# Concorde: Geometry-Aware Link Prediction via Decoupled Energy Minimization

**Ege Demirci**
Department of Computer Science, UC Santa Barbara
Santa Barbara, CA 93106, USA

**Ambuj Singh**
Department of Computer Science, UC Santa Barbara
Santa Barbara, CA 93106, USA

## Abstract

Link prediction is difficult in sparse graphs, where diffusion heuristics lack connectivity and geometric methods suffer spectral instability when coupling learned representations with diffusion operators. We introduce **Concorde**, an energy-based framework that resolves this by decomposing link likelihood into three structurally decoupled branches: **(i)** a parameter-free spectral anchor that establishes permutation-invariant topological stability; **(ii)** a mixed-curvature residual on a learnable $\mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$ product manifold, regularized by a feature-aware Ollivier–Ricci curvature proxy; and **(iii)** a continuous vector field on $\mathcal{T}\mathbb{H}^d_c$ for non-local structural alignment across topological gaps. The branches are fused via learned gating and trained with an $N$-pair ranking loss aligned with retrieval-style evaluation. Across five benchmarks spanning sparse, dense, featureless, and attributed regimes, **Concorde** establishes state-of-the-art early retrieval precision, nearly doubling $P@10\%$ over leading baselines in sparse settings.

## 1 Introduction

Link prediction aims to rank candidate node pairs so that missing but true edges appear above non-edges [25]. Despite strong progress, no single family of methods handles all structural regimes reliably. Neural approaches, such as GNN-based predictors, subgraph reasoning, and neighborhood-aggregation models, learn flexible link rules from node attributes but degrade sharply under sparsity [5, 39, 43, 48, 49]. While classical local overlap heuristics degrade in these regimes due to their strict reliance on dense 2-hop neighborhoods [1, 29, 50], global diffusion-style measures remain competitive, encoding multi-hop connectivity without relying on local neighborhood density [20–22, 27].

Geometric deep learning offers a principled route to unifying these two families. Different graph motifs prefer different curvatures: tree-like hierarchies are faithfully represented in negatively curved hyperbolic spaces, while clique-dense subgraphs are better captured under positive curvature [2, 3, 6, 15, 31, 38]. Discrete graph curvature (e.g., Ollivier–Ricci) provides a complementary *intrinsic* descriptor, distinguishing community edges from bridge-like edges between regions [30, 32, 37, 47]. However, two issues limit the practical impact of geometry for link prediction. **(i)** Most geometric methods use a single manifold and decode distances directly, leaving it unclear how to couple learned geometry with diffusion operators without destabilizing the graph's global spectrum [14, 45]. **(ii)** Most curvature-based methods compute curvature only on observed edges and do not learn a continuous, feature-driven similarity that generalizes to unobserved pairs [35, 37].

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

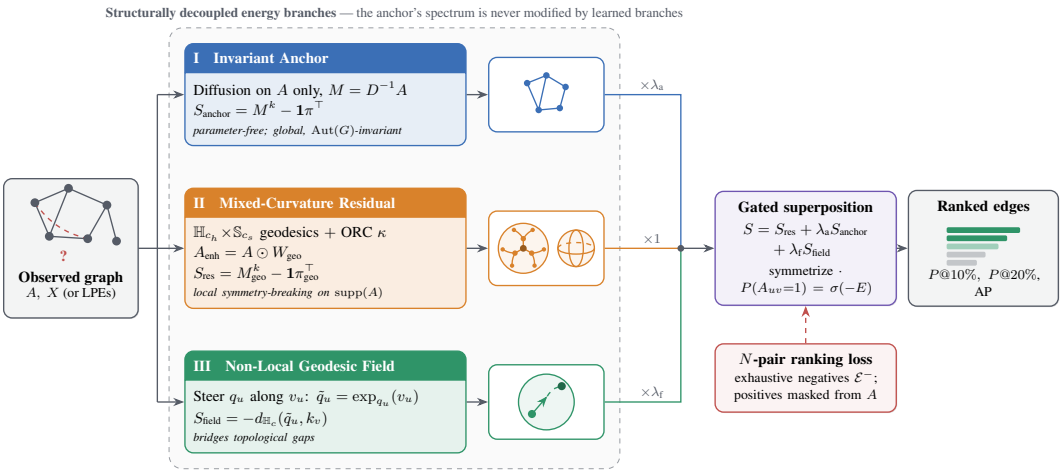


Figure 1: **Overview of Concorde.** Link likelihood is decomposed into three structurally decoupled energy branches: a parameter-free spectral anchor on the unperturbed adjacency (global, $\mathrm{Aut}(G)$-invariant), a mixed-curvature residual on $\mathbb{H} \times \mathbb{S}$ with a feature-aware Ollivier–Ricci guardrail (local symmetry-breaking), and a non-local geodesic steering field on $\mathcal{T}\mathbb{H}^d_c$ (bridging topological gaps). Learned scalar gates fuse the branches, and the model is trained with an $N$-pair ranking loss aligned with early-retrieval evaluation.

**Contributions:**

- **Spectral decoupling hypothesis.** We prove that a single learned adjacency cannot be simultaneously spectrum-preserving (a spectral filter of $A$) and $\mathrm{Aut}(G)$-symmetry-breaking (Proposition 1), and confirm it directly with a coupled single-matrix control that loses 1.8–2.5 AP exactly where the theory predicts (Appendix A.4).
- **A tri-branch energy model.** Concorde decouples link likelihood into a parameter-free spectral anchor, a mixed-curvature residual with a feature-aware Ollivier–Ricci guardrail, and a non-local hyperbolic steering field, fused by learned gates and trained with an $N$-pair ranking loss.
- **State-of-the-art early retrieval.** Across five benchmarks spanning sparse/dense and featureless/attributed regimes, Concorde sets a new state of the art in early precision, nearly doubling $P@10\%$ over the strongest baselines on sparse graphs, with ablations isolating each branch.

### 1.1 Related Work

Classical link predictors based on neighborhood overlap, such as Common Neighbors, Adamic–Adar, and Resource Allocation, remain strong baselines due to their direct encoding of triadic closure [1, 29, 50]. Global diffusion measures such as Katz [22] and SimRank [21] exploit multi-hop connectivity via path-based similarity. Autocovariance (AC) captures random-walk co-visitation structure and can outperform learned GNN baselines on sparse graphs [20]. Gelato refines a diffusion heuristic with a light attribute-driven learner and ranking objective, demonstrating that heuristic reweighting is extremely effective under heavy imbalance [27].

A second line of work performs edge scoring via explicit local-structure reasoning. SEAL popularized enclosing-subgraph classification; subsequent methods improved scalability via neighborhood sketching [5] or modeled explicit common-neighbor structure [48]. NCN/NCNC incorporate common-neighbor structure with false-negative completion [43], and OCN further refines these signals via orthogonalization to address redundancy and over-smoothing [42].

Geometric representation learning embeds nodes in non-Euclidean spaces to match global graph structure. Hyperbolic embeddings compactly model hierarchies via exponential volume growth [31], and hyperbolic GNNs extend message passing to these manifolds [6]. Unified constant-curvature frameworks [2] and mixed-curvature product spaces [15, 38] capture multiple structural motifs

simultaneously, though most rely on single decoders and do not address spectral brittleness when coupling learned geometry to diffusion operators. More recent work combines mixed-curvature geometry with global attention in product-space transformers, or learns per-component curvature importance rather than fixing it *a priori* [10, 15, 38]. Concorde differs from existing literature in two ways: its non-local branch replaces Euclidean self-attention with a *geodesic* steering field native to the hyperbolic manifold (Section 2.3), and its learned gates weight *structurally decoupled* energy branches rather than reweighting components inside one representation, which is precisely what preserves the anchor spectrum (Proposition 2).

Furthermore, discrete Ricci curvature characterizes local graph "shape," distinguishing community edges from topological bridges [30, 32]. Curvature Graph Networks adapt edge weights via Ollivier–Ricci curvature during message passing [47], discrete curvature has been used to characterize over-squashing and bottlenecks in message passing [40], and curvature has been used as an auxiliary signal in applied link prediction [37]. These methods provide valuable intrinsic descriptors but do not learn task-adaptive similarity from node features or integrate with diffusion-based pipelines.

## 2 Methodology

We introduce **Concorde**, an energy-based framework that decouples link likelihood into three structurally restricted branches, each targeting a distinct functional regime.

Let $G = (V, E)$ be an undirected graph with adjacency $A \in \mathbb{R}^{N\times N}$ and features $X \in \mathbb{R}^{N\times d}$; link prediction scores each pair $(u, v)$ by a structural similarity $S(u, v)$ [25]. Rather than deriving $S$ from a distance decoder in a single latent space [18, 31, 36], we treat the negative similarity as a *geometric energy* $E(u, v)$ and model link formation as a Boltzmann distribution with a logistic link:

$$P(A_{uv} = 1) = \sigma(-E(u, v)). \tag{1}$$

Low energy thus corresponds to a high-probability missing link. Standard learned methods minimize this energy within a single representation space or enhanced adjacency $A_{\text{learn}}$ [7, 11]. We argue this forces an unavoidable spectral trade-off between two opposing group-theoretic objectives: **global invariance** (spectral filters on $A$ are permutation-invariant w.r.t. $\mathrm{Aut}(G)$) and **local equivariance** (detecting missing links requires distinguishing structurally symmetric candidates via localized geometric deformations).

We formalize this as **a spectral decoupling hypothesis**. If the single learned operator $A_{\text{learn}}$ acts as a *spectral filter* $\sum_k c_k A^k$ of $A$, as every diffusion operator (including the anchor's $M^k$) does, it is automatically $\mathrm{Aut}(G)$-equivariant, the standard condition satisfied by any message-passing GNN [26, 46], and assigns identical scores within each $\mathrm{Aut}(G)$-orbit. The symmetry-breaking that link prediction requires is thus impossible without leaving this spectral class and distorting the spectrum (Proposition 1, Appendix A). The trade-off is thus structurally generic. Because informative node features can partially break $\mathrm{Aut}(G)$-symmetry, Proposition 1 characterizes the *structural* component of the learned signal: the impossibility is most binding where features are weak or absent relative to topology. Concorde abandons the single-matrix paradigm accordingly, a choice supported by the theory (Appendix A), by the performance collapse under ablation (Section 3.2), and by a direct coupled single-matrix experiment. We define the total geometric energy $E(u, v)$ as a decoupled superposition of three distinct potentials, each restricted to a specific functional space. By setting the final pairwise score $S(u, v)$ to the negative of this joint energy, Concorde computes link probabilities through a Tri-Branch architecture:

$$S(u, v) = -E(u, v) = S_{\text{anchor}}(u, v) + S_{\text{res}}(u, v) + S_{\text{field}}(u, v). \tag{2}$$

The three branches are: (I) a parameter-free spectral anchor that captures global topological invariance; (II) a mixed-curvature residual on a learnable $\mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$ product manifold for local geometric symmetry-breaking; and (III) a continuous vector field on $\mathcal{T}\mathbb{H}^d_c$ for non-local structural alignment across topological gaps. We discuss the three branches next.

### 2.1 Branch I: Global Topological Energy (The Invariant Anchor)

The first branch of the Concorde architecture establishes the macroscopic stability of the model. In multiscale geometric frameworks, stability is typically maintained by proving that a hierarchy of

coarsened graphs remains quasi-isometric to the original topology [24, 51]. Concorde, however, achieves this stability natively in a single scale by strictly isolating the macroscopic topology into an invariant anchor. This branch minimizes the *global topological energy*, measuring structural affinity between nodes from the observed network infrastructure alone. Graph automorphisms, permutations preserving adjacency, act as linear transformations on the disjoint eigenspaces of $A$; consequently, any parameter-free spectral function of $A$ is invariant to these symmetries by construction.

Let $D$ be the diagonal degree matrix of the training graph where $D_{uu} = \sum_v A_{uv}$, and define the row-stochastic transition matrix $M = D^{-1}A$. To evaluate the structural energy between nodes $u$ and $v$ over a receptive field of $k$ hops, we model the multi-step diffusion process. Let $\pi \propto D\mathbf{1}$ represent the stationary distribution column vector of $M$, and let $\mathbf{1}$ be the all-ones column vector. The invariant anchor score matrix is formally defined as a degree-normalized diffusion residual:

$$S_{\text{anchor}} = M^k - \mathbf{1}\pi^T. \tag{3}$$

Each entry $[S_{\text{anchor}}]_{uv} = [M^k]_{uv} - \pi_v$ measures the *excess* $k$-step reachability of $v$ from $u$ above the stationary baseline $\pi_v$. This is related to, but distinct from, the standard Autocovariance (AC) heuristic of Huang et al. [20], which uses the degree-weighted form $\Pi M^k - \pi\pi^T$ where $\Pi = \text{diag}(\pi)$. Our row-stochastic formulation omits the $\Pi$ pre-weighting deliberately: the $\Pi$-weighted form amplifies hub nodes in proportion to their stationary probability, introducing precisely the popularity bias $S_{\text{anchor}}$ is designed to suppress. The next two branches supply the fine-grained symmetry-breaking signal that this parameter-free anchor cannot provide by construction.

**Well-posedness on real graphs:** A unique stationary $\pi$ requires the walk to be aperiodic; we ensure this with self-loops (a standard lazy-walk device, enabled on the datasets most prone to periodicity), which also give dangling low-degree nodes well-defined stationary mass. On disconnected graphs, $M = D^{-1}A$ is block-diagonal, so Branch I operates independently within each connected component ($\pi \propto D\mathbf{1}$ is well-defined per block) and never scores cross-component pairs.

### 2.2 Branch II: Local Geometric Energy (Mixed-Curvature Residual)

While the invariant anchor stabilizes the macroscopic ranking, it is blind to localized symmetry-breaking: nodes in the same structural orbit receive identical scores, yet link prediction hinges on detecting localized perturbations such as a closing triangle or a missing cross-community bridge. Concorde captures these high-frequency corrections without destabilizing the global spectrum through a *local geometric energy* branch on a mixed-curvature manifold.

Real-world graphs are geometrically heterogeneous: tree-like regions embed best in negative curvature, while cyclic or clique-dense subgraphs favor positive curvature [4, 34], and no single Euclidean or purely hyperbolic space accommodates both without distortion [15]. We therefore map node features into a product manifold $\mathcal{M} = \mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$ with learnable curvatures $c_h < 0$ and $c_s > 0$, and compute coordinate-free geodesic distances $d_{\mathbb{H}}(u, v)$ and $d_{\mathbb{S}}(u, v)$ on each component. Learning $c_h, c_s$ end-to-end lets the model scale each component to the dataset's intrinsic geometry rather than fixing it *a priori*.

Geodesic proximity is necessary but not sufficient: distances alone cannot separate a coherent edge from a spurious bridge between incompatible neighborhoods. We therefore add a feature-aware Ollivier–Ricci curvature (ORC) proxy $\kappa(u, v)$ measuring neighborhood alignment, with positive curvature indicating community cohesion and negative curvature a topological bottleneck or bridge.

Formally, we define lazy random-walk measures $\mu_u$ and $\mu_v$ over the one-hop neighborhoods of $u$ and $v$ respectively. Each measure assigns a retention mass $\alpha \in (0, 1)$ to the node itself and distributes the remaining $1 - \alpha$ uniformly over its immediate neighbors $\mathcal{N}(\cdot)$:

$$\mu_u(x) = \alpha \cdot \mathbf{1}[x = u] \; + \; \frac{1-\alpha}{|\mathcal{N}(u)|} \sum_{w \in \mathcal{N}(u)} \mathbf{1}[x = w]. \tag{4}$$

The curvature is then defined via optimal transport between these localized distributions:

$$\kappa(u, v) = 1 - \mathcal{W}_1(\mu_u, \mu_v), \tag{5}$$

where $\mathcal{W}_1$ denotes the Wasserstein-1 distance. To ensure scalability to large graphs, this transport cost is approximated via the highly parallelizable entropy-regularized Sinkhorn algorithm [8] and

precomputed for candidate batches. For $\kappa(u, v)$ to reflect both structural and semantic compatibility, the ground metric must account for more than hop distance. We define a joint ground metric $C(a, b)$ as a convex combination of topological and semantic dissimilarity:

$$C(a,b) = \frac{1}{2} d_{\text{topo}}(a,b) + \frac{1}{2}(1 - \cos_\text{sim}(X_a, X_b)), \quad \text{where } d_{\text{topo}}(a,b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } (a,b) \in E \\ 2 & \text{otherwise} \end{cases}. \tag{6}$$

This coupling ensures that the transport cost, and by extension $\kappa(u, v)$, penalizes pairs whose neighborhoods are geometrically proximate yet semantically incompatible, providing an intrinsic check against structurally implausible link predictions. We fuse the extrinsic manifold distances and the intrinsic ORC guardrail via a deep neural function $\mathcal{F}_\theta$ to compute a local structural modulator:

Let $\bar{d}_{\mathbb{H}}(u)$ and $\bar{d}_{\mathbb{S}}(u)$ be the mean geodesic distances from $u$ to its observed neighbors, encoding local manifold density. The structural modulator $W_{\text{geo}} \in \mathbb{R}^{N \times N}$ has entries:

$$[W_{\text{geo}}]_{uv} = \sigma\Big(\mathcal{F}_\theta\big(\tfrac{X_u + X_v}{2} \oplus |X_u - X_v|,\ d_{\mathbb{H}}(u,v),\ d_{\mathbb{S}}(u,v),\ \kappa(u,v),\ \bar{d}_{\mathbb{H}}(u),\ \bar{d}_{\mathbb{H}}(v),\ \bar{d}_{\mathbb{S}}(u),\ \bar{d}_{\mathbb{S}}(v)\big)\Big). \tag{7}$$

Here $\frac{X_u + X_v}{2}$ and $|X_u - X_v|$ decompose the pair into symmetric and difference feature components. In strictly featureless regimes (e.g., OGBL-DDI), we substitute the absent node attributes $X$ with Laplacian Positional Encodings (LPEs) to provide initial structural coordinates. Because the graph is undirected, the resulting pairwise relationships must be perfectly symmetric. We therefore explicitly symmetrize the structural modulator matrix:

$$W_{\text{geo}} \leftarrow \frac{1}{2}\left(W_{\text{geo}} + W_{\text{geo}}^\top\right). \tag{8}$$

This symmetrized, bounded affinity matrix acts as a localized, equivariant deformation of the graph space. We apply this modulator to reweight the observed topology via the Hadamard product, constructing an enhanced adjacency matrix $A_{\text{enh}} = A \odot W_{\text{geo}}$.

Defining $D_{\text{enh}}$ as the corresponding diagonal degree matrix where $[D_{\text{enh}}]_{uu} = \sum_v [A_{\text{enh}}]_{uv}$, we construct the enhanced transition matrix $M_{\text{geo}} = D_{\text{enh}}^{-1} A_{\text{enh}}$. This dynamically injects high-frequency structural cues into the transition probabilities while strictly preserving the sparsity of the original graph. The residual energy score is then extracted via diffusion over this augmented space (with $\pi_{\text{geo}}$ as its stationary distribution):

$$S_{\text{res}} = M_{\text{geo}}^k - \mathbf{1}\pi_{\text{geo}}^T. \tag{9}$$

The Hadamard form $A \odot W_{\text{geo}}$ amplifies structurally consistent edges and dampens spurious ones while keeping $O(|\mathcal{E}|)$ cost. Adding the unperturbed anchor ($S_{\text{anchor}} + S_{\text{res}}$) acts as a stability regularizer: the global ranking cannot fall below the baseline topological performance. This enforces a separation of concerns, the anchor supplies low-frequency global topology, the residual applies feature-adaptive local reweighting confined to the sparsity support of $A$, and Branch III resolves unobserved gaps.

### 2.3 Branch III: Non-Local Field Energy (Contextual Geodesic Steering)

Branches I and II capture structure within connected subgraphs, but diffusion breaks down under extreme sparsity: two nodes may share identical structural orbits yet be entirely disconnected. Standard attention bridges this via pairwise affinities, but its Euclidean inner products ($QK^T$) are geometrically inconsistent with the curved manifold of Branch II [10, 16, 41]. Concorde instead learns a *continuous feature-to-tangent map* $\mathcal{V} : \mathbb{R}^d \to \mathcal{T}\mathbb{H}_c^d$ that assigns each query node a tangent vector pointing toward its most probable missing link, steering natively on the manifold without neighborhood traversal or costly parallel transport.

First, we project the node features into the tangent space of the origin and lift them to the Lorentz model to form query and key representations $q_u, k_v \in \mathbb{H}_c^d$. In featureless settings, the same Laplacian Positional Encodings (LPEs) used in Branch II serve as the input to the feature-to-tangent map $\mathcal{F}_{\text{steer}}$; crucially, since LPEs are precomputed on the full training graph and fixed throughout training, per-batch edge masking does not require their recomputation, keeping the overhead of Branch III identical across attributed and featureless regimes. For each node $u$, the network generates a specific tangent steering vector:

$$v_u = \text{proj}_{q_u}\big(\mathcal{F}_{\text{steer}}(X_u)\big) \in \mathcal{T}_{q_u}\mathbb{H}_c^d, \tag{10}$$

where the projection operator ensures the steering vector lies strictly in the tangent space of $q_u$ (i.e., the Minkowski inner product $\langle q_u, v_u \rangle_{\mathcal{L}} = 0$). To evaluate the structural alignment between the query's intention and the candidate key, we utilize the Lorentz Exponential Map [28]. This operator transports the query node along the exact geodesic trajectory defined by the semantic steering vector (using $|c_h|$ to denote the magnitude of the negative curvature):

$$\tilde{q}_u = \exp_{q_u}(v_u) = \cosh(\sqrt{|c_h|}||v_u||_{\mathcal{L}})q_u + \frac{\sinh(\sqrt{|c_h|}||v_u||_{\mathcal{L}})}{\sqrt{|c_h|}||v_u||_{\mathcal{L}}}v_u. \tag{11}$$

Finally, the *Non-Local Field Energy* is computed as the geodesic distance between the semantically steered query and the key across the manifold. We convert this energy into an attention-based similarity score:

$$S_{\text{field}}(u, v) = -d_{\mathbb{H}_c}(\tilde{q}_u, k_v). \tag{12}$$

A lower geodesic distance between the steered query $\tilde{q}_u$ and the key $k_v$ corresponds to a higher structural affinity, so the negative distance serves directly as a pairwise similarity score. To capture multi-faceted structural signals, we implement this mechanism using a multi-head attention architecture, where each head learns a distinct steering field and an independent curvature constant (separate from the base manifold in Branch II), which are then fused via a learnable weighted sum.

**Directedness and parameterization:** The query and key representations are produced by an *independent* feature-to-tangent lift, not shared with the Branch II encoder, leaving the steering field free to specialize for non-local alignment. The score $S_{\text{field}}(u, v) = -d_{\mathbb{H}_c}(\tilde{q}_u, k_v)$ is deliberately directed, it steers a query toward its most probable partner and is not symmetric in $(u, v)$. Because the target graph is undirected, symmetry is restored at fusion via $S(u, v) \leftarrow \frac{1}{2}(S(u, v) + S(v, u))$ (Section 2.4): the directed field enriches the hypothesis space during scoring while the final energy stays permutation-invariant.

### 2.4 Optimization & Multi-hop Superposition

To compute the final link likelihood, Concorde integrates the three decoupled energy branches into a unified scoring function. Rather than fixing the relative contribution of each branch, we learn a dynamic superposition via end-to-end gating:

$$S(u, v) = S_{\text{res}}(u, v) + \lambda_{\text{anchor}} S_{\text{anchor}}(u, v) + \lambda_{\text{field}} S_{\text{field}}(u, v), \tag{13}$$

where $\lambda_{\text{anchor}}, \lambda_{\text{field}} \in (0, 1)$ are sigmoid-parameterized scalar gates optimized end-to-end and $S_{\text{res}}$ is fixed at coefficient 1 as the baseline energy scale. Since the graph is undirected, we symmetrize the final score, $S(u, v) \leftarrow \frac{1}{2}(S(u, v) + S(v, u))$, resolving the asymmetry in Branch I's diffusion residual and Branch III's directed steering. The gates let the model adapt to dataset sparsity and homophily: dense graphs upweight $S_{\text{anchor}}$, sparse or heterophilic graphs upweight $S_{\text{field}}$. Empirically the learned gates recover this dependence without supervision, $\lambda_{\text{field}}$ rising monotonically as the graph sparsifies ($0.10$ on dense OGBL-DDI to $0.63$ on sparse Power), and outperform both equal and oracle-tuned fixed gates (Appendix F).

In sparse completion, non-edges grow quadratically with $|V|$, so calibrated binary classification is ill-posed and cross-entropy conflates ranking with calibration. Since our target is retrieval, a ranked shortlist of candidate edges, only the *relative ordering* of scores matters. We therefore supervise ordering directly with an $N$-pair ranking loss over true positives $\mathcal{E}^+$ and the exhaustive complement of unobserved negatives $\mathcal{E}^-$ (no subsampling):

$$\mathcal{L} = \frac{1}{|\mathcal{E}^+|} \sum_{(u,v)\in\mathcal{E}^+} \log\left(1 + \sum_{(u',v')\in\mathcal{E}^-} \exp\Big(S(u', v') - S(u, v)\Big)\right). \tag{14}$$

The exhaustive negative complement is made memory-feasible via graph partitioning and chunked evaluation (Appendix C). Penalizing any negative that scores above a true positive directly optimizes early-retrieval precision ($P@10\%$, $P@20\%$). To prevent structural leakage, we mask each mini-batch's positive edges $\mathcal{E}^+$ from $A$ before constructing the anchor and residual, so the diffusion branches infer link likelihood only from surrounding context. This masking is exact for the diffusion operators; on featureless graphs the fixed LPE inputs are exempt, but their residual leakage is bounded (Davis–Kahan) [9] and empirically negligible ($\leq 0.44$ AP; Appendices D, F).

# 3 Experiments

We evaluate Concorde on five benchmark datasets spanning distinct structural and semantic regimes, following the taxonomy of Mao et al. [25]. *Cora*, *CiteSeer* [33], and *PubMed* [12] are citation networks in which both global structural proximity and bag-of-words node features are informative, providing a rich testbed for methods that integrate structural and semantic signals. *OGBL-DDI* [19] is a dense, featureless drug-drug interaction network governed exclusively by local neighborhood overlap. *Power* [44] is an electrical grid topology at the opposite extreme: extremely sparse, featureless, and dominated by global path-based signals (e.g., Katz index) due to its low mean degree. These five are chosen to span the structural taxonomy along two axes, density (sparse Power to dense OGBL-DDI) and feature availability (featureless to attributed), so that each branch is stressed in isolation; OGBL-DDI's $\sim$1.3M edges additionally test scale, and Concorde's $O(k|E|)$ per-step cost (Appendix C) carries the identical protocol to larger graphs.

We compare against a comprehensive set of link prediction baselines. Enclosing subgraph methods include SEAL [49], which learns structural heuristics from localized subgraphs, and BUDDY [5], a scalable subgraph-sketching alternative. Neighborhood overlap methods, including Neo-GNN [48], NCN, NCNC [43], and the previous state-of-the-art OCNP [42], explicitly incorporate common-neighbor counts and higher-order overlaps on dense, triadic-closure-heavy graphs. We additionally include HGCN [6], a hyperbolic GNN with Fermi-Dirac link decoder, as a single-manifold geometric baseline against which the product manifold design of Branch II is directly contrasted. Finally, representing the previous state-of-the-art among diffusion-based methods, Gelato [27] frames link prediction as an imbalanced retrieval problem, augmenting topological heuristics with attribute-based graph learning.

We report **Average Precision (AP)** as a measure of global ranking quality, with primary emphasis on **Early Precision (**$P@10\%$**,** $P@20\%$**)**. We emphasize early precision because retrieval under extreme class imbalance is governed by the density of true edges at the very top of a *single global ranking* over the entire non-edge complement, which is what $P@r\%$ measures with a denominator held identical across methods, rather than by the sampled-negative shortlists that Hits@$K$/MRR are typically computed over; our exhaustive full-negative evaluation already imposes the hard-negative pressure that per-source protocols such as HeaRT approximate, and AP is reported throughout for global calibration (Appendix D). All baselines are run by us under the *identical* unbiased full-negative protocol and hardware; entries marked OOM or exceeding the time budget reflect a method's cost under this harder protocol rather than selective omission, as the per-method runtime and peak-memory tables in Appendix C make explicit.

## 3.1 Main Results: Breaking the Sparsity Bottleneck

Table 1 summarizes results across all five datasets, where Concorde either establishes a new state of the art or matches models specialized for a specific topological regime. Under extreme sparsity, true missing links frequently span multi-hop gaps with no common neighbors, and neighborhood-overlap methods degrade sharply (NCN $14.77\%$, NCNC $13.42\%$ $P@10\%$). Concorde reaches $27.58\%$ $P@10\%$ on Power, nearly doubling NCN/NCNC; the gain is attributable to Branch III (Section 2.3), whose learned feature-to-tangent map aligns node pairs across topological gaps without any path-based connectivity signal.

OGBL-DDI is the setting most favorable to explicit common-neighbor models: a dense, featureless network governed by local triadic closure. Concorde nonetheless surpasses the strongest overlap baseline OCNP at the top of the retrieval list ($46.45/46.48\%$ $P@10\%/P@20\%$ vs. $46.12/46.23\%$), though OCNP retains an AP advantage ($46.89\%$ vs. $42.32\%$). This gap is structurally expected: the $N$-pair loss optimizes score *ordering* among top candidates rather than global calibration, so it reflects a deliberate bias toward early retrieval, not a failure of discriminability. Notably, Concorde matches these specialized overlap models at top-$k$ precision using only the spherical manifold component and positional encodings, without explicit neighborhood enumeration or feature supervision.

With rich node features, Concorde achieves substantial early-retrieval gains: $P@10\%$ of $25.28\%$ on Cora (vs. $13.21\%$ for Gelato, the strongest similarity-based method) and $20.00\%$ on CiteSeer. The invariant anchor stabilizes global citation structure while the feature-to-tangent map resolves unobserved intra-community links. On PubMed, Concorde leads in early precision ($23.69\%$ vs. $22.07\%$) but Gelato retains a marginal AP edge ($2.88\%$ vs. $2.83\%$); PubMed's sparser intra-community con-

Table 1: Link prediction performance comparison across five benchmark datasets. We report Early Precision ($P@10\%$, $P@20\%$) and Average Precision (AP) using an unbiased evaluation protocol. Results show mean $\pm$ standard deviation over 10 runs, except where noted. ($^*$) Run only once due to runtime $>$ 24 hrs; OOM: Out-of-memory; ($^{**}$) Each run takes $>$ 1000 hrs [27]

| Metric | Model | Cora | CiteSeer | PubMed | OGBL-DDI | Power |
|---|---|---|---|---|---|---|
| P@10 | SEAL | 6.52 ± 1.08 | 5.35 ± 1.12 | ** | 37.45 ± 1.15 | 11.34 ± 1.28 |
| | HGCN | 6.53 ± 1.11 | 7.25 ± 1.23 | 11.10 ± 1.25 | 27.32 ± 1.93 | 18.28 ± 1.19 |
| | Neo-GNN | 6.05 ± 0.32 | 6.23 ± 0.85 | 12.32 ± 0.74 | 25.80* | 10.74 ± 0.39 |
| | BUDDY | 7.23 ± 0.45 | 7.93 ± 0.80 | OOM | 22.59 ± 2.02 | 12.53 ± 0.45 |
| | NCN | 10.12 ± 1.08 | 10.36 ± 0.32 | 3.23 ± 1.09 | 45.30 ± 0.15 | 14.77 ± 0.61 |
| | NCNC | 8.78 ± 0.58 | 10.84 ± 0.34 | 14.56 ± 0.89 | 45.22 ± 0.11 | 13.42 ± 0.31 |
| | OCNP | 9.30 ± 0.32 | 12.76 ± 0.27 | 17.38 ± 0.12 | 46.12 ± 0.13 | 20.60 ± 0.60 |
| | Gelato | 13.21 ± 1.33 | 8.70 ± 1.05 | 22.07 ± 0.12 | 33.34 ± 0.11 | 24.85 ± 4.49 |
| | Concorde | **25.28 ± 1.92** | **20.00 ± 1.58** | **23.69 ± 0.15** | **46.45 ± 0.20** | **27.58 ± 2.78** |
| P@20 | SEAL | 5.57± 0.98 | 5.08 ± 0.87 | ** | 36.23 ± 1.30 | 10.12 ± 0.38 |
| | HGCN | 5.48 ± 1.20 | 6.88 ± 0.92 | 10.53 ± 0.97 | 26.60 ± 2.13 | 17.72 ± 0.98 |
| | Neo-GNN | 5.40 ± 0.45 | 6.12 ± 0.37 | 10.54 ± 0.67 | 24.28* | 11.11 ± 0.25 |
| | BUDDY | 6.23 ± 0.71 | 7.15 ± 0.32 | OOM | 20.34 ± 1.15 | 12.38 ± 0.48 |
| | NCN | 9.08 ± 0.46 | 9.71 ± 0.38 | 2.21 ± 1.84 | 44.30 ± 0.08 | 13.77 ± 0.61 |
| | NCNC | 7.34 ± 0.48 | 9.12 ± 0.75 | 11.28 ± 0.82 | 44.85 ± 0.09 | 14.45 ± 0.87 |
| | OCNP | 8.84 ± 0.23 | 10.11 ± 0.43 | 13.21 ± 0.24 | 46.23 ± 0.08 | 19.91 ± 1.15 |
| | Gelato | 15.47 ± 0.52 | 14.75 ± 0.62 | 16.80 ± 0.24 | 32.25 ± 0.07 | 23.48 ± 2.57 |
| | Concorde | **20.00 ± 0.71** | **16.48 ± 0.69** | **18.92 ± 0.12** | **46.48 ± 0.08** | **24.24 ± 1.36** |
| AP | SEAL | 2.27 ± 0.45 | 2.52 ± 0.22 | ** | 35.20 ± 1.06 | 2.39 ± 0.12 |
| | HGCN | 2.53 ± 0.27 | 2.77 ± 0.38 | 1.84 ± 0.21 | 22.28 ± 2.07 | 3.87 ± 0.49 |
| | Neo-GNN | 2.08 ± 1.68 | 1.68 ± 0.09 | 1.31 ± 0.24 | 23.39* | 2.82 ± 0.18 |
| | BUDDY | 1.32 ± 0.40 | 1.90 ± 0.15 | OOM | 18.89 ± 1.09 | 3.14 ± 0.40 |
| | NCN | 1.98 ± 0.34 | 2.85 ± 0.14 | 0.13 ± 0.05 | 44.75 ± 0.07 | 3.38 ± 0.27 |
| | NCNC | 2.54 ± 0.16 | 3.10 ± 0.12 | 1.90 ± 0.02 | 44.89 ± 0.09 | 3.50 ± 0.24 |
| | OCNP | 2.88 ± 0.20 | 3.30 ± 0.15 | 2.21 ± 0.03 | **46.89 ± 0.07** | 3.85 ± 0.37 |
| | Gelato | 3.90 ± 0.03 | 4.55 ± 0.02 | **2.88 ± 0.00** | 29.42 ± 0.00 | 4.33 ± 0.32 |
| | Concorde | **4.50 ± 0.03** | **5.05 ± 0.18** | 2.83 ± 0.02 | 42.32 ± 0.14 | **6.39 ± 0.28** |

nectivity thins the hard negatives in the upper ranks, narrowing the margin where the $N$-pair objective helps.

### 3.2 Ablations

Table 2: Macro-Ablation: Impact of the Tri-Branch Superposition. We report Average Precision (AP) to demonstrate how each structural operator contributes to hard-negative discrimination.

| Architecture Configuration | | | | Average Precision (AP) | | | | |
|---|---|---|---|---|---|---|---|---|
| $S_{\text{anchor}}$ | $S_{\text{res}}$ | $S_{\text{field}}$ | Description | Cora | CiteSeer | PubMed | OGBL-DDI | Power |
| ✓ | ✓ | ✓ | **Concorde** | **4.50 ± 0.03** | **5.05 ± 0.18** | **2.83 ± 0.02** | **42.32 ± 0.14** | **6.39 ± 0.28** |
| ✓ | × | × | The Invariant Anchor Only | 2.45 ± 0.00 | 2.37 ± 0.00 | 2.30 ± 0.00 | 28.70 ± 0.00 | 3.88 ± 0.00 |
| ✓ | ✓ | × | w/o Geodesic Steering | 3.02 ± 0.27 | 4.70 ± 0.12 | 2.58 ± 0.15 | 38.38 ± 0.30 | 4.17 ± 0.59 |
| ✓ | × | ✓ | w/o Mixed Curvature Residual | 2.87 ± 0.08 | 2.59 ± 0.67 | 1.45 ± 0.28 | 38.22 ± 0.46 | 2.72 ± 0.45 |
| × | × | × | Autocovariance | 2.47 ± 0.00 | 2.65 ± 0.00 | 2.50 ± 0.00 | 29.73± 0.00 | 4.12 ± 0.00 |

To validate the spectral decoupling hypothesis, we systematically ablate each branch of the energy model and report AP in Table 2. Restricting the model to the invariant anchor alone ($S_{\text{anchor}}$) causes performance to drop significantly on structurally demanding datasets, falling to $28.70\%$ AP on OGBL-DDI and $3.88\%$ on Power. We also explicitly separate this configuration from the standard Autocovariance (AC) heuristic ($29.73\%$ and $4.12\%$ AP, respectively). While AC's $\Pi$-weighting introduces a degree-based popularity bias that marginally alters baseline performance, neither isolated heuristic can resolve the complex local symmetries required for accurate link retrieval. Although the spectral anchor provides a permutation-invariant foundation, link prediction requires resolving local ambiguities that the invariant anchor cannot address by construction.

Removing the mixed-curvature residual ($S_{\text{res}}$) produces consistent degradation across all datasets, with the largest drops on CiteSeer ($5.05\% \to 2.59\%$ AP) and Power ($6.39\% \to 2.72\%$ AP). On CiteSeer, the homophilic community structure means that the most informative signal for unobserved links lies in fine-grained local geometry, precisely what $S_{\text{res}}$ encodes via geodesic distances on $\mathcal{M} = \mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$. On Power, the drop indicates that even under extreme sparsity, local curvature-aware representations contribute meaningfully when combined with non-local steering.

Ablating the non-local field energy ($S_{\text{field}}$) produces the most pronounced degradation on Power ($6.39\% \rightarrow 4.17\%$ AP) and a notable drop on OGBL-DDI ($42.32\% \rightarrow 38.38\%$ AP). On Power, where mean degree is low and common neighbors are largely absent, diffusion-based signals carry little information about unobserved links. The structural vector field bypasses this limitation by operating directly on node feature representations in the tangent space of $\mathbb{H}^d_c$, identifying geometrically aligned node pairs without requiring topological connectivity. The comparatively smaller drop on attributed graphs (e.g., Cora: $4.50\% \rightarrow 3.02\%$) indicates that $S_{\text{field}}$ is most critical precisely where topology alone is least informative. The per-dataset pattern of branch sensitivity is further corroborated by a post-hoc enrichment analysis (Appendix B), which shows that Branch II's high-rewiring candidate edges are $1.05$–$1.69\times$ enriched among the globally top-$10\%$ predictions on all datasets where removing Branch II causes the largest ablation drop, while Branch III's static edges dominate on Power, the single dataset where Branch III ablation is most harmful.

To rule out that Concorde's improvement stems from the $N$-pair objective rather than the geometry, we train a geometry-stripped model, learned MLP edge-reweighting plus diffusion with manifold distances, ORC, and local statistics zeroed out, under our *exact* loss and protocol (Table 3). This same-loss reweighting recovers much of the gain over vanilla Autocovariance (e.g. OGBL-DDI $29.73 \rightarrow 38.15$) but plateaus below full Concorde everywhere; the geometry contribution is largest where non-Euclidean structure is strongest ($+4.17$ AP on OGBL-DDI, $+1.87$ on Power). Consistent with this, the $S_{\text{anchor}}/S_{\text{res}}$ rank correlation is only moderate (Spearman $\rho = 0.30$–$0.62$, lowest on sparse Power), confirming the branches rank candidates differently rather than acting as a second diffusion heuristic.

Table 3: Same-loss learned reweighting without geometry, versus vanilla Autocovariance and full Concorde (AP). All three use the identical $N$-pair loss and evaluation protocol; only the geometric inputs differ.

| Configuration | Cora | CiteSeer | PubMed | OGBL-DDI | Power |
|---|---|---|---|---|---|
| Vanilla Autocovariance | 2.47 | 2.65 | 2.50 | 29.73 | 4.12 |
| Same-loss learned reweight(no geometry) | 4.16 | 4.52 | 2.47 | 38.15 | 4.52 |
| **Full Concorde** | **4.50** | **5.05** | **2.83** | **42.32** | **6.39** |

Table 4: Micro-Ablation: Components of the Local Geometric Energy ($S_{\text{res}}$). We isolate the impact of manifold choice and intrinsic structural guardrails on ranking quality. We report Average Precision (AP).

| Configuration | Cora | CiteSeer | PubMed | OGBL-DDI | Power |
|---|---|---|---|---|---|
| **Concorde** | **4.50 ± 0.03** | **5.05 ± 0.18** | **2.83 ± 0.02** | **42.32 ± 0.14** | **6.39 ± 0.28** |
| w/o Hyperbolic Geometry ($\mathbb{H}$) | 4.38 ± 0.05 | 4.72 ± 0.05 | 2.75 ± 0.03 | 39.30 ± 0.14 | 5.78 ± 0.30 |
| w/o Spherical Geometry ($\mathbb{S}$) | 4.34 ± 0.07 | 5.01 ± 0.08 | 2.73 ± 0.04 | 38.32 ± 0.20 | 6.28 ± 0.20 |
| w/o Mixed Curvature (Euclidean Only) | 4.30 ± 0.03 | 4.70 ± 0.03 | 2.67 ± 0.03 | 38.52 ± 0.17 | 5.85 ± 0.32 |
| w/o Feature-Aware ORC ($\kappa$) | 4.37 ± 0.03 | 4.80 ± 0.04 | 2.74 ± 0.04 | 39.22 ± 0.20 | 6.27 ± 0.30 |
| w/o Local Neighborhood Geometry | 4.29 ± 0.07 | 4.56 ± 0.70 | 2.55 ± 0.05 | 39.01 ± 0.44 | 4.85 ± 0.18 |

Replacing the $\mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$ product manifold with a Euclidean latent space of equal capacity degrades AP on all five datasets, with the largest drops on OGBL-DDI ($42.32 \rightarrow 38.52$) and Power ($6.39 \rightarrow 5.85$), the networks with the most pronounced non-Euclidean motifs (dense cliques and tree-like branching). Isolating the two components reveals a clean alignment between curvature and topology: on tree-like Power, ablating $\mathbb{H}^{d_h}_{c_h}$ hurts more than ablating $\mathbb{S}^{d_s}_{c_s}$ ($5.78$ vs. $6.28$), while on triadic-closure-dominated OGBL-DDI the reversal holds ($38.32$ vs. $39.30$). This reversal supports the product parameterization: each component encodes geometrically distinct information.

Removing the feature-aware Ollivier–Ricci proxy ($\kappa$) yields a consistent drop, largest on OGBL-DDI ($42.32 \rightarrow 39.22$): $\kappa$ penalizes pairs that are geodesically close but lack coherent local flow, the false-positive regime pure distance cannot detect. Ablating the local neighborhood-geometry statistics produces the largest single-component drop on Power ($6.39 \rightarrow 4.85$), confirming density-based normalization is critical under sparsity.

## 4 Discussion, Limitations, and Future Work

Concorde resolves the tension between global topological stability and local symmetry-breaking by decomposing link likelihood into three structurally decoupled energy branches, each dominant in the regime it targets: the first (anchor) in dense, featureless graphs; the second (residual) in attributed, homophilic graphs; and the third (field) under extreme sparsity. Removing any single branch produces its largest drop on precisely that regime (Section 3.2), confirming that each contributes significantly.

The ablations also quantify how this division of labor shifts with graph structure. The invariant spectral anchor is the single dominant contributor in dense, featureless regimes: on OGBL-DDI it alone recovers $28.70$ AP, roughly $68\%$ of the full model's $42.32$, and on Power its isolated score ($3.88$ AP) is already competitive with the strongest neighborhood-overlap baseline (OCNP, $3.85$). The two geometric branches then supply consistent, complementary gains on top of this foundation, the mixed-curvature residual contributing most on attributed-homophilic graphs and the non-local field most under extreme sparsity, so that on Cora and CiteSeer the full tri-branch model improves early precision by up to $91\%$ over the strongest baseline. This pattern is corroborated branch-by-branch by the rewiring enrichment analysis (Appendix B): Branch II's high-rewiring candidates are over-represented among the top-$10\%$ predictions precisely on the datasets where its ablation hurts most, while Branch III's static edges dominate on Power, the one regime where removing it is most damaging. Ablation and enrichment thus agree that each branch is both causally necessary and responsible for a structurally distinct slice of the ranked list.

A second recurring theme is the deliberate trade-off between early-retrieval precision and global calibration. Because the $N$-pair objective supervises only the relative ordering of top candidates, Concorde can trail a calibration-oriented baseline on aggregate AP (e.g., OCNP on OGBL-DDI, Gelato on PubMed) while still leading at $P@10\%$/$P@20\%$, the operating point that matters when a human or downstream system inspects only a short ranked shortlist. The metric and the loss should therefore be chosen together: where the application instead consumes a calibrated probability over all candidate pairs, a complementary calibration head or a hybrid ranking–classification objective would be a natural addition, which we leave to future work.

Two limitations warrant acknowledgment. First, the ORC proxy's Sinkhorn iterations, though precomputed and parallelizable, add overhead beyond OGBL-DDI scale; its hyperparameters ($\alpha$, $\varepsilon$, ground-metric mix) are held at standard defaults rather than tuned per dataset (Appendix E), and a feature-driven curvature estimator would remove the overhead. Second, the global scalar gates weight branches uniformly across the graph, suboptimal under pronounced structural heterogeneity; per-subgraph adaptive gating, dynamic-graph extensions via incremental anchor/curvature updates, and cross-graph (out-of-distribution) transfer, which our single-graph protocol does not probe, are natural next steps.

More broadly, our results suggest that geometry should follow the data, not the other way around: matching inductive biases to the structural regime of a graph, rather than relying on a single learned representation space, is a principled and effective strategy for retrieval-style link prediction, and one we expect to generalize across graph learning.

**Funding Statement:** This material is based upon work supported by the National Science Foundation under grant no. IIS-2229876 and is supported in part by funds provided by the National Science Foundation, by the Department of Homeland Security, and by IBM. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or its federal agency and industry partners. The authors declare no competing interests.

## References

[1] Lada A Adamic and Eytan Adar. Friends and neighbors on the web. *Social Networks*, 25(3): 211–230, 2003. ISSN 0378-8733. doi: https://doi.org/10.1016/S0378-8733(03)00009-1. URL `https://www.sciencedirect.com/science/article/pii/S0378873303000091`.

[2] Gregor Bachmann, Gary Bécigneul, and Octavian-Eugen Ganea. Constant curvature graph convolutional networks, 2020. URL `https://arxiv.org/abs/1911.05076`.

[3] Qijie Bai, Changli Nie, Haiwei Zhang, Dongming Zhao, and Xiaojie Yuan. Hgwavenet: A hyperbolic graph neural network for temporal link prediction. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 523–532. ACM, April 2023. doi: 10.1145/3543507.3583455. URL http://dx.doi.org/10.1145/3543507.3583455.

[4] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017. ISSN 1558-0792. doi: 10.1109/msp.2017.2693418. URL http://dx.doi.org/10.1109/MSP.2017.2693418.

[5] Benjamin Paul Chamberlain, Sergey Shirobokov, Emanuele Rossi, Fabrizio Frasca, Thomas Markovich, Nils Hammerla, Michael M. Bronstein, and Max Hansmire. Graph neural networks for link prediction with subgraph sketching, 2023. URL https://arxiv.org/abs/2209.15486.

[6] Ines Chami, Rex Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks, 2019. URL https://arxiv.org/abs/1910.12933.

[7] Yu Chen, Lingfei Wu, and Mohammed J. Zaki. Iterative deep graph learning for graph neural networks: Better and robust node embeddings, 2020. URL https://arxiv.org/abs/2006.13009.

[8] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

[9] Chandler Davis and William M Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[10] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs, 2021. URL https://arxiv.org/abs/2012.09699.

[11] Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning discrete structures for graph neural networks, 2020. URL https://arxiv.org/abs/1903.11960.

[12] Lise Getoor Galileo Mark Namata, Ben London and Bert Huang. Query-driven active surveying for collective classification. In *International Workshop on Mining and Learning with Graphs*, Edinburgh, Scotland, 2012.

[13] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*, volume 207 of *Graduate Texts in Mathematics*. Springer, 2001.

[14] Karish Grover, Haiyang Yu, Xiang song, Qi Zhu, Han Xie, Vassilis N. Ioannidis, and Christos Faloutsos. Spectro-riemannian graph neural networks. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=2MLvV7fvAz.

[15] Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HJxeWnCcF7.

[16] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. Hyperbolic attention networks, 2018. URL https://arxiv.org/abs/1805.09786.

[17] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, 2008.

[18] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[19] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *arXiv preprint arXiv:2005.00687*, 2020.

[20] Zexi Huang, Arlei Silva, and Ambuj Singh. A broader picture of random-walk based graph embedding. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 685–695, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467300. URL https://doi.org/10.1145/3447548.3467300.

[21] Glen Jeh and Jennifer Widom. Simrank: a measure of structural-context similarity. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, page 538–543, New York, NY, USA, 2002. Association for Computing Machinery. ISBN 158113567X. doi: 10.1145/775047.775126. URL https://doi.org/10.1145/775047.775126.

[22] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, Mar 1953. ISSN 1860-0980. doi: 10.1007/BF02289026.

[23] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

[24] Andreas Loukas. Graph reduction with spectral and cut guarantees, 2018. URL https://arxiv.org/abs/1808.10650.

[25] Haitao Mao, Juanhui Li, Harry Shomer, Bingheng Li, Wenqi Fan, Yao Ma, Tong Zhao, Neil Shah, and Jiliang Tang. Revisiting link prediction: A data perspective, 2024. URL https://arxiv.org/abs/2310.00793.

[26] Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4363–4371. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/maron19a.html.

[27] Joao Mattos, Zexi Huang, Mert Kosan, Ambuj Singh, and Arlei Silva. Attribute-enhanced similarity ranking for sparse link prediction. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, page 1044–1055, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712456. doi: 10.1145/3690624.3709314. URL https://doi.org/10.1145/3690624.3709314.

[28] Gal Mishne, Zhengchao Wan, Yusu Wang, and Sheng Yang. The numerical stability of hyperbolic representation learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 24925–24949. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/mishne23a.html.

[29] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2), July 2001. ISSN 1095-3787. doi: 10.1103/physreve.64.025102. URL http://dx.doi.org/10.1103/PhysRevE.64.025102.

[30] Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. Community detection on networks with ricci flow, 2019. URL https://arxiv.org/abs/1907.03993.

[31] Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations, 2017. URL https://arxiv.org/abs/1705.08039.

[32] Yann Ollivier. Ricci curvature of markov chains on metric spaces, 2007. URL https://arxiv.org/abs/math/0701886.

[33] Mustafa Bilgic Lise Getoor Brian Gallagher Prithviraj Sen, Galileo Mark Namata and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[34] Christopher De Sa, Albert Gu, Christopher Ré, and Frederic Sala. Representation tradeoffs for hyperbolic embeddings, 2018. URL https://arxiv.org/abs/1804.03329.

[35] Cedric Sanders, Andreas Roth, and Thomas Liebig. Curvature-based pooling within graph neural networks, 2023. URL https://arxiv.org/abs/2308.16516.

[36] Purnamrita Sarkar and Andrew W. Moore. *Random Walks in Social Networks and their Applications: A Survey*, pages 43–77. Springer US, Boston, MA, 2011. ISBN 978-1-4419-8462-3. doi: 10.1007/978-1-4419-8462-3_3. URL https://doi.org/10.1007/978-1-4419-8462-3_3.

[37] Cong Shen, Pingjian Ding, Junjie Wee, Jialin Bi, Jiawei Luo, and Kelin Xia. Curvature-enhanced graph convolutional network for biomolecular interaction prediction, 2023. URL https://arxiv.org/abs/2306.13699.

[38] Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. Mixed-curvature variational autoencoders, 2020. URL https://arxiv.org/abs/1911.08411.

[39] Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9448–9457. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/teru20a.html.

[40] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations (ICLR)*, 2022.

[41] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018. URL https://arxiv.org/abs/1710.10903.

[42] Juntong Wang, Xiyuan Wang, and Muhan Zhang. OCN: Effectively utilizing higher-order common neighbors for better link prediction. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=E3bMQCKgiX.

[43] Xiyuan Wang, Haotong Yang, and Muhan Zhang. Neural common neighbor with completion for link prediction. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=sNFLN3itAd.

[44] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. doi: 10.1038/30918.

[45] Bo Xiong, Shichao Zhu, Nico Potyka, Shirui Pan, Chuan Zhou, and Steffen Staab. Pseudo-riemannian graph convolutional networks. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=KeIuNChob1H.

[46] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019. URL https://arxiv.org/abs/1810.00826.

[47] Ze Ye, Kin Sum Liu, Tengfei Ma, Jie Gao, and Chao Chen. Curvature graph network. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BylEqnVFDB.

[48] Seongjun Yun, Seoyoon Kim, Junhyun Lee, Jaewoo Kang, and Hyunwoo J. Kim. Neo-gnns: Neighborhood overlap-aware graph neural networks for link prediction, 2022. URL https://arxiv.org/abs/2206.04216.

[49] Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. *CoRR*, abs/1802.09691, 2018. URL http://arxiv.org/abs/1802.09691.

[50] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, October 2009. ISSN 1434-6036. doi: 10.1140/epjb/e2009-00335-8. URL http://dx.doi.org/10.1140/EPJB/E2009-00335-8.

[51] Dongmian Zou and Gilad Lerman. Graph convolutional neural networks via scattering. *Applied and Computational Harmonic Analysis*, 49(3):1046–1074, November 2020. ISSN 1063-5203. doi: 10.1016/j.acha.2019.06.003. URL http://dx.doi.org/10.1016/j.acha.2019.06.003.

# A Theoretical Analysis of the Spectral Decoupling Hypothesis

We provide formal justification for the spectral decoupling hypothesis, establishing that (i) a single-matrix scheme faces an unavoidable trade-off between global invariance and local equivariance, and (ii) the Concorde anchor preserves the graph spectrum by construction.

## A.1 Impossibility of Joint Invariance-Equivariance in a Single Matrix

**Proposition 1** (Invariance–Equivariance Trade-off)**.** *Let $G = (V, E)$ be a graph with adjacency matrix $A$ and automorphism group $\mathrm{Aut}(G)$, and let $A_{\mathrm{learn}} = A + \Delta$ with $\Delta \neq 0$. Consider the two properties:*

1. ***Spectral preservation:*** *$A_{\mathrm{learn}}$ is a* spectral filter *of $A$, i.e. $A_{\mathrm{learn}} = \sum_k c_k A^k$ for some coefficients $\{c_k\}$. Equivalently, $A_{\mathrm{learn}}$ is diagonal in every eigenbasis of $A$ and shares its eigenspaces. Every diffusion operator, including the anchor's $M^k$, is of this form.*
2. ***Symmetry breaking:*** *$A_{\mathrm{learn}}$ assigns distinct scores to some pair of nodes in a common $\mathrm{Aut}(G)$-orbit, i.e. $[A_{\mathrm{learn}}]_{\pi(u)\pi(v)} \neq [A_{\mathrm{learn}}]_{uv}$ for some $\pi \in \mathrm{Aut}(G)$ and some $u, v$.*

*No single matrix $A_{\mathrm{learn}}$ can satisfy both: property (1) implies the negation of property (2). Consequently, breaking orbit symmetry forces $A_{\mathrm{learn}}$ outside the functional calculus of $A$, so it no longer shares $A$'s eigenspaces and its action distorts the graph spectrum.*

*Proof.* Assume property (1). For every $\pi \in \mathrm{Aut}(G)$, the corresponding permutation matrix $\Pi$ satisfies $\Pi A \Pi^\top = A$, so $\Pi$ commutes with $A$ and therefore with every power $A^k$ (as $\Pi$ is orthogonal, $\Pi A^k \Pi^\top = (\Pi A \Pi^\top)^k = A^k$). Hence

$$\Pi A_{\mathrm{learn}} \Pi^\top = \sum_k c_k \, \Pi A^k \Pi^\top = \sum_k c_k \, A^k = A_{\mathrm{learn}}.$$

Reading the identity $\Pi A_{\mathrm{learn}} \Pi^\top = A_{\mathrm{learn}}$ entrywise yields $[A_{\mathrm{learn}}]_{\pi(u)\pi(v)} = [A_{\mathrm{learn}}]_{uv}$ for all $u, v \in V$ and all $\pi \in \mathrm{Aut}(G)$. Thus any two node pairs in the same $\mathrm{Aut}(G)$-orbit receive identical scores, which is exactly the negation of property (2). The final claim is the contrapositive: if $A_{\mathrm{learn}}$ breaks orbit symmetry, it is not a spectral filter of $A$ and hence does not share its eigenspaces. □

**Remark (general equivariant operators).** The obstruction is not special to polynomials in $A$. For *any* $\mathrm{Aut}(G)$-equivariant $A_{\mathrm{learn}}$, orbit-invariance of the scores is precisely the entrywise reading of $\Pi A_{\mathrm{learn}} \Pi^\top = A_{\mathrm{learn}}$. Conversely, on a degenerate eigenspace $V_\lambda$ of $A$, an operator that both preserves $V_\lambda$ (spectral preservation) and is equivariant acts on each irreducible $\mathrm{Aut}(G)$-subrepresentation of $V_\lambda$ as a scalar multiple of the identity, by Schur's lemma [13]. For example, on $K_{2,2}$ the 2-dimensional nullspace carries an irreducible action of $\mathrm{Aut}(K_{2,2}) \cong D_4$, so any spectrum-preserving operator is scalar there and cannot distinguish the two nodes of a bipartite orbit. Escaping the trade-off therefore requires leaving the spectral class altogether, which is exactly what Concorde's feature-driven residual on $A_{\mathrm{enh}}$ does.

**Remark.** This result formalizes why single-matrix learned approaches must sacrifice either global stability or local discriminability. Concorde resolves this by *separating* the two objectives across dedicated branches rather than conflating them in one matrix.

**Remark (connection to GNN expressivity).** $\mathrm{Aut}(G)$-equivariance is precisely the standard invariance condition studied in the GNN expressivity literature [26, 46]: any message-passing GNN satisfying this condition cannot distinguish nodes in the same orbit, which is the same limitation formalized here.

## A.2 Spectrum-Preservation Guarantee of the Invariant Anchor

**Proposition 2** (Spectral Stability of $S_{\mathrm{anchor}}$)**.** *Let $M = D^{-1}A$ be the row-stochastic transition matrix of a connected aperiodic graph $G$, and let $\pi$ be its stationary distribution. The anchor score matrix $S_{\mathrm{anchor}} = M^k - \mathbf{1}\pi^\top$ satisfies:*

1. ***Permutation Invariance:*** *For any $\Pi \in \mathrm{Aut}(G)$, $\Pi S_{\mathrm{anchor}} \Pi^\top = S_{\mathrm{anchor}}$.*

2. ***No Eigenvalue Growth:*** *The spectral radius of $M$ satisfies $\rho(M) = 1$, so no eigenvalue of $M^k$ exceeds 1 in magnitude, and $S_{\text{anchor}}$ introduces no spectral amplification.*

3. ***Spectral Contraction:*** *The non-trivial eigenvalues of $M^k$ satisfy $|\mu_i^k| \leq |\mu_2|^k$, where $|\mu_2| < 1$ is the spectral gap. Thus $S_{\text{anchor}}$ exponentially suppresses non-dominant spectral components as $k$ grows, converging toward the rank-1 matrix $\mathbf{1}\pi^\top$.*

*Proof.* **(1)** Since $\Pi A \Pi^\top = A$ and $\Pi D \Pi^\top = D$ (degree is preserved by automorphisms), we have $\Pi M \Pi^\top = M$. By induction, $\Pi M^k \Pi^\top = M^k$. The stationary distribution satisfies $\pi_{\pi(u)} = \pi_u$ for all automorphisms, so $\Pi \mathbf{1}\pi^\top \Pi^\top = \mathbf{1}\pi^\top$. Combining: $\Pi S_{\text{anchor}} \Pi^\top = S_{\text{anchor}}$.

**(2)** Since $M$ is row-stochastic, $M\mathbf{1} = \mathbf{1}$, so 1 is an eigenvalue with right eigenvector $\mathbf{1}$. The $\ell^\infty$-norm of any row of $M$ equals 1, so $\rho(M) \leq \|M\|_\infty = 1$ [17]. Together: $\rho(M) = 1$, and all eigenvalues $\mu_i$ satisfy $|\mu_i| \leq 1$. Consequently $|\mu_i^k| \leq 1$ for all $i$ and $k \geq 1$, so $M^k$ introduces no eigenvalue growth relative to $M$.

**(3)** For a connected aperiodic graph, the Perron–Frobenius theorem guarantees $\mu_1 = 1$ is the unique eigenvalue of maximum modulus [23]. The corresponding left eigenvector is $\pi^\top$ and right eigenvector is $\mathbf{1}$, so the rank-1 term $\mathbf{1}\pi^\top$ exactly cancels the leading component of $M^k$. The remaining eigenvalues satisfy $|\mu_i| \leq |\mu_2| < 1$, so $\|S_{\text{anchor}}\|_F \to 0$ as $k \to \infty$, with rate $|\mu_2|^k$. □

### A.3 Equivariance of the Geometric Residual

The geometric residual $S_{\text{res}}$ operates on the enhanced adjacency $A_{\text{enh}} = A \odot W_{\text{geo}}$, where $W_{\text{geo}}$ is produced by the learned modulator $\mathcal{F}_\theta$. Unlike a single-matrix scheme, $W_{\text{geo}}$ intentionally breaks $\text{Aut}(G)$-invariance via the feature-aware ORC proxy $\kappa(u, v)$ and geodesic distances on $\mathcal{M}$. The Hadamard product with the fixed sparsity pattern of $A$ ensures:

$$\text{supp}(A_{\text{enh}}) \subseteq \text{supp}(A), \quad \|A_{\text{enh}}\|_0 = |E|. \tag{15}$$

This prevents $S_{\text{res}}$ from introducing new non-edges into the transition structure, bounding its capacity to distort the global topology. Crucially, $S_{\text{res}}$ is computed independently of $S_{\text{anchor}}$: the anchor's spectrum is never modified by the learned branch, preserving all guarantees of Proposition 2.

The three results above collectively justify the spectral decoupling design: (i) a single learned matrix *cannot* simultaneously achieve global invariance and local symmetry-breaking (Proposition 1); (ii) the parameter-free anchor has spectral radius 1, is permutation-invariant, and contracts non-dominant modes at rate $|\mu_2|^k$ (Proposition 2); and (iii) the learned residual operates within the sparsity envelope of $A$, insulating the anchor's spectral properties from gradient updates.

### A.4 Empirical Validation: The Coupled Single-Matrix Model

Proposition 1 predicts that folding the learned dynamics into a *single* coupled adjacency matrix, rather than decoupling them across branches, must either distort the graph spectrum or fail to break orbit symmetry. We test this directly. We construct a genuine coupled single-matrix model that learns one shared enhanced adjacency $A_{\text{learn}} = A + \Delta$ (using Concorde's own reweighting), diffuses on it *alone*, and decodes, i.e. precisely the regime Proposition 1 argues against. We compare against the full decoupled Concorde and additionally report the *spectral drift*, the relative eigenvalue deviation of the coupled matrix, $\|\lambda(A_{\text{enh}}) - \lambda(A)\| / \|\lambda(A)\|$.

The coupled model underperforms decoupled Concorde on **all five** datasets (Table 5), with the largest loss on sparse featureless Power ($-1.85$) and the smallest on feature-rich Cora ($-0.23$). This is exactly the pattern Proposition 1 predicts: node features partially break $\text{Aut}(G)$-symmetry, so the structural impossibility is most binding where features are weakest. The spectral drift is correspondingly largest on Power (0.27), the same graph with the largest accuracy penalty.

**How the decoupled design escapes Proposition 1.** The anchor score $S_{\text{anchor}}$ is computed on the *unperturbed* base adjacency $A$ and added *separately* to the residual $S_{\text{res}}$; the learned reweighting acts on the enhanced adjacency $A_{\text{enh}}$ only, never on the anchor's matrix. The anchor's spectrum is therefore never modified by the learned branch (drift $\approx 0$ by construction), while a single coupled matrix must distort the spectrum to break symmetry (drift up to 0.27 above). Decoupling is precisely what lets Concorde break *local* symmetry ($S_{\text{res}}$ on $A_{\text{enh}}$) while preserving the *global* spectrum ($S_{\text{anchor}}$ on $A$), the two objectives Proposition 1 shows cannot coexist in one matrix.

Table 5: Coupled single-matrix model versus full decoupled Concorde (AP), with the relative spectral drift of the coupled learned matrix. The coupled model underperforms on all five datasets, most on sparse featureless Power and least on feature-rich Cora, matching the prediction of Proposition 1 that node features partially relax the structural impossibility.

| Dataset | Single-matrix AP | Full AP | $\Delta$ | Spectral drift |
|---|---|---|---|---|
| Power | 4.54 | 6.39 | $\mathbf{-1.85}$ | $\mathbf{0.27}$ |
| OGBL-DDI | 39.80 | 42.32 | $-2.52$ | 0.03 |
| CiteSeer | 4.63 | 5.05 | $-0.42$ | 0.10 |
| Cora | 4.27 | 4.50 | $-0.23$ | 0.12 |
| PubMed | 2.72 | 2.83 | $-0.11$ | 0.06 |

# B Branch Contribution Analysis via Geometric Rewiring

To empirically isolate the contribution of each structural branch, we perform a post-hoc rewiring enrichment analysis across all five benchmarks. For each dataset, we quantify the *rewiring magnitude* of a node as the $\ell_1$ deviation between the base adjacency $A$ and the Branch II enhanced adjacency $A_{\text{enh}}$:

$$\rho(u) = \|[A_{\text{enh}}]_{u,:} - A_{u,:}\|_1$$

Each candidate pair $(u, v)$ is assigned an edge-level rewiring score $\rho(u) + \rho(v)$ and labeled as *high-rewiring* (above median) or *static* (below median). We then measure whether high-rewiring edges are over-represented among Concorde's globally top-$10\%$ ranked predictions — the same population as $P@10\%$, via an enrichment ratio relative to the null expectation of $50\%$, with statistical significance assessed by a two-sided Fisher exact test.

Table 6 summarizes the enrichment ratios across all datasets. On attributed graphs (Cora, CiteSeer, PubMed) and on the dense featureless network OGBL-DDI, high-rewiring edges are consistently over-represented in the top-$10\%$ predictions ($1.05$–$1.69\times$, $p < 10^{-4}$ in all cases), indicating that Branch II's mixed-curvature residual actively promotes structurally coherent candidates when node features or dense local structure are available. The effect is strongest on OGBL-DDI ($1.69\times$), where rich triadic-closure structure provides direct signal for the Hadamard reweighting to exploit. On Power, the pattern reverses: static edges are marginally over-represented ($1.10\times$), consistent with the expectation that Branch II's reweighting of *observed* edges cannot reach the topologically disconnected pairs that dominate this extremely sparse graph. In this regime, Branch III's non-local hyperbolic steering field bridges these structural gaps without requiring any path-based connectivity. HGCN, despite operating in a hyperbolic manifold well-suited to tree-like structure, achieves only $18.28\%$ $P@10\%$ on Power, above neighborhood-overlap baselines but $6.57\%$ below Concorde, confirming that curvature alone is insufficient without the non-local field branch to bridge topological gaps.

This enrichment pattern directly corroborates the macro-ablation of Section 3.2. There, removing Branch II causes the largest performance drops on Cora, CiteSeer, PubMed, and OGBL-DDI, while removing Branch III produces the steepest degradation on Power, precisely the datasets where the corresponding branch's edges are enriched in the top-$10\%$ here. The two analyses thus provide complementary evidence: the ablation establishes that each branch is *causally necessary*, while the enrichment analysis reveals *which candidate pairs* each branch is responsible for promoting. Together, they validate the spectral decoupling hypothesis: each branch captures structurally distinct signal, and their relative contributions shift predictably with graph density and feature availability.

# C Scalability Analysis

A central concern in geometric and diffusion-based graph learning is scalability to massive networks, such as the $\sim$1.3 million edges in OGBL-DDI. Concorde is specifically designed to avoid $O(N^2)$ bottlenecks.

**Diffusion Operator ($M^k$):** While the paper formulates the invariant anchor $S_{\text{anchor}}$ and the residual $S_{\text{res}}$ using the dense matrix notation $M^k$, we do not materialize this dense matrix globally in practice. For training and inference, the diffusion scores for a specific candidate pair $(u, v)$ are computed on-the-fly using sparse matrix-vector multiplication (SpMV). Given a sparse transition matrix $M$, we

Table 6: Branch enrichment in globally top-10% ranked predictions. Enrichment $> 1$ indicates over-representation of that branch's candidate edges at the top of the ranked list. All results significant at $p < 10^{-4}$ (Fisher exact test).

| Dataset | Enrichment (Br. II) | Enrichment (Br. III) | Dominant Branch |
|---|---|---|---|
| Cora | $1.05\times$ | $0.95\times$ | Branch II |
| CiteSeer | $1.08\times$ | $0.92\times$ | Branch II |
| PubMed | $1.45\times$ | $0.55\times$ | Branch II |
| OGBL-DDI | $1.69\times$ | $0.31\times$ | Branch II |
| Power | $0.90\times$ | $1.10\times$ | Branch III |

iteratively compute the reachability vector $p_u^{(k)} = M^k e_u$, which requires strictly $O(k|E|)$ operations, keeping the memory footprint minimal.

**Feature-Aware ORC ($\kappa$) and Local Geometry:** Computing the Wasserstein distance via the entropy-regularized Sinkhorn algorithm is computationally intensive. However, the geometric modulator $W_{\text{geo}}$ and the intrinsic curvature proxy $\kappa(u, v)$ are *not* computed for all $N^2$ pairs. They are evaluated exclusively on: (1) the observed edge set $E$ to construct the sparse enhanced adjacency $A_{\text{enh}}$, and (2) the specific batch of candidate pairs $B$ sampled during training or evaluation. Thus, the complexity scales as $O(|E| + |B|)$, completely bypassing the quadratic bottleneck.

**Exact $N$-Pair Loss on Massive Graphs:** A naive implementation of the $N$-pair ranking loss over the exhaustive complement of non-edges requires $O(|V|^2)$ memory, which is computationally infeasible for large graphs. Concorde resolves this without resorting to stochastic subsampling through two mechanisms:

1. **Graph Partitioning (Training):** For highly dense or massive graphs, the training topology is partitioned into localized subgraphs using METIS clustering. The exhaustive negative mask is computed densely *only within each partition*, reducing the localized memory footprint to $O(|V_c|^2)$ where $|V_c| \ll |V|$ is the cluster size. Inter-cluster pairs are extracted and processed sequentially.

2. **Chunked Evaluation (Inference):** During full-pool evaluation, pairwise energy scores are not materialized as a single dense $N \times N$ tensor. Instead, candidate pair distances on the product manifold and tangent spaces are computed in sequential chunks of $50{,}000$ pairs. This guarantees that peak GPU memory usage remains constant regardless of the global graph size, allowing Concorde to execute an exact, full-negative ranking evaluation.

### C.1 Empirical Runtimes and Comparison to Baselines

To empirically validate the scalability of the proposed framework, we record the wall-clock runtimes for both the precomputation of the feature-aware Ollivier-Ricci curvature (ORC) proxy and the per-epoch training time. All timing benchmarks were executed on a single Nvidia H200 GPU instance (`h200-3g.71gb:1`). The precomputation of the ORC proxy introduces minimal overhead, completing in just over a minute even for the dense OGBL-DDI graph (Table 7); it is a *one-time* cost, not a per-epoch one.

A key point is that per-epoch time scales with node count times diffusion steps $k$, *not* edge count. This is why PubMed (19.7k nodes, $k = 3$) is heaviest, on par with Gelato, while OGBL-DDI (4.3k nodes, $k = 1$) is fast despite its $\sim$1.3M edges. Tables 8 and 9 compare Concorde against the strongest baselines on identical hardware under the same unbiased full-negative protocol, all measured by us. On the densest graph (OGBL-DDI), Concorde is 4–5$\times$ faster than the overlap models that are its closest competitors there (7.2s vs. OCNP 28.1s and NCNC 36.2s per epoch). BUDDY, the sketching method cited as scalable, is 40–90$\times$ slower here, out-of-memories on PubMed, and requires 55.4/37.3 GB where Concorde stays under 14 GB, because it must precompute and store subgraph-sketch structural features for every candidate pair, and under the full-negative protocol that candidate set is the entire $O(N^2)$ non-edge complement. Peak memory is otherwise modest and comparable to or below Gelato.

Table 7: Wall-clock runtime for ORC proxy precomputation and average per-epoch training time for Concorde.

| Dataset | ORC Precomputation (s) | Time per Epoch (s) |
|---|---|---|
| Cora | 3.2 | 2.4 |
| CiteSeer | 4.7 | 3.2 |
| PubMed | 32.4 | 54.1 |
| Power | 12.2 | 5.8 |
| OGBL-DDI | 75.9 | 7.2 |

Table 8: Per-epoch training time (s) across methods, identical hardware and full-negative protocol.

| s/epoch | Concorde | NCNC | OCNP | BUDDY | Gelato |
|---|---|---|---|---|---|
| Cora | 2.4 | 4.8 | 3.2 | 0.6 | 3.3 |
| CiteSeer | 3.2 | 8.3 | 4.8 | 485 | 2.9 |
| PubMed | 54.1 | 6.2 | 4.4 | OOM | 49.8 |
| OGBL-DDI | **7.2** | 36.2 | 28.1 | 342 | 1.4 |
| Power | 5.8 | 7.6 | 6.9 | 498 | 7.2 |

# D Evaluation Protocol

We describe the dataset splitting, candidate pool construction, and metric computation used throughout all experiments. The protocol follows the unbiased full-negative convention of Mattos et al. [27], evaluating every method against the exhaustive set of non-edges rather than a sampled subset.

For Cora, CiteSeer, PubMed, edges are split deterministically using a fixed random seed ($s = 0$) via `train_test_split_edges` with validation ratio $5\%$ and test ratio $10\%$. The remaining $85\%$ of edges form the training graph $G_{\text{train}} = (V, E_{\text{train}})$ whose adjacency matrix $A$ is used by all three branches of Concorde. For Power, we load pre-saved official splits and do not re-split. For OGBL-DDI, we use the OGB dataset loader and its prescribed splits directly.

Let $\mathcal{U}$ denote all upper-triangular node pairs $(u, v)$ with $u < v$. The negative pool at each stage is constructed exhaustively with no subsampling:

- **Training negatives** $\mathcal{E}^-_{\text{train}}$: all pairs in $\mathcal{U} \setminus E_{\text{train}}$. This full complement is used in the $N$-pair ranking loss (Eq. 8).
- **Validation pool**: $E_{\text{val}}$ evaluated against $\mathcal{E}^-_{\text{train}} \setminus (E_{\text{val}} \cup E_{\text{test}})$.
- **Test pool**: $E_{\text{test}}$ evaluated against $\mathcal{E}^-_{\text{train}} \setminus (E_{\text{val}} \cup E_{\text{test}})$.

No ground-truth positive edge ever appears in the negative pool at any evaluation stage, eliminating false-negative contamination. To obtain unbiased precision estimates under score ties, the combined pool (positives $\cup$ negatives) is randomly shuffled with a fixed seed before evaluation; ties in predicted scores are therefore broken uniformly at random rather than by index order.

During each mini-batch, the positive edges $\mathcal{E}^+_{\text{batch}}$ are zeroed out of $A$ via an index-put operation before computing $S_{\text{anchor}}$ and $S_{\text{res}}$, so the diffusion-based branches infer link likelihood only from the remaining topological context; the mask is released after each forward pass. This masking is *exact* for the diffusion operators. On featureless graphs, the Laplacian positional encodings that serve as node features are precomputed once on $G_{\text{train}}$ and are *not* re-masked per batch, so a batch's own positive edge remains encoded in its endpoints' LPEs during training. We bound this residual leakage below via a Davis–Kahan argument and verify it is empirically negligible in Appendix F (training-only vs. full-graph LPEs differ by $\leq 0.44$ AP, far below Concorde's margins over baselines).

All metrics are computed in a single forward pass over the full test pool with `torch.no_grad()`. Let $\mathcal{C}_{\text{test}}$ denote the test pool, $n^+ = |E_{\text{test}}|$ the number of positive pairs, and $\hat{s}$ the vector of predicted scores.

- **Average Precision (AP)**: the area under the precision–recall curve over all of $\mathcal{C}_{\text{test}}$, computed via `sklearn.metrics.average_precision_score`.

Table 9: Peak GPU memory (GB) across methods, identical hardware and full-negative protocol.

| peak GB | Concorde | NCNC | OCNP | BUDDY | Gelato |
|---|---|---|---|---|---|
| Cora | 2.45 | 1.85 | 1.23 | 1.62 | 1.83 |
| CiteSeer | 2.14 | 2.93 | 1.32 | 12.84 | 1.18 |
| PubMed | 12.84 | 4.15 | 3.77 | OOM | 18.12 |
| OGBL-DDI | 13.89 | 6.82 | 5.56 | 55.39 | 7.49 |
| Power | 3.05 | 2.84 | 1.42 | 37.28 | 3.08 |

- **Early Precision** $P@r\%$: let $k_r = \lceil n^+ \cdot (r/100) \rceil$. We retrieve the top-$k_r$ candidates by score and compute:

$$P@r\% = \frac{1}{k_r} \sum_{i \in \text{top-}k_r} \mathbf{1}[y_i = 1]. \tag{16}$$

  We report $r \in \{10, 20\}$ as primary metrics. Critically, $k_r$ is defined relative to $n^+$, not the pool size, so the denominator is identical across datasets and methods given the same positive set.

**Choice of metrics.** We report AP and early precision rather than MRR/Hits@$K$ because our target is retrieval under *extreme* class imbalance, where the quantity of interest is the density of true edges at the very top of a single global ranking over the entire non-edge complement, exactly what $P@r\%$ measures with a denominator held identical across methods. This regime is strictly harder than the sampled-negative protocols that MRR/Hits@$K$ are usually computed under: our exhaustive full-negative evaluation already imposes the hard-negative pressure that recent per-source protocols (e.g., HeaRT) approximate, since every non-edge competes with every positive rather than a sampled shortlist. AP additionally captures global calibration, and we report it throughout (including where Concorde trades AP for early precision), so no ranking behavior is hidden by the metric choice.

Model selection is performed by monitoring $P@100\%$ on the validation pool after each epoch (equivalent to precision over exactly $n^+_{\text{val}}$ top-ranked candidates). The checkpoint achieving the highest validation precision is used for all reported test metrics.

OGBL-DDI and Power carry no node attributes. For both, we substitute absent features with Laplacian Positional Encodings (LPEs): the $k$ non-trivial eigenvectors of the normalized graph Laplacian $\tilde{L} = I - D^{-1/2}AD^{-1/2}$, computed via dense `scipy.linalg.eigh` and row-$\ell_2$-normalized. We use $k = 128$ for OGBL-DDI and $k = 32$ for Power. LPEs are precomputed once on the full training graph topology $G_{\text{train}}$ and fixed throughout training. While fixing the LPEs introduces a mild structural approximation, we argue the induced leakage is negligible in practice. By the Davis–Kahan theorem, the perturbation to eigenvector $\phi_i$ of the normalized Laplacian from removing a single edge $(u, v)$ is bounded by $\|\delta\phi_i\| \leq \|\delta\tilde{L}\|/\delta_i$, where $\|\delta\tilde{L}\|$ is $O(1/\sqrt{d_u d_v})$ for the normalized Laplacian and $\delta_i$ is the corresponding eigengap. For graphs of moderate density, this perturbation is negligible: the LPEs computed on the unmasked training graph and on any single-edge-masked subgraph are effectively identical. The per-batch adjacency masking of target edges from $A$ remains exact for the diffusion operators (the primary anti-leakage mechanism), which is unaffected by this LPE approximation.

All runs are seeded via a unified call setting `torch`, `numpy`, and `random` seeds simultaneously. Each dataset is run for the number of independent seeds reported in Table 1; mean and standard deviation are computed across those runs.

# E Implementation Details and Hyperparameters

Table 10 summarises all dataset-specific hyperparameters. All models are trained with Adam. Unless otherwise stated, the tri-branch architecture uses $k = 3$ diffusion steps, ORC retention mass $\alpha = 0.5$, Sinkhorn regularisation $\varepsilon = 0.05$ with 30 iterations, and a maximum neighbourhood degree of 32 for ORC computation.

The mixed-curvature product manifold $\mathbb{H}^{d_h}_{c_h} \times \mathbb{S}^{d_s}_{c_s}$ allocates equal dimensionality to both components. We use $d_h = d_s = 16$ for Cora and CiteSeer (smaller graphs where over-parameterisation hurts) and $d_h = d_s = 32$ for PubMed, OGBL-DDI, and Power. Curvature scalars $c_h$ and $c_s$ are initialised at $\approx 0.694$ (via $\log c = 0$) and optimised end-to-end via $c = \text{softplus}(\log c) + 10^{-3}$.

Table 10: Dataset-specific hyperparameter configurations for Concorde.

| Dataset | $d_h{=}d_s$ | Hidden | LR | WD | $k$ | Scale | Self-loops | Epochs |
|---|---|---|---|---|---|---|---|---|
| Cora | 16 | 64 | $10^{-3}$ | $5{\times}10^{-4}$ | 3 | 3 | ✓ | 100 |
| CiteSeer | 16 | 64 | $10^{-4}$ | $5{\times}10^{-4}$ | 3 | 3 | ✓ | 100 |
| PubMed | 32 | 128 | $10^{-3}$ | 0 | 3 | – | × | 100 |
| OGBL-DDI | 32 | 512 | $10^{-3}$ | $5{\times}10^{-5}$ | 1 | 3 | ✓ | 250 |
| Power | 32 | 256 | $10^{-3}$ | $10^{-5}$ | 1 | 1 | ✓ | 250 |

The feature-to-tangent map $\mathcal{F}_{\text{steer}}$ is implemented as a two-layer MLP with LeakyReLU(0.2) activations, projecting node features into a space of dimension `hidden_dim`. The multi-head architecture uses $H = 4$ independent heads, each with its own learnable curvature scalar (initialised identically to the base manifold). For each head, the Lorentz exponential map and pairwise geodesic distances are computed independently. The resulting head-specific attention scores are then fused via a softmax-normalised learnable weighted sum. Scores are computed in chunks of $50{,}000$ edge pairs to bound peak GPU memory.

**Numerical stability.** The Lorentz operations are guarded against their standard failure modes: tangent-vector norms are floored at $10^{-8}$ before the square root in the exponential map, the argument of the $\operatorname{arccosh}$ geodesic distance is clamped to $\geq 1{+}10^{-5}$, spherical $\cos\theta$ is clamped to $[-1{+}\epsilon,\ 1{-}\epsilon]$, and every learnable curvature is floored at $10^{-5}$ through the $\operatorname{softplus}(\log c) + 10^{-3}$ parameterization. These keep each map and distance strictly within its valid domain throughout training, with no manual precision tuning or gradient clipping required.

The anchor gate $\lambda_{\text{anchor}}$ is initialised to $\sigma(0) = 0.5$ and the field gate $\lambda_{\text{field}}$ to $\sigma(-2) \approx 0.12$, reflecting the prior that the non-local field energy is a small correction. Both are optimised end-to-end with the remaining parameters.

The feature-aware Ollivier–Ricci curvature proxy uses lazy random-walk measures with retention mass $\alpha = 0.5$, distributing the remaining $1 - \alpha$ uniformly over one-hop neighbours. The joint ground metric weights topological and semantic dissimilarity equally ($\beta = 0.5$). The Sinkhorn algorithm uses regularisation $\varepsilon = 0.05$ and runs for 30 iterations. ORC values are precomputed in batches of $5{,}000$ candidate pairs before training begins and held fixed throughout.

All datasets use dropout $p = 0.5$ applied after each hidden layer in the DeepFusionMLP. Spectral normalisation is applied to the first two linear layers of the DeepFusionMLP to stabilise training under the $N$-pair ranking loss.

# F Additional Ablations and Robustness Analyses

## F.1 Learned Gate Values and the Gating Policy

The two scalar gates $\lambda_{\text{anchor}}$ and $\lambda_{\text{field}}$ are the only mechanism by which Concorde adapts its branch mixture to a dataset. Table 11 reports the values learned end-to-end (both initialised at $\lambda_{\text{anchor}} = 0.50$, $\lambda_{\text{field}} = 0.12$). The field gate rises monotonically as the graph sparsifies, from $0.10$ on dense OGBL-DDI to $0.63$ on sparse Power, a $6\times$ increase over its initialisation, and is dominant only on the sparsest graph, while the anchor gate is highest on dense OGBL-DDI. The gates thus recover, via an independent unsupervised parameter, the regime dependence hypothesised in Section 2.4 and established causally by the macro-ablation (Table 2) and rewiring enrichment (Appendix B). We report this as a consistent trend ($n = 5$ datasets), not a significance claim.

Table 11: Learned gate values by dataset, ordered by mean degree. Both gates are initialised at $(\lambda_{\text{anchor}}, \lambda_{\text{field}}) = (0.50, 0.12)$.

| Dataset | Mean degree | $\lambda_{\text{anchor}}$ | $\lambda_{\text{field}}$ |
|---|---|---|---|
| OGBL-DDI | dense | 0.62 | 0.10 |
| CiteSeer | 2.33 | 0.48 | 0.13 |
| Cora | 3.32 | 0.44 | 0.28 |
| PubMed | $\sim$4.5 | 0.47 | 0.33 |
| Power | 2.27 | 0.47 | 0.63 |

To confirm that adaptive gating is not cosmetic, we ablate only the gate policy (branches held fixed; mean AP over 3 runs) against two alternatives: equal weighting ($\lambda = 1$) and a per-dataset oracle-tuned fixed gate (Table 12). Learned $>$ tuned $>$ equal on all five datasets. The gap is largest where the optimal weighting is most skewed: on OGBL-DDI the learned field gate is driven to $0.10$, so forcing $\lambda = 1$ collapses AP from $42.32$ to $30.45$ ($-11.87$), and even the best per-dataset oracle-tuned fixed gate reaches only $40.18$. Adaptive gating therefore recovers, with no hyperparameter search, weightings that outperform hand-tuned constants, and does so most where it matters.

Table 12: Gating-policy ablation (AP, mean over 3 runs). Learned gating versus a per-dataset oracle-tuned fixed gate and equal weighting ($\lambda = 1$).

| Dataset | Learned | Tuned fixed | Equal ($\lambda$=1) | Learned − Equal |
|---|---|---|---|---|
| OGBL-DDI | 42.32 | 40.18 | 30.45 | $\mathbf{+11.87}$ |
| Power | 6.39 | 6.18 | 4.20 | $+2.19$ |
| PubMed | 2.83 | 2.73 | 2.32 | $+0.51$ |
| CiteSeer | 5.05 | 4.92 | 4.78 | $+0.27$ |
| Cora | 4.50 | 4.40 | 4.24 | $+0.26$ |

### F.2 Robustness of Laplacian Positional Encodings

On the featureless graphs (OGBL-DDI, Power), Concorde substitutes absent node attributes with Laplacian Positional Encodings (LPEs). We verify that results are neither fragile to the LPE dimension nor materially contaminated by leakage; both checks are single-run and reported as trends.

**Dimension sweep.** Table 13 sweeps the LPE dimension $k$. Power is stable for $k \geq 32$ and only under-parameterised at $k = 16$ (its 32 non-trivial eigenvectors already span the low-degree structure), so the reported $k = 32$ sits on a flat plateau rather than a tuned optimum. OGBL-DDI improves monotonically and saturates near the reported $k = 128$. The sweep therefore supports, rather than cherry-picks, the chosen dimensions.

Table 13: LPE dimension sweep (AP) on the featureless graphs.

| LPE dim $k$ | 16 | 32 | 64 | 128 |
|---|---|---|---|---|
| Power | 4.55 | 6.36 | 6.12 | 6.47 |
| OGBL-DDI | 38.33 | 40.47 | 42.10 | 43.28 |

**Leakage under masking.** Because LPEs are precomputed on the full graph, they could in principle encode connectivity that is masked per-batch during training. We tested this directly by recomputing LPEs on the *training-only* graph, excluding all validation and test connectivity, and retraining (Table 14). The change is small and in the expected direction (training-only sees strictly less structure and is marginally lower): $0.24$ AP on Power and $0.44$ on OGBL-DDI, far below the method's margins over baselines in Table 1. This matches the Davis–Kahan bound in Appendix D, and the per-batch adjacency masking of target edges, which is exact across all three branches, remains the primary anti-leakage mechanism and is unaffected.

Table 14: LPE leakage check (AP): encodings computed on the full graph (as reported) versus the training-only graph.

| LPE source | Power AP | OGBL-DDI AP |
|---|---|---|
| Full graph (as reported) | 6.36 | 43.28 |
| Training-only graph | 6.12 | 42.84 |

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the framework's capabilities (Tri-branch decoupled energy minimization) and explicitly highlight that the framework nearly doubles early retrieval precision over baseline methods in sparse settings, which perfectly aligns with the experimental results detailed in Section 3.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are explicitly discussed in Section 4, specifically addressing the computational overhead of the ORC proxy Sinkhorn iterations on larger graphs and the potential suboptimality of uniform branch weighting via global scalar gates.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Full proofs for the Invariance-Equivariance Trade-off (Proposition 1) and Spectral Stability (Proposition 2) are provided in Appendix A, complete with clearly stated structural assumptions (e.g., A1 and A2).

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Appendix D (Evaluation Protocol) and Appendix E (Implementation Details and Hyperparameters) thoroughly detail the dataset splits, negative pool construction, masking strategies, and exact hyperparameter configurations used for all models.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The complete codebase is provided in the supplementary material (Concorde.zip). It includes the primary execution scripts, alongside automated data downloading, preparation, and splitting protocols.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix E provides a comprehensive breakdown (Table 6) of dataset-specific hyperparameters, hidden dimensions, learning rates, and optimization strategies.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 1 and Table 2 report the mean and standard deviation over 10 independent runs for all primary metrics, ensuring statistical reliability across the random seeds.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix C.1 explicitly details the hardware used (a single Nvidia H200 GPU instance) and provides exact wall-clock runtimes for both ORC precomputation and per-epoch training in Table 5.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our work uses only publicly available data with no personal or sensitive information, and follows all NeurIPS ethical guidelines for data use and reporting.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper focuses on foundational theoretical advancements in geometric deep learning and link prediction. We foresee no direct, immediate negative societal applications of improving retrieval precision in sparse network topologies.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The work does not release pretrained models or datasets with a high risk for dual-use or malicious misuse.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All benchmark datasets (Cora, CiteSeer, PubMed, OGBL-DDI, Power) are standard public benchmarks and are properly cited in Section 3 and the Bibliography.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The newly introduced Concorde framework code is released as an open-source asset. The supplementary material includes comprehensive documentation (via a detailed README.md) outlining the repository structure, architectural components, dependencies, and a minimal execution example.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The research does not involve human subjects or crowdsourcing.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The research does not involve human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: Large Language Models were not used as an original or non-standard component of the core methods developed in this research.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.