of a matrix polynomial in A (capturing cycle/walk homomorphisms), (ii) a quadratic form $u^\top q(A)v$ with u, v derived from feature channels (capturing feature-decorated walks), or (iii) a low-order moment of diffused features $P^k X$ (capturing distributional summaries over k-hop neighborhoods). Each of these can be realized using $O(1)$ working vectors and a small number of sparse matrix–vector products (SpMVs), each costing $O(m)$ time under adjacency-list access.

Stochastic trace estimation for cycle/walk traces. For $\ell \leq s$, we include coordinates of the form $\tau_\ell := \operatorname{tr}(A^\ell)$ (or analogously $\operatorname{tr}(P^\ell)$ for a normalized diffusion). Since A^ℓ is dense even when A is sparse, we do not form it explicitly. Instead we use Hutchinson's estimator: for i.i.d. Rademacher probes $z_1, \ldots, z_R \in \{\pm 1\}^n$,

$$\widehat{\tau}_\ell \; := \; \frac{1}{R} \sum_{r=1}^{R} z_r^\top A^\ell z_r, \qquad \mathbb{E}[\widehat{\tau}_\ell] = \operatorname{tr}(A^\ell). \tag{9}$$

Each term $z_r^\top A^\ell z_r$ is computed by the pipeline $v^{(0)} := z_r$, $v^{(t)} := Av^{(t-1)}$ for $t = 1, \ldots, \ell$, followed by the inner product $z_r^\top v^{(\ell)}$. This costs ℓ SpMVs per probe, hence total time $O(Rm\ell)$ for a fixed ℓ and $O(Rms^2)$ if naively repeated for all $\ell \leq s$. In practice we reuse intermediate powers for each probe (keeping $v^{(t)}$ as we increment t), so that computing all $\{\widehat{\tau}_\ell\}_{\ell \leq s}$ costs $O(Rms)$ time for the SpMVs plus $O(Rns)$ for inner products, dominated by $O(Rms)$ in the sparse regime.

Concentration bounds for (9) are standard. For symmetric M and Rademacher probes, one has

$$\Pr\big(\big| \widehat{\operatorname{tr}}(M) - \operatorname{tr}(M) \big| \geq \varepsilon \|M\|_F \big) \; \leq \; 2\exp(-cR\varepsilon^2), \tag{10}$$

for a universal constant $c > 0$. Applying (10) with $M = A^\ell$ and a union

bound over $\ell = 1, \ldots, s$ yields the parameter choice

$$R = \Theta\left(\varepsilon^{-2} \log(s/\delta)\right) \qquad \implies \qquad \max_{\ell \leq s} |\widehat{\tau}_\ell - \tau_\ell| \leq \varepsilon \max_{\ell \leq s} \|A^\ell\|_F \text{ w.p. } \geq 1 - \delta. \tag{11}$$

If one replaces Hutchinson by Hutch++ (or related variance-reduced estimators), the same target accuracy can be achieved with fewer probes when A^ℓ is approximately low-rank; the computational interface remains SpMV-based, and our use of R may be interpreted as a generic sketch budget.

Feature-decorated walk energies as SpMV pipelines. For feature channels $x_j := X[:,j] \in \mathbb{R}^n$, we include (a selected subset of) coordinates of the form

$$e_{\ell,j,j'} := x_j^\top A^\ell x_{j'}. \tag{12}$$

These are computed deterministically by repeated SpMV: set $y^{(0)} := x_{j'}$, iterate $y^{(t)} := A y^{(t-1)}$ for $t = 1, \ldots, \ell$, and output $x_j^\top y^{(\ell)}$. Computing all $\{e_{\ell,j,j}\}_{\ell \leq s, j \leq d}$ by this method costs $O(msd)$ time and $O(nd)$ space to store X (or streaming access if X is stored externally). If cross-channel terms $j \neq j'$ are desired, we either restrict to a prescribed set of pairs (application-dependent) or compress the feature dimension first: choose a sketching matrix $S \in \mathbb{R}^{d \times d'}$ (e.g. CountSketch or a random sign projection with $d' \ll d$) and replace X by $\widetilde{X} := XS$. Then energies computed on $\widetilde{X}$ summarize many original channels while keeping the same SpMV cost with d' in place of d. Since the map $X \mapsto XS$ is linear and applied identically at every node, permutation invariance is unaffected.

We record the basic invariance for completeness: for any permutation matrix P,

$$x_j^\top A^\ell x_{j'} = (Px_j)^\top (PAP^\top)^\ell (Px_{j'}),$$

so (12) is a valid coordinate of a graph signature.

Diffusion moments and bounded feature arithmetic. Let P be a fixed diffusion operator derived from A (e.g. $D^{-1}A$ or $D^{-1/2}AD^{-1/2}$). For $k \leq s$ we compute $M_k := P^k X$ via the recurrence $M_0 := X$ and $M_k := PM_{k-1}$. This uses s SpMVs per feature channel in the worst case, but in matrix form it is simply s sparse matrix–dense matrix multiplies (SpMMs), realizable as s passes over edges with $O(md)$ work per pass, hence total $O(msd)$ time. From M_k we compute moment coordinates

$$\mu_{k,r} := \frac{1}{n} \sum_{v=1}^n \left(M_k[v,:]\right)^{\odot r}, \qquad 1 \leq k \leq s, \ \ 1 \leq r \leq s.$$

These require only entrywise powers and summations. Under our standing boundedness assumption $X \in [-1,1]^{n \times d}$ and for standard choices of P with

$\|P\|_{\infty \to \infty} \leq 1$, the iterates satisfy $\|M_k\|_\infty \leq 1$, so the moments are numerically stable at fixed s and do not require additional randomization. If one wishes to reduce computation further, one may subsample moment orders r, subsample feature coordinates, or compute moments after feature sketching $X \mapsto XS$ as above.

Compression of the assembled signature. After computing the chosen coordinates, we may apply a linear sketch C (e.g. CountSketch) to obtain a fixed-width vector $C\widehat{\Phi}_s(G) \in \mathbb{R}^{p'}$. Since C is linear, it can be applied in streaming form as coordinates are generated, requiring $O(p')$ additional memory. For norms such as ℓ_2, standard guarantees imply that with $p' = \Theta(\varepsilon^{-2}\log(1/\delta))$, pairwise distances between signatures are preserved up to $(1\pm\varepsilon)$ for a fixed set of graphs (or up to additive ε in an appropriate scaling), which is sufficient when $\widehat{\Phi}_s$ is used as an expressivity proxy or as an input to a Lipschitz downstream functional.

Complexity and parameter selection. Collecting the preceding components, we obtain the following operational bound: if $\widehat{\Phi}_s$ consists of (a) all $\ell \leq s$ trace coordinates estimated with R probes, (b) a prescribed set of $O(d)$ feature energies per $\ell \leq s$, and (c) diffusion moments up to order s, then the total work is

$$\tilde{O}(Rms) \quad \text{(trace coordinates)} \qquad + \qquad \tilde{O}(msd) \quad \text{(energies and diffusion)}$$

with space $\tilde{O}(Rn+nd)$ (storing probes and a constant number of working vectors/matrices). Choosing R as in (11) yields coordinate-wise additive error $\varepsilon\|A^\ell\|_F$ for all trace-like coordinates simultaneously with probability at least $1-\delta$, while the remaining coordinates are computed exactly given the chosen feature sketching/compression. This completes the efficient-computability component needed for the subsequent representation theorem: we can approximate a nontrivial, monotone family of invariant joint statistics using only sparse linear-algebra primitives and controlled randomness.

Representation theorem: invariant degree-s polynomials factor through Φ_s. We now justify the claim that the *full* joint feature–topology signature $\Phi_s(G)$ is an algebraic sufficient statistic for permutation-invariant polynomials of bounded degree. Fix n and d, and consider the action of the symmetric group S_n on node indices, acting on (A, X) by simultaneous relabeling $A \mapsto PAP^\top$ and $X \mapsto PX$ for the permutation matrix P. Let $\mathbb{R}[A, X]_{\leq s}$ denote the real polynomials in the entries of A and X of total degree at most s, and let $\mathbb{R}[A, X]_{\leq s}^{S_n}$ be the subspace of S_n-invariant polynomials.

Theorem 4.1 (Representation by decorated homomorphism polynomials). *For every $f \in \mathbb{R}[A, X]_{\leq s}^{S_n}$ there exists a finite collection of feature-decorated*

pattern graphs (H, α) with $|V(H)| \leq s$ and coefficients $\{c_{H,\alpha}\}$ such that, for all $G = (A, X) \in \mathcal{G}_{n,d}$,

$$f(G) \;=\; \sum_{(H,\alpha)} c_{H,\alpha}\, p_{H,\alpha}(G),$$

where each $p_{H,\alpha}(G)$ is a feature-decorated homomorphism polynomial and appears as a coordinate of the full signature $\Phi_s(G)$. Equivalently, there exists a linear map ℓ (depending on f) such that $f(G) = \ell(\Phi_s(G))$ for all G.

Proof sketch. We proceed by reducing to orbit sums of monomials. Any $f \in \mathbb{R}[A, X]_{\leq s}$ can be written as a finite linear combination of monomials of the form

$$m(A, X) \;=\; \prod_{t=1}^{T} A_{u_t v_t} \,\cdot\, \prod_{q=1}^{Q} X_{w_q, j_q}, \qquad T + Q \leq s, \tag{13}$$

where (u_t, v_t) and w_q are node indices in $[n]$ and $j_q \in [d]$ are feature coordinates. Under the S_n action, monomials are permuted by relabeling all node indices simultaneously. Since f is invariant, it lies in the span of *orbit sums* of monomials: for each monomial m, define its orbit sum

$$\mathrm{Orb}(m)(A, X) \;:=\; \sum_{\pi \in S_n} m(\pi \cdot (A, X)),$$

or, equivalently, sum over distinct relabelings of the indices appearing in m (dividing by stabilizers only changes the coefficient and is immaterial for spanning). Standard invariant-theoretic arguments for finite group actions imply that the orbit sums of monomials span $\mathbb{R}[A, X]_{\leq s}^{S_n}$.

It remains to identify each orbit sum with a decorated homomorphism polynomial. Given a monomial as in (13), consider the finite set of node variables actually used by the monomial,

$$U \;:=\; \{u_1, v_1, \ldots, u_T, v_T, w_1, \ldots, w_Q\},$$

and let $k := |U| \leq T + Q \leq s$. We form a *pattern graph* H whose vertex set is an abstract copy of U (with repeated indices identified), and whose edge multiset records the adjacency factors: each factor $A_{u_t v_t}$ contributes an edge between the abstract vertices corresponding to u_t and v_t. (If the same pair appears multiple times, this produces parallel edges; algebraically this simply represents multiplicity in the monomial. One may also eliminate parallel edges by allowing edge labels/multiplicities, which does not change the conclusion.) Likewise, the feature factors $\prod_q X_{w_q, j_q}$ define a decoration α that assigns to each pattern vertex $a \in V(H)$ a feature monomial in the coordinates of X at the image of a, namely the product of those $X_{\cdot, j}$ terms whose node index equals a.

With this construction, summing m over relabelings is the same as summing over all node assignments $\varphi : V(H) \to [n]$ of the product of corresponding adjacency and feature terms, i.e.,

$$\text{Orb}(m)(A, X) \;=\; \sum_{\varphi : V(H) \to [n]} \; \prod_{(a,b) \in E(H)} A_{\varphi(a)\varphi(b)} \cdot \prod_{a \in V(H)} \alpha_a\big(X_{\varphi(a),:}\big) \;=:\; p_{H,\alpha}(G).$$

The right-hand side is exactly a feature-decorated homomorphism polynomial: it is a homomorphism sum from the pattern H into G, with vertex weights given by feature monomials. Since $k = |V(H)| \le s$, this coordinate is included in the full signature $\Phi_s(G)$ by definition. Linear combinations over orbit sums therefore yield the desired representation of f as a linear functional of $\Phi_s(G)$, proving Theorem 4.1.

Specialization to restricted signatures. The preceding theorem is exact but uses the full basis of patterns up to size s, which is not intended for direct computation. For the *restricted* signature $\widehat{\Phi}_s(G)$, we obtain an immediate, though correspondingly restricted, factorization statement: any function that depends on G only through the chosen coordinates of $\widehat{\Phi}_s$ trivially factors through $\widehat{\Phi}_s$. Formally, if we write $\widehat{\Phi}_s(G) = (q_1(G), \ldots, q_p(G))$ for the selected invariant coordinates (cycle/walk traces, feature-decorated quadratic forms, diffusion moments, and possibly their linear sketches), then for any polynomial $h \in \mathbb{R}[t_1, \ldots, t_p]$ the composite

$$G \;\longmapsto\; h\big(\widehat{\Phi}_s(G)\big)$$

is a permutation-invariant polynomial in (A, X) whose algebraic complexity is controlled by the degrees of the individual coordinates q_i. In particular, for families of targets known *a priori* to be functions of counts of short cycles, low-order walk energies, and low-order diffusion moments, the restricted signature is already sufficient: such targets admit exact representations of the form $g \circ \widehat{\Phi}_s$ for an appropriate g.

Implications for universality under symmetry constraints. Theorem 4.1 provides an "upper bound" on expressive power: if an architecture can (approximately) recover $\Phi_s(G)$ and then apply a sufficiently rich readout, it can realize any invariant polynomial of degree at most s. When n is fixed, this algebraic statement can be combined with polynomial approximation to obtain a standard universality consequence for continuous invariant functions. Indeed, the domain

$$\mathcal{D}_{n,d} \;:=\; \{(A, X) : A \in \{0,1\}^{n \times n} \text{ symmetric, } \text{diag} = 0, \; X \in [-1,1]^{n \times d}\}$$

is compact (as a finite union of compact sets), and any continuous invariant function $F : \mathcal{D}_{n,d} \to \mathbb{R}$ can be uniformly approximated by polynomials in

the entries of (A, X) (e.g. by Stone–Weierstrass on a surrounding hypercube, then restricting to $\mathcal{D}_{n,d}$). Averaging the approximating polynomials over S_n preserves uniform approximation and yields invariant polynomials. Consequently, for every $\epsilon > 0$ there exists an invariant polynomial $f_\epsilon \in \mathbb{R}[A, X]^{S_n}$ such that $\sup_{G \in \mathcal{D}_{n,d}} |F(G) - f_\epsilon(G)| \leq \epsilon$. Truncating to some finite degree s (depending on ϵ and F) and applying Theorem 4.1 shows that f_ϵ factors through Φ_s; hence, up to ϵ, F factors through Φ_s as well.

This is precisely the form of universality relevant to symmetry-constrained learning: the obstacle is not the readout class (one may take a generic multilayer perceptron on the signature coordinates), but rather the ability of the representation mechanism to capture the requisite invariant coordinates. In the sequel, we therefore treat $\widehat{\Phi}_s$ as a computationally accessible proxy for Φ_s, and we measure a model F by its signature reconstruction error $\|\widehat{\Phi}_s(G) - \psi(F(G))\|$. The next section turns this reconstruction viewpoint into task-level approximation and generalization bounds for any Lipschitz downstream functional.

Approximation bounds via signature reconstruction. We now make explicit the basic transfer principle implicit in our reconstruction viewpoint: whenever a model can recover the restricted signature $\widehat{\Phi}_s(G)$ up to small error, it can approximate *any* sufficiently regular downstream functional of that signature with a commensurate error. Fix a norm $\| \cdot \|$ on the signature space and let $g : \mathbb{R}^p \to \mathbb{R}$ be L-Lipschitz with respect to this norm. Consider any representation map F (e.g. an MPNN, a graph transformer, or a hybrid architecture) and a fixed decoder ψ mapping $F(G)$ into $\mathbb{R}^p$. If we have the uniform reconstruction guarantee

$$\sup_{G \in \mathcal{C}} \left\| \widehat{\Phi}_s(G) - \psi(F(G)) \right\| \leq \eta \tag{14}$$

on a graph class $\mathcal{C} \subseteq \mathcal{G}_{n,d}$, then we obtain an immediate task-level bound.

Proposition 4.2 (Lipschitz transfer bound)**.** *Under* (14)*, we have*

$$\sup_{G \in \mathcal{C}} \left| g(\widehat{\Phi}_s(G)) - g(\psi(F(G))) \right| \leq L\eta.$$

Proof. For each $G \in \mathcal{C}$, Lipschitz continuity gives $|g(\widehat{\Phi}_s(G)) - g(\psi(F(G)))| \leq L\|\widehat{\Phi}_s(G) - \psi(F(G))\|$. Taking the supremum over $G \in \mathcal{C}$ and applying (14) yields the claim. $\square$

Randomized signatures and error decomposition. In practice, the computed signature is itself random, due to trace probes, sketches, or feature subsampling. Let $\widetilde{\Phi}_s(G)$ denote the (random) output of the signature

algorithm, and suppose we have a high-probability uniform approximation

$$\mathbb{P}\left[\ \sup_{G \in \mathcal{C}} \big\|\widetilde{\Phi}_s(G) - \widehat{\Phi}_s(G)\big\| \le \varepsilon\ \right] \ \ge\ 1 - \delta. \tag{15}$$

Assume in addition that the model-decoder pair recovers the *computed* signature up to η, i.e.,

$$\sup_{G \in \mathcal{C}} \big\|\widetilde{\Phi}_s(G) - \psi(F(G))\big\| \ \le\ \eta, \tag{16}$$

either deterministically or with high probability (e.g. over training randomness). Then on the event in (15), we have by the triangle inequality

$$\sup_{G \in \mathcal{C}} \big\|\widehat{\Phi}_s(G) - \psi(F(G))\big\| \ \le\ \sup_{G} \|\widehat{\Phi}_s(G) - \widetilde{\Phi}_s(G)\| + \sup_{G} \|\widetilde{\Phi}_s(G) - \psi(F(G))\| \ \le\ \varepsilon + \eta.$$

Applying Proposition 4.2 yields a task error bound of $L(\varepsilon + \eta)$ on this event. This decomposition is operationally useful: it cleanly separates *computational* approximation (controlled by R, sparsification, etc.) from *model* approximation (controlled by architecture and training).

Dataset-level diagnostic variants. Uniform bounds such as (14) are convenient for theory but often too strong as diagnostics. For a finite dataset $\mathcal{D} = \{G_i\}_{i=1}^N$, we consider empirical reconstruction losses such as

$$\widehat{\eta}_{\mathrm{avg}} \ := \ \frac{1}{N} \sum_{i=1}^N \big\|\widehat{\Phi}_s(G_i) - \psi(F(G_i))\big\|, \qquad \widehat{\eta}_{\mathrm{max}} \ := \ \max_{1 \le i \le N} \big\|\widehat{\Phi}_s(G_i) - \psi(F(G_i))\big\|.$$

By the same Lipschitz argument applied pointwise and then averaged, we obtain the empirical task discrepancy bound

$$\frac{1}{N} \sum_{i=1}^N \big| g(\widehat{\Phi}_s(G_i)) - g(\psi(F(G_i))) \big| \ \le\ L\,\widehat{\eta}_{\mathrm{avg}},$$

and similarly a uniform-on-dataset bound with $\widehat{\eta}_{\mathrm{max}}$. This yields a simple diagnostic procedure: compute $\widehat{\Phi}_s(G_i)$ once (or a high-accuracy proxy), train ψ to predict these signature coordinates from $F(G_i)$, and report $\widehat{\eta}_{\mathrm{avg}}$ or $\widehat{\eta}_{\mathrm{max}}$. Any downstream evaluation function g with a known Lipschitz constant L (in the chosen norm) immediately inherits a certified dataset-level approximation guarantee.

A particularly transparent case is when g is *coordinate projection*, e.g. $g(u) = u_k$. Under the ℓ_∞ norm, g is 1-Lipschitz, and the bound reduces to

$$\big| [\widehat{\Phi}_s(G)]_k - [\psi(F(G))]_k \big| \ \le\ \big\|\widehat{\Phi}_s(G) - \psi(F(G))\big\|_\infty,$$

so signature reconstruction directly measures the error in predicting that invariant statistic. This specialization is what makes coordinate-wise lower bounds (as in the next section) immediately interpretable as task lower bounds for suitable targets.

Generalization from finite samples (one simple route). When ψ is learned from data, one may combine the transfer principle with standard generalization bounds for regression. For instance, suppose ψ ranges over linear decoders $\psi_W(h) = Wh$ with a norm constraint $\|W\|_{\text{op}} \leq B$ (operator norm under the relevant pair of norms), and suppose $\|F(G)\| \leq M$ almost surely. Then $\|\psi_W(F(G))\| \leq BM$. If the signature coordinates are bounded (which holds for our coordinates on $\mathcal{G}_{n,d}$ with $A \in \{0,1\}^{n \times n}$ and $X \in [-1,1]^{n \times d}$ at fixed s), we may apply uniform convergence tools (e.g. Rademacher complexity for linear predictors) to bound the population reconstruction error by the empirical reconstruction error plus an $O(BM/\sqrt{N})$ complexity term (up to constants depending on the loss and the dimension). Composing with Proposition 4.2 yields population task guarantees of the schematic form

$$\mathbb{E}\big[\,|g(\widehat{\Phi}_s(G)) - g(\psi_W(F(G)))|\,\big] \;\lesssim\; L\Big(\widehat{\eta}_{\text{avg}} + \frac{BM}{\sqrt{N}}\Big),$$

with the understanding that the precise constant depends on the chosen learning setup. The salient point is that the only task dependence enters through L; the representation-dependent quantity is the signature reconstruction performance.

When the bound is tight. The Lipschitz transfer bound is sharp in the minimax sense: for any fixed error vector $\Delta = \widehat{\Phi}_s(G) - \psi(F(G))$, the worst-case L-Lipschitz functional can realize discrepancy essentially $L\|\Delta\|$ (up to norm duality). Concretely, for a linear functional $g(u) = \langle a, u \rangle$ with $\|a\|_* = L$ (dual norm), we have

$$|g(\widehat{\Phi}_s(G)) - g(\psi(F(G)))| = |\langle a, \Delta \rangle| \leq \|a\|_* \|\Delta\| = L\|\Delta\|,$$

and equality is attained when a aligns with Δ. Thus, absent structural assumptions on g, one cannot improve Proposition 4.2 beyond constant factors. Conversely, the bound can be loose when (i) g depends only on a small subset of stable coordinates, (ii) the reconstruction error concentrates on coordinates to which g is insensitive, or (iii) g is effectively lower-Lipschitz on the data manifold than on all of $\mathbb{R}^p$. These observations motivate our use of coordinate-wise diagnostics and targeted signatures: if a task is known to depend on specific coordinates (e.g. short-cycle traces or decorated walk energies), then controlling reconstruction error on those coordinates yields a correspondingly tighter and more interpretable guarantee.

Strict refinement: a concrete WL-indistinguishable family separated by cycle traces. We exhibit an explicit pair of featured graphs on which any 1-WL-equivalent MPNN (run on constant input features) collapses, while a single low-order coordinate of our signature separates. Fix n

divisible by 12, set $d = 1$, and take constant features $X = \mathbf{1} \in \mathbb{R}^{n \times 1}$. Let

$$G_n := \bigsqcup_{i=1}^{n/3} C_3, \qquad G'_n := \bigsqcup_{i=1}^{n/4} C_4,$$

i.e. G_n is the disjoint union of $n/3$ triangles and G'_n is the disjoint union of $n/4$ 4-cycles. Both graphs are 2-regular, and with constant features every vertex has the same rooted t-hop neighborhood multiset type for every fixed t. Consequently, 1-WL produces the trivial coloring on both graphs, and the standard indistinguishability argument implies that any 1-WL-equivalent message passing architecture F assigns identical node embeddings within each graph at every layer, hence identical graph embeddings after any permutation-invariant readout. In particular,

$$F(G_n) = F(G'_n) \qquad \text{for all such } F \text{ (on constant features).}$$

On the other hand, the cycle-trace coordinate at $\ell = 3$ separates the pair. Since $\mathrm{tr}(A^\ell) = \hom(C_\ell \to G)$, we have for G_n that each triangle contributes $\mathrm{tr}(A_{C_3}^3) = 6$ (indeed the eigenvalues of C_3 are $2, -1, -1$), and therefore

$$\mathrm{tr}(A(G_n)^3) = \frac{n}{3} \cdot 6 = 2n.$$

For G'_n, each C_4 is bipartite, hence has no closed walks of odd length, so $\mathrm{tr}(A(G'_n)^3) = 0$. Thus for any $s \geq 3$ the full signature Φ_s (and likewise the sketched signature $\widehat{\Phi}_s$, which includes $\mathrm{tr}(A^3)$ as a basic coordinate) satisfies

$$[\Phi_s(G_n)]_{\mathrm{tr}(A^3)} \neq [\Phi_s(G'_n)]_{\mathrm{tr}(A^3)}.$$

This already witnesses strict refinement over 1-WL-limited expressivity in a particularly transparent form: the obstruction is global (odd-cycle presence) and is invisible to local multiset aggregation starting from fully symmetric features.

Lower bounds for signature recovery by WL-limited models. We now translate the above separation into a reconstruction lower bound for any attempt to predict signature coordinates from a 1-WL-equivalent embedding. Consider the normalized coordinate

$$c(G) := \frac{1}{n} \mathrm{tr}(A^3),$$

so that $c(G_n) = 2$ and $c(G'_n) = 0$. Let F be any 1-WL-equivalent MPNN run on constant features and let ψ be any decoder. Since $F(G_n) = F(G'_n)$, we must have $\psi(F(G_n)) = \psi(F(G'_n))$, hence at least one of the two reconstruction errors for the scalar target $c(\cdot)$ is large:

$$\max\left\{|c(G_n) - \psi(F(G_n))|, \; |c(G'_n) - \psi(F(G'_n))|\right\} \geq \frac{|c(G_n) - c(G'_n)|}{2} = 1.$$

Equivalently,

$$\inf_{\psi} \sup_{G \in \{G_n, G'_n\}} |c(G) - \psi(F(G))| \geq 1,$$

which is a constant lower bound independent of n. The same argument applies verbatim to any separating coordinate of $\widehat{\Phi}_s$ (e.g. $\mathrm{tr}(A^\ell)$ for any fixed ℓ, or a decorated quadratic form $x^\top A^\ell x$ when x is part of the signature): whenever two graphs are 1-WL-indistinguishable under the admissible inputs, any 1-WL-equivalent F identifies them, forcing any decoder to incur at least half the coordinate gap on that pair. In view of the coordinate-projection specialization discussed earlier (take $g(u) = u_k$ and $\|\cdot\|_\infty$), such reconstruction lower bounds are immediately interpretable as task lower bounds for Lipschitz targets depending on those coordinates.

Beyond triangles: families separated by low-order spectral moments. The preceding example uses $\ell = 3$, but the phenomenon is not tied to triangles. For any fixed degree $\Delta \geq 2$, the class of Δ-regular graphs with constant features is collapsed by 1-WL into a single color class, hence is indistinguishable to any 1-WL-equivalent MPNN. Within this class, one may choose graph families with different low-order spectral moments, i.e. different values of $\mathrm{tr}(A^\ell)$ for some constant ℓ. Since $\mathrm{tr}(A^\ell)$ is the ℓ-th power-sum of eigenvalues, equality of $\mathrm{tr}(A^\ell)$ for all $\ell \leq s$ is a strong constraint; by selecting non-cospectral regular graphs (or even cospectral graphs differing in feature-decorated coordinates), one obtains separations at some bounded order. Thus, even at fixed small s, the signature coordinates include invariants that are provably inaccessible to purely local 1-WL-type aggregation on symmetric inputs.

Matching upper bounds: architectures that realize trace/decorated-walk coordinates. We next indicate why the lower bound is not an artifact of the coordinate choice but rather of the model class. The cycle-trace coordinates and feature-decorated walk energies in $\widehat{\Phi}_s$ admit near-linear randomized estimators based on repeated applications of A and global reductions, exactly as in the signature algorithm. Concretely, for $\mathrm{tr}(A^\ell)$ one may use Hutchinson estimators:

$$\widehat{\tau}_\ell(G) := \frac{1}{R} \sum_{r=1}^{R} z_r^\top A^\ell z_r, \qquad z_r \in \{\pm 1\}^n \text{ i.i.d.}$$

With $R = \Theta(\varepsilon^{-2} \log(1/\delta))$ we obtain additive error $O(\varepsilon \|A^\ell\|_F)$ with probability at least $1 - \delta$. Importantly, each term $z_r^\top A^\ell z_r$ can be computed by maintaining node-wise states $v_r^{(0)} := z_r$ and iterating the linear message passing update $v_r^{(t+1)} := A v_r^{(t)}$ for $t = 0, \ldots, \ell - 1$, followed by the global sum $\sum_i z_r(i) v_r^{(\ell)}(i)$. This computation is permutation-equivariant in the natural sense (permuting node labels permutes the sampled probe entries), and

it fits within architectures that possess either (i) an explicit global mixing channel capable of aggregating node-wise products into a graph-level scalar, together with internal randomized probes, or (ii) a stable positional/identifier channel (e.g. canonicalized Laplacian features or other symmetry-breaking node encodings) that supplies nontrivial initial vectors from which the same linear-algebraic computation can proceed deterministically.

Similarly, feature-decorated energies $x_j^\top A^\ell x_j$ are computed by the same ℓ rounds of linear message passing on the input feature channel x_j, followed by a global inner product. Hence, the coordinates that separate G_n and G_n' are efficiently computable by models with suitable global reductions and either internal randomness or a positional mechanism that prevents total symmetry collapse. In this sense, the above lower bounds are tight with respect to the computational primitives underlying $\widehat{\Phi}_s$: they separate WL-limited architectures from those that can stably implement the corresponding global linear-algebraic probes.

Summary of the separation mechanism. The separation may be stated succinctly: on fully symmetric inputs, 1-WL-equivalent message passing is constrained to functions constant on each 1-WL class, whereas $\widehat{\Phi}_s$ contains low-order homomorphism/spectral statistics (and their feature-decorated analogues) that vary within these classes. Any model class capable of reconstructing these coordinates must, in some form, break the WL symmetry—either by augmenting the computation with global mixing plus suitable probe vectors, or by providing additional positional structure that is not expressible by 1-WL refinement alone.

Empirical protocol (optional): probing for signature recovery and its relation to task accuracy. While the preceding results are formal, we may strengthen the story empirically by treating $\widehat{\Phi}_s(G)$ as a *measurable expressivity target* and asking to what extent different architectures recover it from their learned embeddings. Concretely, we fix an order s and sketch budget R, compute $\widehat{\Phi}_s(G)$ for each graph G in a dataset, train a family of models F_θ on standard downstream tasks, and subsequently evaluate whether a fixed decoder ψ can predict $\widehat{\Phi}_s(G)$ from the learned representation $F_\theta(G)$. This yields a quantitative proxy for the reconstruction error η appearing in the Lipschitz transfer bound, and allows us to test whether smaller empirical η aligns with improved downstream generalization.

Signature computation and normalization. For each graph $G = (A, X)$, we compute $\widehat{\Phi}_s(G)$ using the same coordinate family across all methods (e.g. cycle/walk traces $\widehat{\tau}_\ell$ for $\ell \le s$, feature-decorated energies $x_j^\top A^\ell x_j$, and k-hop diffusion moments). Since coordinates may differ in scale across datasets, we

define a normalized signature

$$\widetilde{\Phi}_s(G) \ := \ D^{-1}\big(\widehat{\Phi}_s(G) - \mu\big),$$

where μ and the diagonal scaling D are computed on a training split (coordinate-wise mean and standard deviation, or robust alternatives such as median/MAD). We report errors in both ℓ_2 and ℓ_∞ norms, and we include a coordinate-wise evaluation for interpretable invariants (e.g. $\mathrm{tr}(A^3)$, $\mathrm{tr}(A^4)$, and selected feature moments). When randomized trace estimators are used, we either (i) fix the probe vectors across runs to reduce evaluation noise, or (ii) average over independent probe seeds and report confidence intervals as a function of R.

Embeddings to be probed. We probe graph representations produced by several model classes trained on the same downstream supervision. For message passing we include representative 1-WL-equivalent architectures (e.g. GCN/GraphSAGE/GIN with sum aggregation), with identical treatment of node features and global pooling. As stronger baselines we include subgraph-based models (e.g. k-GNN variants or architectures that explicitly aggregate over small tuples/subgraphs) and graph transformers with global mixing (e.g. sparse attention, structural encodings, or GPS-type hybrids). To avoid confounding by parameter count, we match models by either (a) total parameter budget, or (b) wall-clock/SpMV-equivalent compute. We additionally consider feature regimes that are known to induce symmetry collapse: constant features $X = \mathbf{1}$, low-entropy categorical features, and (as a control) randomized features.

Probe design and reconstruction metrics. Given a trained model F_θ, we freeze θ and fit a probe ψ to predict $\widetilde{\Phi}_s(G)$ from $F_\theta(G)$. The probe may be linear (ridge regression) to measure linearly accessible information, or a small MLP to measure information recoverable with modest nonlinearity. Denoting predictions by $\widehat{Y}(G) = \psi(F_\theta(G))$, we report:

$$\eta_2 \ := \ \Big(\mathbb{E}\| \widetilde{\Phi}_s(G) - \widehat{Y}(G) \|_2^2\Big)^{1/2}, \qquad \eta_\infty \ := \ \mathbb{E}\| \widetilde{\Phi}_s(G) - \widehat{Y}(G) \|_\infty,$$

estimated on a held-out test split. For coordinate-wise diagnostics we report R^2 and Pearson/Spearman correlation between each signature coordinate and its prediction. For trace-like coordinates we also report relative error against the (near-)exact value on small graphs where exact computation is feasible, thereby separating model error from sketch error.

Relating signature recovery to downstream accuracy. We then test whether empirical task performance aligns with signature recovery. For a

given dataset and supervised target $y(G)$, we record each model's test accuracy (or RMSE/AUROC) and its reconstruction error η_2 or η_∞. Across architectures (and across random restarts), we compute the correlation between performance and $-\eta$, and we optionally fit a simple regression of the form

$$\mathrm{Perf}(F_\theta) \approx a - b\,\eta_2(F_\theta),$$

to quantify how much of the variance in generalization is explained by signature accessibility. We emphasize that this is not a causal claim, but it operationalizes the qualitative prediction suggested by the Lipschitz transfer statement: when a task is well-approximated by a function of $\widehat{\Phi}_s$, models that better reconstruct $\widehat{\Phi}_s$ should tend to perform better. To further isolate this effect, we include synthetic targets $y(G) = g(\widehat{\Phi}_s(G))$ for known Lipschitz g (e.g. a linear functional with bounded norm, or a clipped polynomial), in which case the hypothesis becomes directly testable.

Ablations over the signature and over computational primitives. We carry out two complementary ablations. First, we ablate the signature basis by removing coordinate families (cycle traces only; decorated energies only; neighborhood moments only) and by varying s and R. This identifies which components are actually present in learned embeddings, and whether improvements come from capturing short cycles, feature-propagation statistics, or both. Second, we ablate architectural primitives that, by the preceding theory, should matter for global coordinates: (i) remove global attention/mixing from transformers; (ii) restrict readouts to purely local pooling; (iii) remove positional encodings or random-probe channels; (iv) constrain depth to be less than s. For each ablation we re-run probing and record the degradation in coordinate recovery, with special attention to trace-like coordinates that are provably inaccessible to 1-WL-equivalent message passing on symmetric inputs.

Benchmarks: WL-hard synthetic families and standard real-world datasets. To stress-test separations, we include synthetic datasets designed to break 1-WL (and in some cases higher WL), such as: regular graphs with differing short-cycle statistics; CFI-type constructions; and paired graph families that are 1-WL-indistinguishable but have different $\mathrm{tr}(A^\ell)$ for small ℓ. In these settings we can directly verify that MPNNs collapse (both in downstream classification and in signature probing), while models with global mixing or subgraph mechanisms recover separating coordinates. We then complement with real datasets where topology–feature interactions matter, such as molecular regression/classification benchmarks (e.g. ZINC, QM9, OGB molecular tasks) and citation/social graph-derived graph classification benchmarks (or ego-graph extracts from large networks). In molecular data, decorated walk energies and diffusion moments provide feature-aware

analogues of cycle statistics; in citation-style data, neighborhood moments probe oversmoothing/oversquashing phenomena through higher-hop feature propagation summaries.

Reporting and reproducibility. For each dataset we report: downstream performance, probe reconstruction errors (η_2, η_∞), coordinate-wise correlations for salient invariants, and ablation curves in s and R. We keep train/validation/test splits fixed, disclose probe hyperparameters, and include multiple random seeds for both model training and sketching. The resulting suite serves as an empirical map from *model class $\to$ recoverable joint feature–topology statistics*, complementing the formal separations by quantifying how the proposed signature behaves as an evaluative instrument in practical regimes.

Discussion, limitations, and extensions. The proposed joint feature–topology signature is intentionally positioned between two extremes: on the one hand, the full basis Φ_s provides a clean universality statement for invariant polynomials of bounded degree; on the other hand, the sketched object $\widehat{\Phi}_s$ is designed to be computable with near-linear primitives (SpMV, global sums, and randomized probes). This gap is not merely technical: it expresses an explicit trade-off between (i) basis completeness and (ii) the computational model one is willing to assume. In particular, any fixed, efficiently computable coordinate family necessarily excludes large portions of Φ_s (by hardness of exact subpattern counting), so the role of $\widehat{\Phi}_s$ should be understood as an *instrument* for measuring and separating model classes, rather than as a canonical or uniquely optimal representation.

Dependence on the choice of basis and on (s, R). A first limitation is that $\widehat{\Phi}_s$ is only as informative as the coordinates we choose to include. While cycle/walk traces and feature-decorated quadratic forms capture non-1-WL information and feature–structure interactions, they are not exhaustive even at fixed degree. Moreover, the utility of larger s is constrained by both compute and variance: increasing s increases the number of SpMV steps and can inflate estimator variance for trace-like terms involving A^ℓ when $\|A^\ell\|_F$ grows quickly with ℓ. Similarly, R controls estimator concentration but may be costly when many coordinates are probed. In practice, there is an additional statistical issue: when we normalize coordinates across a dataset, rare patterns may have heavy-tailed empirical distributions, making ℓ_2-based errors sensitive to a small number of graphs. These considerations motivate adaptive schemes (e.g. selecting coordinates by stability, mutual information with labels, or sensitivity to architectural ablations) rather than committing to a single fixed basis.

Inductive settings and varying graph sizes. Our definitions are compatible with inductive learning across graphs with varying n and m, but care is required to ensure that coordinates are comparable across sizes. Some coordinates (e.g. $\mathrm{tr}(A^\ell)$) scale superlinearly in n for typical sparse families, while others (e.g. averaged diffusion moments) remain $O(1)$. Thus, for inductive benchmarks we typically require size-normalized variants, such as $n^{-1}\mathrm{tr}(A^\ell)$, $m^{-1}\mathrm{tr}(A^\ell)$, or normalization by powers of an estimated spectral radius. A related limitation is that the signature, being global, can confound "more structure" with "more nodes" unless normalization is explicitly enforced. A systematic treatment would specify a normalization convention as part of the definition of $\widehat{\Phi}_s$, possibly conditioned on a graph class (bounded degree, bounded expansion, etc.) to avoid degenerate scaling regimes.

Feature privacy and leakage in probing-style evaluations. Because coordinates of $\widehat{\Phi}_s$ are explicit global statistics, they can leak information about sensitive attributes present in X or about rare substructures. This is not an artifact of our method: any sufficiently expressive invariant is potentially identifying. However, our setting makes the issue concrete because we propose to *measure* recoverability of these statistics from learned embeddings. If privacy is a concern, one may (i) restrict to coordinates that are provably insensitive to individual features (e.g. clipping feature monomials, enforcing bounded influence), (ii) publish only compressed sketches (e.g. CountSketch outputs) rather than raw coordinates, and/or (iii) add calibrated noise to obtain approximate differential privacy guarantees for released signatures. The last option is particularly natural because many coordinates are sums of bounded contributions; for instance, if $X \in [-1,1]^{n \times d}$ and we use clipped moments, then Laplace or Gaussian mechanisms can be applied with sensitivity bounds depending on the clipping parameters and on the operator norm of the diffusion. A principled privacy analysis would require specifying an adjacency notion (node-level vs. edge-level vs. feature-level) and is an open direction.

Beyond simple undirected graphs: weights, directions, and edge features. We have focused on simple graphs with $A \in \{0,1\}^{n \times n}$ symmetric and zero diagonal. Many applications involve weighted edges, directed edges, and edge features. Most of our coordinates extend formally by replacing A with a general sparse matrix W (possibly asymmetric) and defining trace-like quantities $\mathrm{tr}(W^\ell)$ for directed cycle weights, along with bilinear forms $x_j^\top W^\ell x_j$. When W is asymmetric, Hutchinson estimators still apply to $\mathrm{tr}(W^\ell)$, but variance and stability may differ, and one may prefer symmetrizations (e.g. $WW^\top$) depending on the semantics. Edge features can be incorporated by considering multiple adjacency channels $\{A_r\}_{r=1}^{|\mathcal{R}|}$ and including mixed monomials such as $\mathrm{tr}(A_{r_1} \cdots A_{r_\ell})$ or feature-decorated

forms $x^\top A_{r_1} \cdots A_{r_\ell} x$. This yields a multi-relational signature appropriate for knowledge graphs, at the cost of a basis that grows with $|\mathcal{R}|$ and with the number of mixed products.

Heterogeneous graphs and typed nodes. For heterogeneous graphs with node types, one extension is to treat type indicators as additional feature channels and reuse the same coordinate families. This yields invariants that capture type-conditioned walk statistics (e.g. how often a length-ℓ walk begins and ends in a given type). A second, more structured extension is to enforce type constraints at the pattern level by restricting homomorphisms to type-respecting maps; equivalently, we consider a typed pattern H and count only maps $h \in \mathrm{Hom}(H \to G)$ that preserve types. The algorithmic analogue is to insert diagonal masks between adjacency multiplications (e.g. $A M_{\mathrm{type}} A$), which remains compatible with SpMV primitives but introduces additional design degrees of freedom (which masks, in what order, and how to control coordinate growth).

Dynamic graphs and streaming updates. For dynamic graphs (edge insertions/deletions and evolving features), recomputing $\widehat{\Phi}_s(G)$ from scratch may be wasteful. Some coordinates admit incremental updates: for example, diffusion moments based on $P^k X$ can be updated approximately using iterative methods when P changes slowly, and quadratic forms $x^\top A^\ell x$ can be maintained via truncated Krylov subspace techniques. Trace-like coordinates are more delicate under local edits, but there exist streaming estimators for spectral moments that update sketches in sublinear time under certain sparsity assumptions. Establishing rigorous error propagation bounds for such incremental schemes, especially when s is moderate and the graph changes adversarially, is an open direction.

Optimal bases and compute–expressivity frontiers. A central open question is to characterize, for a given computational budget (e.g. $\tilde{O}(m)$ or $\tilde{O}(ms)$ time), what is the maximal separation power achievable by *any* permutation-invariant signature. In our framework this becomes: among all coordinate families computable by a prescribed set of primitives (SpMV, bounded-depth message passing, sparse attention, randomized probing), which subspace of $\mathbb{R}[A, X]^{S_n}_{\leq s}$ can be approximated to accuracy ε with failure probability δ? Relatedly, one may ask for minimax lower bounds: for a class of graphs $\mathcal{C}$, what sketch dimension p and probe budget R are necessary to distinguish or approximate all functions in a target function class (e.g. Lipschitz functions of Φ_s)? Finally, there is a modeling question: can one *learn* a basis adaptively, selecting coordinates (or random features) that best explain downstream tasks while maintaining provable invariance and concentration? Addressing these questions would convert the present construction from a

principled measuring device into a theory of optimal invariant computation under resource constraints.