MORSEFORMER: SCALABLE TOPOLOGICAL GRAPH TRANSFORMERS VIA DISCRETE MORSE THEORY

Anonymous authors
Paper under double-blind review

ABSTRACT

Higher-order structure such as cycles and cliques is invisible to standard message passing, whose expressivity is bounded by the Weisfeiler–Lehman test. Topological graph learning recovers it by lifting a graph to a cell complex in which such structures become explicit cells, and models that pass messages or attend over these cells gain expressive power. The lift, however, multiplies the number of cells, most of which carry no new topology, and global attention over the lifted complex becomes expensive, and the topological descriptors computed on it are even more expensive. We introduce MorseFormer, a graph transformer that reduces the lifted complex with discrete Morse theory before learning on it. An acyclic matching collapses cells in pairs and leaves a small set of critical cells whose Morse complex has the same homotopy type of the lifted complex; the Morse retraction carries the attributes of the removed cells onto the critical cells, which become the tokens of a transformer, positioned in the reduced topology by its Hodge Laplacian positional encoding. MorseFormer++ uses the resulting representation together with a graph backbone, so higher-order information enters without attention over the full lift. Morse reduction turns a lifted complex into a token sequence an order of magnitude shorter while preserving its homology, and attention over that sequence is correspondingly cheaper. Across graph classification and molecular benchmarks, MorseFormer++ reaches the best reported accuracy on several datasets, and descriptors read from the Morse complex alone separate more pairs of graphs that Weisfeiler–Lehman cannot tell apart. Discrete Morse reduction thus serves as a representation mechanism for scalable higher-order topological deep learning.

1 INTRODUCTION

A well-known limitation of message-passing graph neural networks (GNNs) (Gori et al., 2005; Hamilton et al., 2017) is their limited ability to capture potentially meaningful topological information, such as cycles and cliques (Chen et al., 2020; Garg et al., 2020; Horn et al., 2022). This limitation stems from their expressive power, which is bounded by the Weisfeiler–Lehman test (Xu et al., 2019; Morris et al., 2019). To address this, topological deep learning (TDL) extends message passing to higher-order relational structures, known as cells, constructed from the original graph through a procedure called lifting. For instance, *clique lifting* (Bodnar et al., 2021b) constructs a simplicial complex from the cliques of the input graph, whereas cycle lifting (Bodnar et al., 2021a) identifies independent cycles to construct a cell complex. Thus, by explicitly representing these structures, TDL methods make higher-order relational information directly accessible to message-passing layers over generalized neighborhood structures (Papillon et al., 2023). Beyond their theoretical gains in expressivity (Bodnar et al., 2021a; Akbari et al., 2026), these models have also demonstrated benefits on real-world downstream tasks (Bodnar et al., 2021a; Giusti et al., 2023; Eitan et al., 2025).

However, depending on the graph structure, lifting can generate a large number of cells, substantially increasing both memory requirements and computational cost. This challenge is particularly pronounced in Transformer-based architectures, whose self-attention mechanism scales quadratically with the number of cells. To illustrate this issue, Figure 1 shows the impact of the lifting procedure in terms of memory (sequence lenght) and time (forward/backward step) on a representative TDL model, the Cellular Transformer (Ballester et al., 2024), across popular graph classification benchmarks.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

A cycle lifting turns a ZINC molecule of 23 atoms into 51 cells and a PROTEINS graph of 39 nodes into 168, with single graphs reaching 2,444; because every cell attends to every other, the time of one attention layer grows faster than the token count, by $2\times$ on the molecular datasets and $9\times$ on PROTEINS, and on a lift of 38,000 cells one attention pass takes 285 ms on one GPU (Section 4.3). Yet most of these cells are redundant as carriers of topology: the 51 cells of a ZINC lift have the homology of a point. Attention-based models (Ballester et al., 2024; Uddin et al., 2026) therefore face a tension: the lifts that carry the most information are the ones on which attention is least affordable.

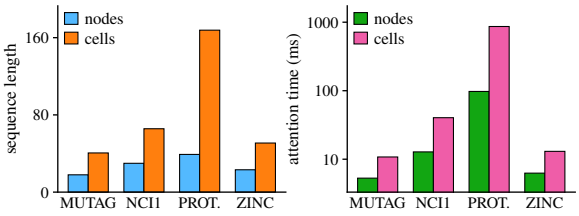

Figure 1: Mean sequence length and single-layer runtime (in log scale) for graph versus cellular Transformers — measured on CPU with batch size equal to 32.

In this work, we aim to improve the computational performance of TDL methods while preserving relevant higher-order topological information for the task at hand. In doing so, we propose shrinking the lifted complex before applying the message-passing or attention operations. However, this shrinking cannot be arbitrary: it must keep the topology that motivated the lift, or the higher-order cells were created for nothing. Discrete Morse theory (Forman, 1998) provides a reduction with exactly this guarantee. An *acyclic matching* pairs cells of adjacent ranks that can be collapsed together. The cells that remain unpaired are *critical*, and they span a smaller complex, the *Morse complex*, with the same homotopy type as the lifted complex (Harker et al., 2014; Ebli et al., 2024).

Importantly, homotopical equivalence alone does not tell us how to learn on the reduced complex. The cells removed by the collapse may carry task-relevant attributes, and the surviving critical cells must be represented in a way that reflects their role in the reduced topology. Discrete Morse theory supplies one more object that makes this possible: the *Morse retraction*, a linear map from the lifted complex to the Morse complex. Its nonzero entries identify, for every critical cell, a *basin* of original cells that flow into it. We use the retraction to transport attributes from the basin to the critical cell. We then encode each critical cell by its transported attributes, by eigenvectors of the reduced Hodge Laplacian (which position the cell in the reduced topology), and by its rank.

With these ideas in mind, we introduce MorseFormer — a Transformer-based model that operates on the resulting critical cells. Our empirical results further suggest that retaining the original graph structure is often beneficial. Motivated by this observation, we propose MorseFormer++, an architecture that combines MorseFormer with standard Graph Transformers (Rampášek et al., 2022).

We demonstrate the strong predictive performance of MorseFormer(++) under multiple evaluation protocols and graph-level prediction tasks (Section 4.1). We also evaluate our models on expressivity benchmarks (BREC datasets) (Wang & Zhang, 2024) in Section 4.2. Finally, we conduct ablation studies regarding the choice of attribute transport (Appendix D), and we report results on the compression and attention savings of the proposed method (Section 4.3).

2 Background

2.1 Cell complexes and their algebra

Cell complexes. A *regular cell complex* (Bodnar et al., 2021a) is a space built by attaching cells of increasing dimension, each k-cell being homeomorphic to an open k-ball whose boundary is glued to cells of lower dimension. Vertices are 0-cells, edges are 1-cells, and triangles or polygons are 2-cells. We write σ, τ for cells, $\dim \sigma$ for dimension, and n_k for the number of k-cells. The cells of a complex are related by incidence. The cell σ is a *face* of τ, written $\sigma < \tau$, when σ lies on the boundary of τ; when in addition $\dim \tau = \dim \sigma + 1$, σ is a *facet* of τ and τ a *coface* of σ. A cell with exactly one coface is a *free face*. The *Hasse diagram* is the directed graph on the cells with an arc $\sigma \to \tau$ for every facet relation; it records the whole combinatorial structure of the complex. The 1-*skeleton* is the graph formed by the 0- and 1-cells. A *simplicial complex* is the special case in which every k-cell is a k-simplex, the convex hull of $k+1$ vertices, closed under taking faces. Following Ebli et al. (2024), we equip complexes with *attribute matrices* $X = \{X_k \in \mathbb{R}^{n_k \times d}\}_{k \geq 0}$, where row i of X_k holds the feature vector of the i-th k-cell. Input graphs provide X_0.

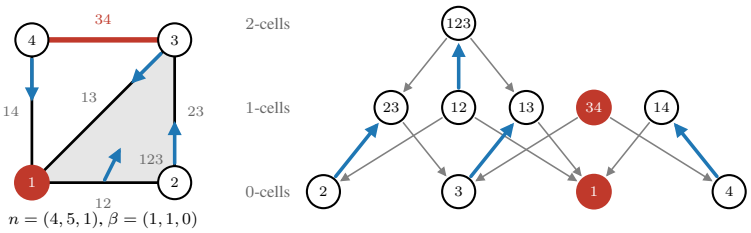

Figure 2: **Acyclic matching.** Left: a complex with a filled triangle 123 and an open triangle 134, with $\beta = (1, 1, 0)$. Blue arrows indicate matched pairs. Right: the gradient graph G_μ, with matched arcs reversed (blue) and two critical cells (red), yielding $c = \beta$ and reducing ten cells to two. The path $2 \to 23 \to 3 \to 13 \to 1$ illustrates a gradient V-path. Acyclicity prevents such paths from forming cycles.

Boundary matrices. The boundary operator accounts for how cells of adjacent ranks are attached. Each cell complex induces a *chain complex* (Forman, 1998) (C_*, B_*): the *chain space* $C_k \cong \mathbb{R}^{n_k}$ has one basis vector per k-cell, and the *boundary matrix* $B_k \in \mathbb{R}^{n_{k-1} \times n_k}$ has $B_k[\sigma, \tau] = \pm 1$ when σ is a facet of τ, with the sign fixed by a choice of orientation, and 0 otherwise. Each column of B_k therefore describes the oriented boundary of one k-cell. A key property is that $B_{k-1}B_k = 0$.

Homology and Betti numbers. The kernel of B_k consists of k-chains without boundary, the *cycles*; the image of B_{k+1} consists of the cycles that bound a $(k+1)$-chain. Because $B_k B_{k+1} = 0$, every boundary is a cycle, and the quotient $H_k = \ker B_k / \operatorname{im} B_{k+1}$ is the *k-th homology*: the cycles that are not filled in. Its dimension, the *k-th Betti number* $\beta_k = \dim \ker B_k - \operatorname{rank} B_{k+1}$, counts connected components (β_0), independent unfilled loops (β_1) and enclosed voids (β_2). Two complexes are *homotopy equivalent* when one can be continuously deformed into the other; homotopy-equivalent complexes have the same homology, while their numbers of cells may differ.

Remark 1. *In general, a chain complex does not fully determine its underlying cell complex. However, the regular lifts considered here have their cell incidences and orientations encoded in the boundary matrices. We therefore adopt the algebraic formulation of discrete Morse theory (Sköldberg, 2006), as MorseFormer operates entirely on (C_*, B_*) and cell attributes X. The resulting Morse complex $(D_*, \tilde{B}_*)$ is chain-homotopy equivalent to (C_*, B_*) and, by Forman's theorem (Forman, 1998), corresponds to a cell complex homotopy-equivalent to the original one.*

The Hodge Laplacian. The boundary matrices also define a Laplacian at every rank, $\Delta_k = B_k^\top B_k + B_{k+1} B_{k+1}^\top \in \mathbb{R}^{n_k \times n_k}$. The first term relates two k-cells that share a facet, the second relates two k-cells that lie on a common coface, so Δ_k summarises the incidence structure at rank k. For $k = 0$ it is the graph Laplacian. It is symmetric positive semidefinite and its kernel has dimension exactly β_k, so its spectrum carries both the local incidence pattern and the global homology.

Liftings. A *lifting* decides which higher-order structures of a graph become cells. Formally it maps a graph $G = (V, E)$ to a cell complex $\mathcal{L}(G)$ with $C_0 = \mathbb{R}^{|V|}$, $C_1 = \mathbb{R}^{|E|}$, and n_k attached k-cells at each rank $k \geq 2$ (Bodnar et al., 2021a). The *clique lifting* attaches a k-simplex to every $(k+1)$-clique up to a maximal dimension $k_{\max}$ and yields a simplicial complex. The *cycle lifting* attaches a 2-cell to every chordless cycle of length at most $\ell_{\max}$ and yields a cell complex whose 2-cells are polygons.

2.2 Discrete Morse Theory

Collapses. An *elementary collapse* removes a k-cell and an incident $(k+1)$-cell without changing the homotopy type of the complex. Discrete Morse theory organizes these collapses through a *matching* of incident cells, leaving unmatched cells *critical*. To ensure a valid reduction, the matching must be *acyclic*: reversing matched arcs in the Hasse diagram must yield a directed acyclic graph, which determines an order in which the reduction can be processed. Figure 2 shows a matching on the Hasse diagram of a small complex. It leaves one critical vertex and one critical edge and every other cell is paired. The maps that make this collapse precise are introduced in the following.

Definition 2 (Acyclic Matching (Forman, 1998))**.** *A* matching *μ on (C_*, B_*) is a set of pairs (σ, τ) with σ a k-cell, τ a $(k+1)$-cell, $B_{k+1}[\sigma, \tau] \neq 0$, and each cell appearing in at most one pair. Cells in no pair are* critical*, and we write c_k for the number of critical k-cells. The* gradient graph *G_μ is the Hasse diagram with each matched pair's arc reversed; μ is* acyclic *if G_μ has no directed cycles.*

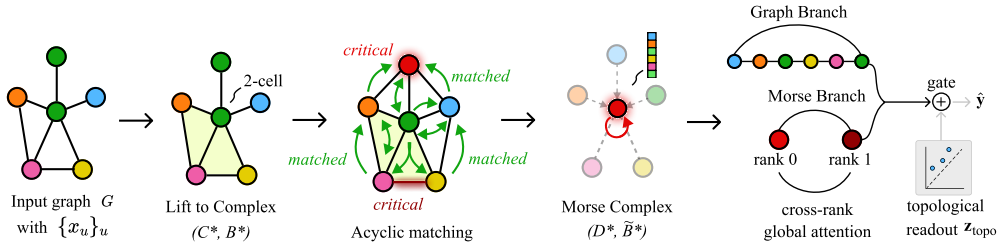

Figure 3: **MorseFormer.** A graph is lifted to a cell complex; an acyclic matching (green arrows) pairs collapsible cells and leaves the critical cells (red), which span the Morse complex. The Morse retraction transports attributes onto the critical cells, which are encoded with the reduced boundary and Hodge geometry and processed by the Morse encoder, whose pooled output is the Morse representation. In MorseFormer++ (Section 3.5) this representation corrects the output of a graph backbone on the original graph through a gate before the prediction.

Gradient paths. A *gradient V-path* (Forman, 1998) is a directed path in G_μ that alternates between reversed (matched) arcs and Hasse arcs, $\sigma_0 \to \tau_0 \to \sigma_1 \to \tau_1 \to \cdots$, with every $(\sigma_i, \tau_i) \in \mu$ and every σ_{i+1} a facet of τ_i other than σ_i. Such a path describes how a collapse at one pair pushes structure onto the next. The matching is acyclic exactly when no V-path is closed, in which case the matched cells admit a topological order, in which collapses can be carried out.

Importantly, we also need a map to transport information from removed cells to surviving ones. The discrete Morse *flow* provides this map. Let $V_k : C_k \to C_{k+1}$ send each matched k-cell σ to $e_\tau / B_{k+1}[\sigma, \tau]$, where τ is its partner, and all other basis vectors to 0. The flow operator $\Phi_k = I - B_{k+1}V_k - V_{k-1}B_k$ moves chains along gradient paths. Acyclicity ensures finite convergence (Proposition 4) to $\Phi_k^\infty = I - B_{k+1}\Sigma^{(k)} - \Sigma^{(k-1)}B_k$, where $\Sigma^{(k)}$ is Sköldberg's *splitting homotopy*, computed recursively in topological order (Sköldberg, 2006) (Section E.2). Restricting the converged flow to the critical cells gives the maps between the two complexes.

Definition 3 (Morse Complex and Retraction (Sköldberg, 2006; Ebli et al., 2024)). *Let $\mathcal{I}_k \subset \{1, \ldots, n_k\}$ index the critical k-cells and $D_k := \mathbb{R}^{c_k}$ have basis $\{\hat{e}_j\}_{j \in \mathcal{I}_k}$. The* Morse retraction *and* inclusion *are the submatrices $\phi_k = \Phi_k^\infty[\mathcal{I}_k, :] \in \mathbb{R}^{c_k \times n_k}$ and $\psi_k = \Phi_k^\infty[:, \mathcal{I}_k] \in \mathbb{R}^{n_k \times c_k}$. The* Morse boundary *$\tilde{B}_k = \phi_{k-1}B_k\psi_k$, and $(D_*, \tilde{B}_*)$ is the* Morse complex*: a chain complex homotopy-equivalent to (C_*, B_*) with $\beta_k(D_*) = \beta_k(C_*)$ for all k (Forman, 1998).*

The dimensions say what these maps do. There are n_k cells of rank k in the lift and c_k critical ones, so $\phi_k \in \mathbb{R}^{c_k \times n_k}$ maps the original rank-k chain space onto the smaller critical-cell space, and $\psi_k \in \mathbb{R}^{n_k \times c_k}$ embeds the critical cells back. In Figure 2, ϕ_0 is the row of ones that sends every vertex to the critical vertex, and ϕ_1 selects the critical edge. The Morse boundary $\tilde{B}_k = \phi_{k-1}B_k\psi_k$ is the boundary operator the critical cells inherit. Its entries can be integers of magnitude larger than one, because a single critical cell may be reached along several gradient paths.

3 MorseFormer

MorseFormer has two steps (Figure 3). First, the input graph is lifted and the lifted complex is reduced with an acyclic Morse matching. Second, attributes are transported from the lifted cells onto the critical cells, each critical cell is encoded with the reduced topology, and a transformer over the critical cells produces the Morse representation of the graph. MorseFormer++ adds a third step, in which this representation corrects a graph backbone.

3.1 Reducing the lifted complex

The ideal matching leaves as few critical cells as possible, but finding it is NP-hard (Joswig & Pfetsch, 2006). We use a three-phase heuristic adapted from the reduction–cancellation pipeline of computational topology (Kaczyński et al., 1998; Mischaikow & Nanda, 2013; Harker et al., 2014), with an initialisation for graph liftings: free-face coreduction pairs the immediately collapsible cells (Kaczyński et al., 1998), spanning-tree pairing leaves one critical vertex per connected component, and V-path cancellation removes pairs of critical cells joined by a unique gradient path (Forman, 1998, §10). Each phase preserves acyclicity; Section E.2 analyses them.

Three-phase acyclic matching for graph liftings

Input: chain complex (C_*, B_*). **Output:** matching μ, critical index sets $\{\mathcal{I}_k\}_k$, retraction ϕ_k, inclusion ψ_k, and Morse boundary $\tilde{B}_k$.

Phase 1 (free-face coreduction). Match any unmatched cell σ with a unique unmatched coface τ, repeating to fixpoint (Harker et al., 2014, Alg. 3.6).
Phase 2 (spanning-tree pairing). For each connected component, run BFS on the 1-skeleton from a root vertex; for every non-root v with parent edge e_v, add (v, e_v) to μ when both are unmatched. When Phase 1 has not consumed any tree edge, this saturates $c_0 = \beta_0$.
Phase 3 (V-path cancellation). While there exists a unique gradient V-path between a critical $(k{+}1)$-cell τ and a critical k-cell σ in G_μ, reverse it (Forman, 1998, §10); each reversal reduces $\sum_k c_k$ by 2.

Retraction. Solve the recursion for $\Sigma^{(k)}$ by topological-sort dynamic programming (Harker et al., 2014), then read off $\phi_k, \psi_k, \tilde{B}_k$ as submatrices of Φ_k^∞.

Computing the retraction. Building the converged flow by repeated multiplication of dense $n_k \times n_k$ operators be costly. Acyclicity results that in a topological order of the matched cells, the flow operator acting on them is strictly triangular, so a single dynamic-programming sweep gives the converged flow without forming a dense operator.

Proposition 4 (Exact linear-time Morse retraction)**.** *Let (C_*, B_*) carry an acyclic matching μ, and order the matched k-cells by a topological sort of the matched-cell digraph induced by G_μ (a DAG since μ is acyclic). Then (i) the matched-cell block of the flow operator Φ_k has zero diagonal and is strictly lower-triangular in this order, hence nilpotent of index at most $n_k - c_k$; (ii) consequently $\Phi_k^m = \Phi_k^\infty$ for every $m \geq n_k - c_k$, and Φ_k^∞ is obtained* exactly *by a single back-substitution sweep over the topological order (the gradient-path recursion of Definition 3); (iii) the sweep returns ϕ_k, ψ_k and $\tilde{B}_k = \phi_{k-1} B_k \psi_k$ identical to the dense fixed-point iteration, in $O(c_k \cdot \mathrm{nnz}(B_*))$ time and $O(c_k\, n_k)$ memory versus $O(n_k^3)$. (Proof in Section E.2.)*

The sweep returns the same ϕ_k, ψ_k and $\tilde{B}_k$ as the dense fixed-point formulation of Ebli et al. (2024) at a cost of $O(c_k \cdot \mathrm{nnz}(B_*))$ per rank, cheap enough under a learnable lifting (Section C).

3.2 Transporting attributes to the critical cells

The reduction keeps the topology but does not say which feature vector each surviving cell should carry: n_k cells of rank k enter and c_k remain. Before that, the higher-rank cells need attributes at all, since the input graph provides features only on vertices.

Attributes on the lifted complex. We set $X_0 = [x_v]_{v \in V} \in \mathbb{R}^{n_0 \times d}$ from the node features and fill the higher ranks from the boundary. The nonzero entries of column j of B_k identify the $(k{-}1)$-cells in the boundary of cell j, so we define, for $k \geq 1$,

$$X_k[j,:] \;=\; \frac{\sum_i |B_k[i,j]|\, X_{k-1}[i,:]}{\sum_i |B_k[i,j]|}. \tag{1}$$

Morse basins. For a critical cell $i \in \mathcal{I}_k$, the nonzero entries of row i of the retraction $\phi_k \in \mathbb{R}^{c_k \times n_k}$ identify the original cells that flow into it. We call $\mathcal{B}_k(i) = \{j : \phi_k[i,j] \neq 0\}$ the *Morse basin* of i and consider three ways to summarise the attributes inside it.

Fixed Morse projection. The coefficients $\phi_k[i,j]$ are already weights over the basin, so the retraction itself can aggregate:

$$X_k^{\mathrm{morse}}[i,:] \;=\; \sum_{j \in \mathcal{B}_k(i)} \phi_k[i,j]\, X_k[j,:], \qquad \text{that is,} \qquad X_k^{\mathrm{morse}} = \phi_k X_k \in \mathbb{R}^{c_k \times d}. \tag{2}$$

This introduces no parameters, and the triple $(D_*, \tilde{B}_*, X^{\mathrm{morse}})$ is the *attributed Morse complex*.

DeepSets aggregation. A learned set function keeps nonlinear information about the basin while remaining invariant to the order of its cells. A rank-specific encoder is applied to every cell, the encodings are weighted by the flow coefficients, summed and decoded:

$$X_k^{\mathrm{morse}}[i,:] \;=\; \rho_k\left(\sum_{j\in\mathcal{B}_k(i)} \phi_k[i,j]\,\mathrm{MLP}_k^{\mathrm{set}}\big(X_k[j,:]\big)\right), \tag{3}$$

where $\mathrm{MLP}_k^{\mathrm{set}}$ and ρ_k are learnable maps (Zaheer et al., 2017).

Attention aggregation. The critical cell can also select the cells of its basin that matter. The fixed projection provides the query and the basin cells the keys and values: with $Q_k = W_k^Q(\phi_k X_k)$, $K_k = W_k^K X_k$, $V_k = W_k^V X_k$ and scores $s_{ij} = Q_k[i,:]\,K_k[j,:]^\top/\sqrt{d}$,

$$X_k^{\mathrm{morse}}[i,:] \;=\; \sum_{j\in\mathcal{B}_k(i)} \frac{\exp(s_{ij})}{\sum_{j'\in\mathcal{B}_k(i)}\exp(s_{ij'})}\,V_k[j,:]. \tag{4}$$

The aggregators change only how attributes are transported; the critical cells and the reduced boundary are the same (Section D compares them). Under a learnable lifting (Franco et al., 2026), the candidate-cell selectors pass through a straight-through estimator (Bengio et al., 2013); the matching and retraction are recomputed on every forward pass and detached on the backward pass, so gradients flow through X_k and the aggregator but not through ϕ_k. Under a static lifting the retraction is computed once as preprocessing.

3.3 Encoding critical cells

A token should also say where its cell sits in the reduced complex and what kind of cell it is. For each critical k-cell with index $i \in \mathcal{I}_k$ we form

$$z_i \;=\; \mathrm{MLP}_k(X_k^{\mathrm{morse}}[i,:]) \;+\; \mathrm{HodgePE}_k(i) \;+\; \mathrm{RankEmb}(k), \tag{5}$$

The first term encodes the transported attributes through a rank-specific encoder. The second positions the cell: the Hodge Laplacian of the Morse complex,

$$\tilde{\Delta}_k \;=\; \tilde{B}_k^\top \tilde{B}_k \;+\; \tilde{B}_{k+1}\,\tilde{B}_{k+1}^\top \;\in\; \mathbb{R}^{c_k\times c_k}, \tag{6}$$

summarises the incidence structure among critical k-cells as Δ_k does for the lift (Section 2.1), and its eigenvectors of smallest eigenvalue vary slowly across the reduced complex, so they give spectral coordinates for the critical cells. $\mathrm{HodgePE}_k(i)$ is the i-th row of the matrix of the p such eigenvectors, made invariant to the sign of each eigenvector by SignNet (Lim et al., 2023). Because the decomposition is taken after reduction, it describes the topology the model operates on and costs an eigendecomposition of size c_k rather than n_k. The third term, $\mathrm{RankEmb}(k) \in \mathbb{R}^d$, is a learnable embedding of the rank, needed because all ranks share one sequence.

3.4 Communication among critical cells

Critical cells communicate in two ways. Locally, two critical cells are neighbours when $\tilde{B}_*$ has a nonzero entry between them, and the entry is used as an edge attribute. Globally, self-attention runs over the whole critical-cell sequence, across all ranks. Each layer of the *Morse encoder* combines the two; writing $\mathcal{N}(i)$ for the Morse-boundary neighbourhood of token i,

$$\begin{aligned} m_i^{(\ell)} &= \mathrm{MPNN}^{(\ell)}\big(z_i^{(\ell)},\,\{z_j^{(\ell)}\}_{j\in\mathcal{N}(i)}\big), \qquad a_i^{(\ell)} = \mathrm{Attn}^{(\ell)}\big(z_i^{(\ell)},\,\{z_j^{(\ell)}\}_j\big), \\ z_i^{(\ell+1)} &= \mathrm{FFN}\Big(\mathrm{Norm}\big(z_i^{(\ell)} + m_i^{(\ell)} + a_i^{(\ell)}\big)\Big), \end{aligned} \tag{7}$$

where the local block is a GatedGCN or GINE layer and the attention index j ranges over every critical cell of the graph. This is the layer structure of GPS (Rampášek et al., 2022) applied to critical cells; the rank embedding lets the attention tell a vertex token from a face token. Global attention also has a structural role: the Morse complex can be non-regular, so $\tilde{B}_*$ may have a zero entry between two critical cells that are incident in the lift, because the contributions of several gradient paths cancel. The local block cannot connect such cells; attention can. Mean-pooling the critical-cell representations gives the Morse representation h_M of the graph.

3.5 MORSEFORMER++

MorseFormer learns from the critical cells alone. MorseFormer++ also keeps a *graph backbone*, a GPS stack (Rampášek et al., 2022) over the original graph with Laplacian positional encodings, and lets the Morse representation correct it. Let h_G be the mean-pooled backbone output. A gate computed from both decides, coordinate by coordinate, how much of the correction to apply:

$$g \;=\; \mathrm{sigmoid}\big(W_g\,[h_G;\, h_M] + b_g\big), \qquad h \;=\; h_G \;+\; g \odot (h_M - h_G). \tag{8}$$

which equals the convex combination $(1-g)\odot h_G + g \odot h_M$. The predictor also receives a fixed-size summary z_{topo} of the Morse complex: persistence diagrams along an edge filtration, Betti curves, the smallest Hodge eigenvalues and statistics of the Morse boundary (Section E.1). The prediction is $\hat{y} = \mathrm{MLP}([\,h;\, z_{\mathrm{topo}}\,])$; Section 4.2 evaluates the summary on its own.

4 EXPERIMENTS

We ask four questions. **RQ1:** does the Morse representation improve accuracy under established graph-classification and molecular protocols? **RQ2:** how much does the choice of basin aggregator matter? **RQ3:** how much graph structure do descriptors read from the Morse complex distinguish without training? **RQ4:** how much does the reduction shorten the cell sequence, and what does that save? Section 4.1 answers RQ1 and RQ2, Section 4.2 RQ3 and Section 4.3 RQ4.

We use the protocol of each paper setup: including splits, seeds, metric and model selection, and baseline numbers are copied from that paper; numbers from different protocols are not comparable. The one baseline we train ourselves is GPS (Rampášek et al., 2022), the backbone of MorseFormer++ and therefore the reference for the Morse correction. The domains of the tables are the three liftings of Section 2.1: graph (no lifting), simplicial (clique lifting) and cellular (chordless-cycle lifting). Within each domain we report **MorseFormer**, the critical-cell transformer alone, and **MorseFormer++**, which adds the GPS backbone through the gate of Equation (8). The gap between the two rows is the contribution of the backbone; the gap between GPS and MorseFormer++ is the contribution of the Morse representation (hyperparameters in Section E).

4.1 PREDICTIVE PERFORMANCE

Graph classification. We follow Bernárdez et al. (2025) and Papillon et al. (2025): stratified $50/25/25$ splits, five seeds, best-validation test selection, accuracy on MUTAG, PROTEINS, NCI1 and NCI109 and MAE on ZINC-12K without edge features. The baselines, from Table 1 of Bernárdez et al. (2025), are GCN (Kipf & Welling, 2017) and GAT (Veličković et al., 2018) with and without positional and structural encodings (Cantürk et al., 2024), SCCNN (Yang et al., 2022), and the GCCN (Papillon et al., 2025) and HOPSE (Bernárdez et al., 2025) families on simplicial and cellular liftings. Table 1 reports the results by lifting domain.

In the graph domain, the GPS row is the backbone alone, so its distance to MorseFormer++ is the effect of the Morse correction with everything else fixed: $+7.8$ points on MUTAG, $+7.2$ on PROTEINS, $+7.6$ on NCI1 and $+5.7$ on NCI109, while on ZINC-12K the MAE moves from 0.11 to 0.13. MorseFormer alone trails MorseFormer++ by 8.4 and 8.1 points on NCI1 and NCI109: the Morse complex of the bare graph does not carry these tasks on its own. After lifting the gap closes. In the cellular domain MorseFormer alone is within 2.1 to 3.5 points of MorseFormer++ on the four classification datasets and matches or exceeds HOPSE-G on MUTAG, NCI1 and NCI109. The cellular MorseFormer++ is the strongest entry in the table on every dataset: 89.80 on MUTAG, 77.14 on PROTEINS, 83.53 on NCI1, 80.93 on NCI109 and 0.10 MAE on ZINC-12K, ahead of the strongest cellular baseline by 1.7, 0.2, 4.5 and 3.1 points and 0.09 MAE. In the simplicial domain MorseFormer++ leads on four of five datasets; on PROTEINS the three top entries (MorseFormer 76.32, HOPSE-G 76.27, MorseFormer++ 75.67) lie within one standard deviation.

Table 1: **Predictive performance under HOPSE/TopoTune.** Test accuracy % (↑) on MUTAG, PROTEINS, NCI1 and NCI109 and MAE (↓) on ZINC-12K; mean and standard deviation over five seeds. Rows are grouped by lifting domain. Best per domain in bold, best across the table in blue.

Domain	Method	MUTAG↑	PROTEINS↑	NCI1↑	NCI109↑	ZINC-12K↓
Graph	GCN	$69.79_{\pm 6.80}$	$75.70_{\pm 2.14}$	$72.86_{\pm 0.69}$	$72.20_{\pm 1.22}$	$0.62_{\pm 0.01}$
	GCN+PSEs	$84.26_{\pm 3.56}$	$75.77_{\pm 2.98}$	$74.88_{\pm 0.54}$	$74.70_{\pm 0.79}$	$0.43_{\pm 0.03}$
	GAT	$72.77_{\pm 2.77}$	$76.34_{\pm 1.66}$	$75.00_{\pm 1.00}$	$73.80_{\pm 0.73}$	$0.61_{\pm 0.01}$
	GAT+PSEs	$80.43_{\pm 5.08}$	$75.48_{\pm 3.44}$	$77.86_{\pm 1.02}$	$75.57_{\pm 0.60}$	$0.39_{\pm 0.02}$
	GPS	$76.84_{\pm 8.81}$	$69.64_{\pm 3.57}$	$74.65_{\pm 1.47}$	$74.38_{\pm 2.35}$	$\mathbf{0.11}_{\pm 0.00}$
	MorseFormer	$84.21_{\pm 13.42}$	$74.29_{\pm 1.94}$	$73.85_{\pm 1.98}$	$71.90_{\pm 1.23}$	$0.42_{\pm 0.05}$
	MorseFormer++	$\mathbf{84.59}_{\pm 5.19}$	$\mathbf{76.83}_{\pm 4.89}$	$\mathbf{82.21}_{\pm 1.40}$	$\mathbf{80.04}_{\pm 1.60}$	$0.13_{\pm 0.01}$
Simplicial	SCCNN	$76.17_{\pm 6.63}$	$74.19_{\pm 2.86}$	$76.60_{\pm 1.75}$	$77.12_{\pm 1.07}$	$0.36_{\pm 0.02}$
	GCCN	$85.96_{\pm 4.66}$	$74.91_{\pm 2.51}$	$76.85_{\pm 1.50}$	$75.76_{\pm 1.28}$	$0.35_{\pm 0.01}$
	HOPSE-M	$86.38_{\pm 6.31}$	$75.13_{\pm 3.66}$	$77.35_{\pm 0.81}$	$76.34_{\pm 1.40}$	$0.39_{\pm 0.02}$
	HOPSE-G	$85.53_{\pm 2.77}$	$76.27_{\pm 2.10}$	$77.70_{\pm 1.40}$	$76.36_{\pm 1.75}$	$0.43_{\pm 0.01}$
	MorseFormer	$82.11_{\pm 10.26}$	$\mathbf{76.32}_{\pm 2.67}$	$72.87_{\pm 1.92}$	$73.06_{\pm 0.85}$	$0.33_{\pm 0.02}$
	MorseFormer++	$\mathbf{86.56}_{\pm 6.85}$	$75.67_{\pm 2.67}$	$\mathbf{79.61}_{\pm 2.48}$	$\mathbf{77.97}_{\pm 2.77}$	$\mathbf{0.13}_{\pm 0.07}$
Cellular	GCCN	$86.38_{\pm 6.49}$	$74.41_{\pm 1.77}$	$78.23_{\pm 1.47}$	$77.19_{\pm 0.21}$	$0.19_{\pm 0.01}$
	HOPSE-M	$88.09_{\pm 5.55}$	$76.99_{\pm 2.63}$	$78.09_{\pm 1.44}$	$77.27_{\pm 1.01}$	$0.24_{\pm 0.01}$
	HOPSE-G	$88.09_{\pm 1.90}$	$76.42_{\pm 2.19}$	$79.07_{\pm 0.65}$	$77.81_{\pm 1.26}$	$0.21_{\pm 0.01}$
	MorseFormer	$87.72_{\pm 2.88}$	$73.65_{\pm 2.98}$	$80.47_{\pm 0.89}$	$78.35_{\pm 2.36}$	$0.22_{\pm 0.02}$
	MorseFormer++	$\mathbf{89.80}_{\pm 4.87}$	$\mathbf{77.14}_{\pm 6.10}$	$\mathbf{83.53}_{\pm 2.90}$	$\mathbf{80.93}_{\pm 1.02}$	$\mathbf{0.10}_{\pm 0.01}$

Table 3: **Molecular benchmarks.** ZINC-12K with edge features (MAE, ↓) and ogbg-molhiv (ROC-AUC %, ↑); mean and standard deviation over seeds. Best per column in bold.

Family	Method	ZINC↓	MolHIV↑
MPNN	GCN (Kipf & Welling, 2017)	$0.321_{\pm 0.009}$	$76.06_{\pm 0.97}$
	GIN (Xu et al., 2019)	$0.252_{\pm 0.014}$	$77.07_{\pm 1.49}$
	PNA (Corso et al., 2020)	$0.188_{\pm 0.004}$	$79.05_{\pm 1.32}$
	DGN (Beaini et al., 2021)	$0.168_{\pm 0.003}$	$79.70_{\pm 0.97}$
	GPS (Rampášek et al., 2022)	$0.070_{\pm 0.004}$	$78.80_{\pm 1.01}$
Expressive	HIMP (Fey et al., 2020)	$0.151_{\pm 0.006}$	$78.80_{\pm 0.82}$
	GSN (Bouritsas et al., 2023)	$0.108_{\pm 0.018}$	$77.99_{\pm 1.00}$
	DS-GNN (Bevilacqua et al., 2022)	$0.087_{\pm 0.003}$	$76.54_{\pm 1.37}$
	DSS-GNN (Bevilacqua et al., 2022)	$0.102_{\pm 0.003}$	$76.78_{\pm 1.66}$
	SUN (Frasca et al., 2022)	$0.083_{\pm 0.003}$	$80.03_{\pm 0.55}$
	GNN-SSWL+ (Zhang et al., 2023a)	$0.070_{\pm 0.005}$	$79.58_{\pm 0.35}$
	Subgraphormer (Bar-Shalom et al., 2024)	$0.067_{\pm 0.007}$	$80.38_{\pm 1.92}$
	Subgraphormer+PE (Bar-Shalom et al., 2024)	$0.063_{\pm 0.001}$	$79.48_{\pm 1.28}$
Topological	Cellular Transformer (Ballester et al., 2024)	0.080	79.46
	CIN (Bodnar et al., 2021a)	$0.079_{\pm 0.006}$	$80.94_{\pm 0.57}$
	CIN++ (Giusti et al., 2023)	$0.077_{\pm 0.004}$	$80.63_{\pm 0.94}$
	SMCN (Eitan et al., 2025)	$\mathbf{0.060}_{\pm 0.004}$	$\mathbf{81.16}_{\pm 0.90}$
	MorseFormer++	$0.064_{\pm 0.002}$	$79.18_{\pm 1.37}$

10-fold cross-validation. We follow Uddin et al. (2026): 10-fold cross-validation on BZR, COX2, MUTAG and PROTEINS against DMP (Bodnar et al., 2021b), FC-V (O'Bray et al., 2021), G-Mix (Han et al., 2022), RGCL (Li et al., 2022), EMP (**?**), TopoGCL (Chen et al., 2024), DASP (Ye et al., 2025) and TopoFormer, all from its Table 2. Table 2 shows the same pattern. Against GPS trained under the same protocol (84.19, 81.14, 83.48 and 69.00 on the four datasets), MorseFormer++ gains 7.3 points on BZR, 5.5 on COX2, 7.3 on MUTAG and 10.6 on PROTEINS. Against the published pool the datasets split evenly: TopoFormer leads on BZR and MUTAG by 0.9 and 3.9 points, MorseFormer++ on COX2 and PROTEINS by 2.7 and 1.9 points, where it is the highest value reported in Uddin et al. (2026).

Table 2: **Predictive performance under the TopoFormer protocol.** 10-fold cross-validation accuracy % (↑); mean and standard deviation over folds. Baselines and TopoFormer from Uddin et al. (2026), Table 2. Best per dataset in bold blue.

Method	BZR	COX2	MUTAG	PROTEINS
DMP^{TF}	—	$80.90_{\pm \text{NA}}$	$89.80_{\pm 3.6}$	$75.80_{\pm 3.7}$
FC-V^{TF}	$85.61_{\pm 0.9}$	$81.01_{\pm 0.8}$	$84.00_{\pm 1.6}$	$75.30_{\pm 3.3}$
G-Mix^{TF}	$84.15_{\pm 2.3}$	$83.83_{\pm 2.2}$	$81.96_{\pm 0.6}$	$76.83_{\pm 1.4}$
RGCL^{TF}	$84.54_{\pm 1.7}$	$79.31_{\pm 0.6}$	$87.66_{\pm 1.0}$	$75.03_{\pm 0.4}$
EMP^{TF}	$88.78_{\pm 0.6}$	$85.53_{\pm 1.7}$	—	$78.79_{\pm 0.6}$
$\text{TopoGCL}^{\text{TF}}$	$87.17_{\pm 0.8}$	$81.45_{\pm 0.9}$	$90.09_{\pm 0.9}$	$77.30_{\pm 0.9}$
DASP^{TF}	$89.40_{\pm 1.6}$	$84.80_{\pm 4.7}$	$91.90_{\pm 0.6}$	$77.60_{\pm 0.6}$
TopoFormer	**92.36** ± 4.11	83.93 ± 4.30	**94.68** ± 4.30	77.64 ± 3.64
MorseFormer++	91.47 ± 4.86	**86.65** ± 4.79	90.80 ± 5.87	**79.57** ± 4.30

Molecular benchmarks. On ZINC-12K with edge features (canonical split, MAE) and ogbg-molhiv (OGB scaffold split, ROC-AUC), against the baselines of Eitan et al. (2025) and Bodnar et al. (2021a) that report both columns, MorseFormer++ attains 0.064 ± 0.002 MAE and 79.18 ± 1.37 ROC-AUC (Table 3). On ZINC, SMCN (0.060) and Subgraphormer+PE (0.063) are ahead by 0.004 and 0.001; MorseFormer++ is ahead of every other entry, including CIN (0.079), CIN++ (0.077), the Cellular Transformer (0.080) and its backbone GPS (0.070). On MolHIV, SMCN (81.16) leads, and MorseFormer++ is 1.98 behind it and within one standard deviation of Subgraphormer (Bar-Shalom et al., 2024) and GNN-SSWL+ (Zhang et al., 2023a).

Attribute transport. Section D compares the three aggregators of Section 3.2 with everything else fixed, on ZINC under the cellular Morse complex and on PROTEINS under the simplicial one. DeepSets improves on the fixed projection in both, from 0.070 to 0.064 MAE and from 72.68 to 74.11; attention is the strongest choice on PROTEINS (75.67) and the weakest on ZINC (0.079). The aggregator moves ZINC by 0.015 MAE and PROTEINS by 3 points, and DeepSets is the safe default.

4.2 Structural information in the Morse complex

We evaluate descriptors read from the Morse complex without training on BREC (Wang & Zhang, 2024): 400 pairs of graphs that 1-WL cannot distinguish. A method is credited with a pair when its descriptors for the two graphs differ; trained models follow the BREC protocol and training-free descriptors are compared exactly, as in Ji et al. (2025). More details can be found in Section E.1. The *full fingerprint* concatenates the three. The static descriptors of graph Morse complex separate 97 pairs and the clique lift raises this to 115. Betti curves alone separate 219, more than SpectRe (216), the strongest untrained descriptor reported so far. Together the three concatenated separate 314 pairs, more than every trained model in the table, including I^2-GNN and 3-WL.

Table 4: Pairs of 1-WL-indistinguishable graphs, of 400, whose descriptors differ. Trained models from Wang & Zhang (2024); descriptor baselines from Ji et al. (2025).

Method	**/ 400**
Trained models	
3-WL test	270
Graphormer (Ying et al., 2021)	79
PPGN (Maron et al., 2019)	233
GSN (Bouritsas et al., 2023)	254
$SSWL_P$ (Zhang et al., 2023a)	248
I^2-GNN (Huang et al., 2023)	281
Training-free descriptors	
Standard PH (Ji et al., 2025)	12
RePHINE (Immonen et al., 2023)	164
SpectRe (Ji et al., 2025)	216
Morse-reduced descriptors (this work)	
Static, Morse complex of G	97
Static, clique lift	115
Betti curves, Morse-reduced	219
Full fingerprint	**314**

4.3 Efficiency

We time the same attention kernel on the full lifted and on the Morse-reduced cell sequence of Watts–Strogatz graphs lifted to chordless 2-complexes (Figure 4). Critical cells stay close to the Betti numbers while the lift grows faster than the graph, so the compression grows from $10\times$ at $n = 500$ to $17.5\times$ at $n = 10$K, where the lift has 38,076 cells and the Morse complex 2,174. Since attention is quadratic in the sequence length, the speedup per attention pass grows from $13\times$ to $250\times$: 285 ms over the lift against 1.1 ms over the Morse complex. End to end, MorseFormer++ reaches 0.064 MAE on ZINC at 9.1 s per epoch on one GPU, between GraphGPS (0.070, 7.2 s) and SMCN (0.060, 11.0 s), and MorseFormer alone runs at 3.1 s per epoch (Section B). Lifting and matching cost 1–10 ms per molecular graph, against 88–121 ms for the auxiliary summary (Section C).

5 Conclusion

MorseFormer uses discrete Morse theory to reduce a lifted complex to its critical cells and learns on the resulting Morse complex; MorseFormer++ combines this representation with a graph backbone through a learned gate. Across our experiments the reduced complex retains useful structural information at a fraction of the attention cost. Two limitations point to future work: the descriptors depend on the choice of Morse matching, which an invariant matching would make canonical, and the retraction is detached during backpropagation, which differentiable Morse chain maps (Harker et al., 2014) would remove.

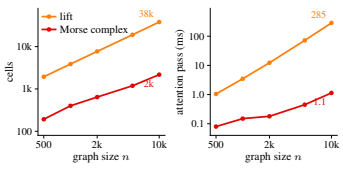

Figure 4: Watts–Strogatz graphs lifted to chordless 2-complexes, n from 500 to 10K: cells of the lift against critical cells of the Morse complex, and the wall-clock of one attention pass over each sequence. Log axes.

REPRODUCIBILITY STATEMENT

All code is released anonymously at `https://anonymous.4open.science/r/morseformer-B21B`: the lifting and Morse-engine modules (three-phase matching, retraction sweep, attribute transport), the MorseFormer and MorseFormer++ models, and one training script per experiment. The matching and the retraction are specified in Section 3.1, Algorithm 1 and Section E.2, and Proposition 4 is proved in Section E.2. Every protocol follows its reference paper, including splits, seeds, metric and model selection: HOPSE/TopoTune (stratified $50/25/25$ splits, five seeds), TopoFormer (10-fold cross-validation), and the canonical ZINC-12K and OGB scaffold splits (Section 4). Baseline numbers are copied from the cited tables; the only baseline we train, GPS, uses the published recipe of Rampášek et al. (2022) under the same protocol. The MorseFormer-specific hyperparameter grid is in Section E, the auxiliary topological summary in Section E.1, and per-stage preprocessing costs in Section C. Training used NVIDIA V100 GPUs, the ZINC timing comparison one RTX 4060, and the preprocessing timings one CPU core.

AI USE STATEMENT

We used large language models to polish and s with coding and plotting. All methods, experiments, results, statements and, written and verified by the authors, who take full responsibility for the content of the paper.

REFERENCES

Amirreza Akbari, Amauri H Souza, and Vikas K Garg. The logical expressiveness of topological neural networks. In *International Conference on Learning Representations (ICLR)*, 2026.

Rubén Ballester, Pablo Hernández-García, Mathilde Papillon, Claudio Battiloro, Nina Miolane, Tolga Birdal, Carles Casacuberta, Sergio Escalera, and Mustafa Hajij. Attending to topological spaces: The cellular transformer. *arXiv preprint arXiv:2405.14094*, 2024.

Linus Bao, Emily Jin, Michael M. Bronstein, İsmail İlkan Ceylan, and Matthias Lanzinger. Homomorphism counts as structural encodings for graph learning. In *International Conference on Learning Representations (ICLR)*, 2025.

Guy Bar-Shalom, Beatrice Bevilacqua, and Haggai Maron. Subgraphormer: Unifying subgraph GNNs and graph transformers via graph products. In *International Conference on Machine Learning (ICML)*, 2024.

Dominique Beaini, Saro Passaro, Vincent Létourneau, William L. Hamilton, Gabriele Corso, and Pietro Liò. Directional graph networks. In *International Conference on Machine Learning (ICML)*, 2021.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

Guillermo Bernárdez, Marco Montagna, Louis Van Langendonck, Martin Carrasco, Amirreza Akbari, Louisa Cornelis, Mathilde Papillon, Pere Barlet-Ros, Nina Miolane, and Lev Telyatnikov. HOPSE: Scalable higher-order positional and structural encoder for combinatorial representations. *arXiv preprint arXiv:2505.15405*, 2025.

Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M. Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. In *International Conference on Learning Representations (ICLR)*, 2022.

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Liò, Guido F. Montúfar, and Michael Bronstein. Weisfeiler and Lehman go cellular: CW networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021a.

Cristian Bodnar, Fabrizio Frasca, Yu Guang Wang, Nina Otter, Guido F. Montúfar, Pietro Liò, and Michael Bronstein. Weisfeiler and Lehman go topological: Message passing simplicial networks. In *International Conference on Machine Learning (ICML)*, 2021b.

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Semih Cantürk, Renming Liu, Olivier Lapointe-Gagné, Vincent Létourneau, Guy Wolf, Dominique Beaini, and Ladislav Rampášek. Graph positional and structural encoder. In *International Conference on Machine Learning (ICML)*, 2024.

Yuzhou Chen, Jose Frias, and Yulia R. Gel. TopoGCL: Topological graph contrastive learning. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures? In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 10383–10395. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/75877cb75154206c4e65e76b88a12712-Paper.pdf.

Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Stefania Ebli, Celia Hacker, and Kelly Maggs. Morse theoretic signal compression and reconstruction on chain complexes. *Journal of Applied and Computational Topology*, 2024.

Yam Eitan, Yoav Gelberg, Guy Bar-Shalom, Fabrizio Frasca, Michael Bronstein, and Haggai Maron. Topological blindspots: Understanding and extending topological deep learning through the lens of expressivity. In *International Conference on Learning Representations (ICLR)*, 2025.

Matthias Fey, Jan-Gin Yuen, and Frank Weichert. Hierarchical inter-message passing for learning on molecular graphs. In *ICML Workshop on Graph Representation Learning and Beyond (GRL+)*, 2020.

Robin Forman. Morse theory for cell complexes. *Advances in Mathematics*, 1998.

Jorge Luiz Franco, Gabriel Duarte, Alexander V. Nikitin, Moacir A. Ponti, Diego Mesquita, and Amauri H. Souza. Differentiable lifting for topological neural networks. In *International Conference on Learning Representations (ICLR)*, 2026.

Fabrizio Frasca, Beatrice Bevilacqua, Michael M. Bronstein, and Haggai Maron. Understanding and extending subgraph GNNs by rethinking their symmetries. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Vikas Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *International conference on machine learning*, pp. 3419–3430. PMLR, 2020.

Lorenzo Giusti, Teodora Reu, Francesco Ceccarelli, Cristian Bodnar, and Pietro Liò. CIN++: Enhancing topological message passing. *arXiv preprint arXiv:2306.03561*, 2023.

Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*, volume 2, pp. 729–734. IEEE, 2005. doi: 10.1109/IJCNN.2005.1555942.

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

Xiaotian Han, Zhimeng Jiang, Ninghao Liu, and Xia Hu. G-Mixup: Graph data augmentation for graph classification. In *International Conference on Machine Learning (ICML)*, 2022.

Shaun Harker, Konstantin Mischaikow, Marian Mrozek, and Vidit Nanda. Discrete Morse theoretic algorithms for computing homology of complexes and maps. *Foundations of Computational Mathematics*, 2014.

Max Horn, Edward De Brouwer, Michael Moor, Yves Moreau, Bastian Rieck, and Karsten Borgwardt. Topological graph neural networks. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=oxxUMeFwEHd.

Yinan Huang, Xingang Peng, Jianzhu Ma, and Muhan Zhang. Boosting the cycle counting power of graph neural networks with I^2-GNNs. In *International Conference on Learning Representations (ICLR)*, 2023.

Johanna Immonen, Amauri H. Souza, and Vikas Garg. Going beyond persistent homology using persistent homology. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Mattie Ji, Amauri H. Souza, and Vikas Garg. Graph persistence goes spectral. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Michael Joswig and Marc E. Pfetsch. Computing optimal Morse matchings. *SIAM Journal on Discrete Mathematics*, 2006.

Tomasz Kaczyński, Marian Mrozek, and Maciej Ślusarek. Homology computation by reduction of chain complexes. *Computers & Mathematics with Applications*, 1998.

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

Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Sihang Li, Xiang Wang, An Zhang, Yingxin Wu, Xiangnan He, and Tat-Seng Chua. Let invariant rationale discovery inspire graph contrastive learning. In *International Conference on Machine Learning (ICML)*, 2022.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations (ICLR)*, 2023.

Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H.S. Torr, and Ser-Nam Lim. Graph inductive biases in transformers without message passing. In *International Conference on Machine Learning (ICML)*, 2023.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Konstantin Mischaikow and Vidit Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry*, 2013.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

Leslie O'Bray, Bastian Rieck, and Karsten Borgwardt. Filtration curves for graph representation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2021.

Mathilde Papillon, Sophia Sanborn, Mustafa Hajij, and Nina Miolane. Architectures of topological deep learning: A survey of message-passing topological neural networks. *arXiv preprint arXiv:2304.10031*, 2023.

Mathilde Papillon, Guillermo Bernárdez, Claudio Battiloro, and Nina Miolane. TopoTune: A framework for generalized combinatorial complex neural networks. In *International Conference on Machine Learning (ICML)*, 2025.

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

Emil Sköldberg. Morse theory from an algebraic viewpoint. *Transactions of the American Mathematical Society*, 2006.

Md Joshem Uddin, Astrit Tola, Cuneyt Gurcan Akcora, and Baris Coskunuzer. TopoFormer: Topology meets attention for graph learning. In *International Conference on Learning Representations (ICLR)*, 2026.

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

Xiyuan Wang, Pan Li, and Muhan Zhang. Graph as point set. In *International Conference on Machine Learning (ICML)*, 2024.

Yanbo Wang and Muhan Zhang. An empirical study of realized GNN expressiveness. In *International Conference on Machine Learning (ICML)*, 2024.

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

Maosheng Yang, Elvin Isufi, and Geert Leus. Simplicial convolutional neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

Wei Ye, Shuhao Tang, Hao Tian, and Qijun Chen. Beyond histogram comparison: Distribution-aware simple-path graph kernels. *IEEE Transactions on Artificial Intelligence*, 2025.

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

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R. Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Bohang Zhang, Guhao Feng, Yiheng Du, Di He, and Liwei Wang. A complete expressiveness hierarchy for subgraph GNNs via subgraph Weisfeiler-Lehman tests. In *International Conference on Machine Learning (ICML)*, 2023a.

Bohang Zhang, Shengjie Luo, Liwei Wang, and Di He. Rethinking the expressive power of GNNs via graph biconnectivity. In *International Conference on Learning Representations (ICLR)*, 2023b.

A Related Work

Higher-order message passing. CIN (Bodnar et al., 2021a), CIN++ (Giusti et al., 2023) and SC-CNN (Yang et al., 2022) pass messages between the cells of a lifted complex, and SMCN (Eitan et al., 2025) extends this family to close expressivity gaps. GCCN (Papillon et al., 2025) runs a base GNN on the Hasse graph of the lift, and HOPSE (Bernárdez et al., 2025) derives positional and structural encodings from decompositions of that graph. All of them learn on the lifted complex. MorseFormer learns on its Morse complex: the lift exposes higher-order structure, and the representation is built on the critical cells that survive the reduction.

Attention over higher-order structure. Attention-based methods differ in what becomes a token. The Cellular Transformer (Ballester et al., 2024) attends over every cell of the lift, within and across ranks through the incidence relations or over all cells at once, with positional encodings computed on the lifted complex, among them barycentric-subdivision, random-walk and Slepian encodings and the eigenvectors of the Hodge Laplacian of the lift. Its authors note that attention is quadratic in the number of cells and that the model is therefore limited to complexes with few cells. TopoFormer (Uddin et al., 2026) attends over a short sequence of tokens derived from a filtration. Graph transformers keep the tokens at the nodes and inject structure through encodings or biases: spectral encodings in SAN (Kreuzer et al., 2021), centrality and distance biases in Graphormer (Ying et al., 2021; Zhang et al., 2023b), relative random-walk encodings in GRIT (Ma et al., 2023), a

point-set encoder in PST (Wang et al., 2024), homomorphism counts in MoSE (Bao et al., 2025), and a local message-passing block next to the attention in GPS (Rampášek et al., 2022). The tokens of the Morse encoder are the critical cells of the Morse complex, across ranks. The graph backbone keeps its node tokens, and any of these transformers could serve as that backbone.

Topological descriptors and discrete Morse theory. Persistent homology adds structural signal to graph learners (Immonen et al., 2023; Ji et al., 2025); its cost usually confines it to vertex and edge filtrations computed as preprocessing. MorseFormer uses such descriptors only in the auxiliary summary of Section 3.5. Discrete Morse theory has long served to simplify complexes and to accelerate homology computation (Forman, 1998; Sköldberg, 2006; Mischaikow & Nanda, 2013; Harker et al., 2014); its closest use in machine learning is the Morse projection of Ebli et al. (2024), which compresses signals on chain complexes. Here the reduced complex is the domain on which a representation is learned.

B Accuracy–cost trade-off on ZINC

Figure 5 places MorseFormer++ and MorseFormer on the ZINC accuracy/cost frontier discussed in Section 4.3: test MAE against complete per-epoch wall-clock time, all methods timed on the same GPU with their published configurations.

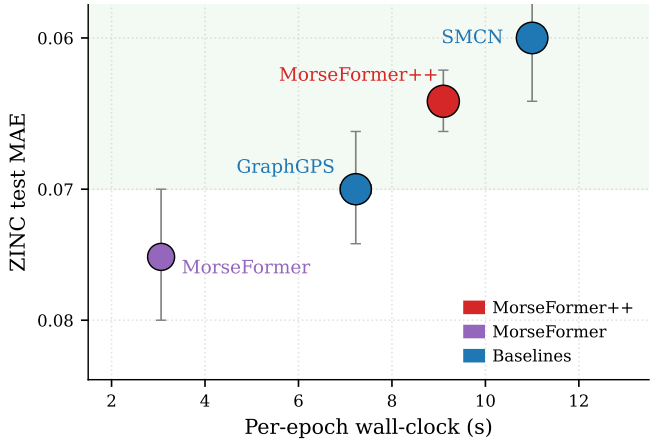

Figure 5: **Accuracy–cost trade-off on ZINC.** Test MAE against per-epoch wall-clock time on the same GPU (RTX 4060, batch 32), published configurations for the baselines; lower is better on both axes. The shaded band marks MAE below 0.07.

C Runtime of the topological pipeline

Table 5 times the three preprocessing stages the data loader runs on an unseen graph (no caching): chordless-cycle lifting, the three-phase Morse matching with retraction, and the topological readout (persistence, Betti, and Hodge blocks). Means are over the test split (twelve graphs for PROTEINS), measured on one CPU core of a lab workstation (NVIDIA TITAN RTX node); all stages are CPU-bound and cached after the first pass, so they are one-shot per graph. Lifting and Morse reduction cost 1–10 ms per molecular graph; the readout dominates throughout. PROTEINS is the outlier that explains its absence from the cellular rows of Table 1: its larger, denser graphs push the Morse stage to 4.1 s per graph, roughly $40\times$ the NCI1 pipeline.

D Critical-cell aggregation ablation

The basin-wise aggregator of Section 3.2 maps cell attributes X_k to critical-cell attributes X_k^{morse}. We compare the three variants introduced there, Fixed ($X_k^{\mathrm{morse}} = \phi_k X_k$), DeepSets (Zaheer et al., 2017) and multi-head Attention, keeping every other choice fixed. ZINC is ablated under the cellular Morse

Table 5: Per-graph cost (ms) of the topological pipeline, mean over the test split; chordless lifting with $\ell_{\max} = 6$ (18 for ZINC-12K).

Dataset	mean $\lvert V\rvert$	**Lift**	**Morse**	**Readout**	**Total**
MUTAG	18	1.7	7.1	88.2	97.0
NCI1	26	1.7	9.2	117.9	128.8
NCI109	27	1.7	9.7	121.4	132.9
ZINC-12K	23	1.5	8.2	106.3	116.0
PROTEINS	68	23.3	4054.9	586.6	4664.8

Table 6: **Attribute-transport ablation.** ZINC-12K (cellular Morse complex, MAE ↓) and PROTEINS (simplicial Morse complex, accuracy % ↑), five seeds each. Only the basin aggregator changes.

Aggregator	**ZINC ↓**	**PROTEINS ↑**
Fixed ($\phi_k X_k$)	$0.070_{\pm 0.003}$	$72.68_{\pm 3.60}$
DeepSets (Zaheer et al., 2017)	$\mathbf{0.064}_{\pm 0.002}$	$74.11_{\pm 4.06}$
Attention (multi-head)	$0.079_{\pm 0.005}$	$\mathbf{75.67}_{\pm 2.67}$

complex and PROTEINS under the simplicial one, the liftings that give each dataset its strongest result in Section 4.1.

DeepSets improves on Fixed in both regimes ($0.070 \to 0.064$ on ZINC; $72.68 \to 74.11$ on PROTEINS). On PROTEINS, multi-head Attention is the strongest of the three; on ZINC it is worse than Fixed. The two regimes differ in task (regression vs. classification) and in lifting (cellular vs. simplicial Morse complex). The best aggregator is regime-dependent, and DeepSets is the safer default.

E Implementation Details

Code. The full training code, lifting and Morse-engine modules, and per-experiment training scripts are released at `https://anonymous.4open.science/status/morseformer-B21B`.

Hyperparameter grid. The GPS backbone (local MPNN, positional encodings, optimizer, learning rate, scheduler, batch size, epochs) follows the configurations of Rampášek et al. (2022), available in their public repository. We sweep the following MorseFormer-specific axes:

- **Lifting.** *graph* (no lifting); *cycle basis*, *chordless* (induced cycles) and *clique complex* with max ring size $\ell_{\max} \in \{6, 18\}$; *∂lift* (Franco et al., 2026) with backbone $\in \{$cycle_basis, induced_cycles, $\}$, $k_{\max} \in \{6, 8\}$.
- **Critical-cell aggregation.** Fixed ($\phi_k X_k$); DeepSets (Zaheer et al., 2017); Attention.
- **Morse encoder.** Morse layers $L_M \in \{2, 4\}$; cross-rank attention enabled/disabled.

Hardware For all benchmarking experiments, we use a cluster with Nvidia V100 GPUs — details regarding the compute infrastructure are omitted for anonymity.

E.1 Topological readout details

This section specifies the auxiliary graph-level summary z_{topo} of Section 3.5. It is computed on the Morse complex, once per graph under a static lifting, and enters the predictor next to the fused representation h. Concrete values of the hyperparameters introduced below (K, T, E, D_∞) are listed alongside the rest of the MorseFormer-specific hyperparameter grid.

Filtration construction. The edge filtration assigns each edge $e = (u, v)$ the weight

$$w(u, v) \;=\; 4 \;-\; \deg(u) \;-\; \deg(v) \;+\; 3 \cdot |N(u) \cap N(v)|,$$

where $|N(u) \cap N(v)|$ is the number of triangles incident to e. Edges are inserted in ascending order of w to obtain $\emptyset = \mathcal{X}_0 \subset \mathcal{X}_1 \subset \cdots \subset \mathcal{X}_{|E|} = \mathcal{X}$. When the lifting includes 2-cells, cycles up

to length $\ell_{\max}$ are added as 2-cells once their bounding edges have all been inserted, coupling the filtration's maximum ring size to that of the lifting.

Persistence-diagram block. Let $\mathrm{PD}_k(\mathcal{X}) = \{(b_i, d_i)\}_{i=1}^{m_k}$ be the rank-k persistence diagram of size m_k. Infinite deaths are replaced by a fixed constant D_∞ chosen to lie outside the normalised finite-death range. Per rank, $\mathbf{v}_k$ concatenates the top-K (b, d) pairs ranked by lifetime $\ell_i = d_i - b_i$ (zero-padded when $m_k < K$) with summary statistics over the diagram — mean and standard deviation of lifetimes, maximum lifetime, mean birth, mean death, and the count m_k. Stacking across ranks yields $\mathbf{v}_{\text{PD}}$.

Betti-curve block. $\mathbf{v}_\beta$ samples the rank-k Betti numbers $\beta_k(t) = \dim\ker \tilde{\Delta}_k^{(t)}$ at T uniformly spaced filtration thresholds, Morse-reducing each $\mathcal{X}_t$ independently.

Static block. $\mathbf{v}_{\text{stat}}$ records, per rank, the Betti number $\beta_k(\mathcal{X})$ of the original complex together with the E smallest eigenvalues of the rank-k reduced Hodge Laplacian $\tilde{\Delta}_k$ on the Morse complex.

Boundary-statistics block. $\mathbf{v}_{\text{bdry}}$ stores, per rank, summary statistics over the nonzero entries of the Morse boundary $\tilde{B}_k$: nonzero count, mean magnitude, maximum magnitude, and standard deviation of magnitudes.

Total readout. Concatenating the four blocks yields

$$z_{\text{topo}} \;=\; \big[\, \mathbf{v}_{\text{PD}};\; \mathbf{v}_\beta;\; \mathbf{v}_{\text{stat}};\; \mathbf{v}_{\text{bdry}} \,\big],$$

fed to the final MLP together with the fused representation h of Equation (8).

E.2 Matching algorithm details

This appendix gives per-phase analysis and pseudocode for the three-phase matching of Section 3.1.

Phase 1 (free-face coreduction). This is the coreduction routine of Harker et al. (2014, Alg. 3.6); each addition preserves acyclicity (Kaczyński et al., 1998). The phase runs in time linear in the total Hasse-diagram size.

Phase 2 (spanning-tree pairing). BFS from each connected-component root touches every vertex and every 1-cell at most once, so it terminates in $O(|V| + |E|)$ time. When Phase 1 has not consumed any tree edge, the only unmatched 0-cells after Phase 2 are the BFS roots, so $c_0 = \beta_0$. This initialisation is the piece tailored to graph liftings.

Phase 3 (V-path cancellation). This is Forman's classical V-path cancellation (Forman, 1998, §10). Each path reversal preserves acyclicity of the gradient graph and reduces $\sum_k c_k$ by 2.

Retraction assembly. The recursion for $\Sigma^{(k)}$ is solved by topological-sort dynamic programming over the gradient graph G_μ restricted to matched-lower cells. Per dimension, the sweep runs in $O(c_k \cdot \mathrm{nnz}(B_*))$ time and $O(c_k \cdot n_k)$ memory (Harker et al., 2014).

Proof of Proposition 4. Every off-diagonal entry $\Phi_k[i, j]$ records a length-two alternating V-path in G_μ; a nonzero matched-to-matched entry from j to i corresponds to an arc $i \to j$ of the matched-cell digraph. Acyclicity makes that digraph a DAG, so in a topological order the matched block of Φ_k is strictly lower-triangular with zero diagonal, which is (i). A strictly triangular $(n_k - c_k)$-square block is nilpotent of index $\le n_k - c_k$, so the entries of Φ_k^m—signed sums of gradient V-paths, whose length is bounded by $n_k - c_k$—stop changing once $m \ge n_k - c_k$; the limit Φ_k^∞ therefore exists and, evaluated in reverse topological order, is exactly Sköldberg's splitting homotopy (Sköldberg, 2006), i.e. the recursion of Definition 3. Each cell touches only its incident boundary entries, so the sweep costs $O(c_k \cdot \mathrm{nnz}(B_*))$, which is (ii). The dense iteration and the sweep converge to the same operator Φ_k^∞ and hence extract identical $\phi_k, \psi_k, \tilde{B}_k$, which is (iii). □

Algorithm 1 Three-phase acyclic matching with Morse retraction.

Require: Chain complex (C_*, B_*)
Ensure: Acyclic matching μ, critical index sets $\{\mathcal{I}_k\}$, retraction ϕ_k, inclusion ψ_k, Morse boundary $\tilde{B}_k$

$\mu \leftarrow \emptyset$
// Phase 1: free-face coreduction
while $\exists$ unmatched cell σ with unique unmatched coface τ **do**
    $\mu \leftarrow \mu \cup \{(\sigma, \tau)\}$
end while
// Phase 2: spanning-tree pairing on 1*-skeleton*
for each connected component **do**
    Run BFS from a root vertex r
    for each non-root v with parent edge e_v **do**
        if v and e_v are both unmatched **then**
            $\mu \leftarrow \mu \cup \{(v, e_v)\}$
        end if
    end for
end for
// Phase 3: V-path cancellation
while $\exists$ critical $(k{+}1)$-cell τ and critical k-cell σ joined by a unique directed path in G_μ **do**
    Reverse that path
end while
$\mathcal{I}_k \leftarrow$ unmatched k-cells
// Retraction assembly: Sköldberg DP
for $k = 0, 1, \ldots$ **do**
    Solve $\Sigma^{(k)}$ via topological-sort DP over matched-lower k-cells
    $\Phi_k^\infty \leftarrow I - B_{k+1}\Sigma^{(k)} - \Sigma^{(k-1)}B_k$
    $\phi_k \leftarrow \Phi_k^\infty[\mathcal{I}_k, :], \quad \psi_k \leftarrow \Phi_k^\infty[:, \mathcal{I}_k]$
    $\tilde{B}_k \leftarrow \phi_{k-1}\, B_k\, \psi_k$
end for