RESTORING COMPONENT LOCALITY IN SPECTRAL POSITIONAL ENCODINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spectral positional encodings should preserve the representation of a component when an unrelated component is added. A fixed-rank prefix of a merged Laplacian spectrum can violate this principle: the added component competes for the same eigenvector slots. We formalize *Component Locality*, prove that a global prefix cannot guarantee it, and provide a sufficient direct-sum criterion. Component-local selection restores the identity by assigning each component its own spectral quota. We test this mechanism with shared checkpoints and a component-local downstream path on real graphs placed in controlled disjoint unions. CL-SPE restores encoding, hidden-state, and prediction locality in an independent holdout across all seven main cells; six cells show practically significant prediction-drift reductions. BudgetMatched and ClusterComplete separate total budget, allocation, and exact-eigenspace effects. An additional USPTO study establishes the mechanism on natural reaction graphs: 49.17% of nonboundary primary training reactions exhibit displacement under BudgetMatched. Component-local selection reduces reaction-macro kernel error from 0.206 under BudgetMatched to numerical zero and restores coverage to 1. Reaction-center prediction meets the overall noninferiority criterion on all 3,512 primary test reactions. Together, controlled interventions and natural reaction graphs establish component-wise allocation as a concrete design principle for spectral positional encodings.

1 INTRODUCTION

Laplacian eigenvectors provide structural coordinates for graph neural networks and graph transformers (Dwivedi & Bresson, 2021; Kreuzer et al., 2021; Rampášek et al., 2022; Huang et al., 2024). Their computation often ends with a fixed-rank truncation. On a disconnected input, this truncation allocates a shared budget across a merged spectrum. An unchanged component can therefore receive different coordinates.

Chemical reactions provide a natural example: source molecules form separate components before new bonds form (Jin et al., 2017). Controlled disjoint unions expose the same issue by evaluating a target C with an unrelated graph D. Our analysis concerns inputs eigendecomposed as one matrix. Spectra precomputed separately for each graph retain their original scope when collated into a mini-batch.

We call the desired identity *Component Locality*: the representation of C inside $C \sqcup D$ agrees with its standalone representation under the known node correspondence. We study this identity at the selected eigensystem, encoder, and prediction layers while holding the target fixed.

Theorem 1 proves a global-prefix counterexample and gives a sufficient direct-sum criterion. A sufficiently long path contributes enough small positive eigenvalues to displace every target mode from a finite global prefix. Component-wise selection preserves each component's quota. Away from cutoff boundaries, displacement d_C determines kernel error through $\mathrm{KLE} = \sqrt{d_C/k_C}$, where $k_C = \min(k, |C| - 1)$.

Our controlled study trains on connected graphs, where the selectors coincide. GlobalSelect-LocalSPE and CL-SPE then share a checkpoint, component-local aggregation, attention, and target readout. Changing selector scope isolates its effect. BudgetMatched, ClusterComplete, and contextual-SPE interventions separate budget, allocation, and aggregation effects. The natural USPTO study retains

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.

reaction-wide attention while assigning spectral coordinates within each molecule. It tests allocation and complete training-algorithm behavior on naturally disconnected graphs (Appendix N).

Contributions.

1. A Component Locality formulation, a global-prefix counterexample, and a sufficient direct-sum criterion for spectral encoding.
2. A shared-checkpoint study with independent holdout evidence of encoding, hidden-state, and prediction locality across seven cells.
3. Budget, quota, context, and eigenspace-completion interventions that identify distinct sources of spectral coupling.
4. Natural reaction-graph evidence with matched actual budgets, reaction-level statistics, and prediction equivariance under node permutations.

2 Component Locality

2.1 Spectral coordinates and selector scope

For an undirected graph with nonnegative edge weights, components follow the binary support and $L_{\mathrm{sym}} = I - D^{-1/2}AD^{-1/2}$. PyG uses the isolate block 1; zero-diagonal isolates provide a sensitivity check. A spectral encoding selects eigenpairs, applies an eigenvalue map φ, and aggregates through ρ (Dwivedi & Bresson, 2021; Lim et al., 2022; Huang et al., 2024).

Global-Positive-k removes structural zeros and keeps the first $\min(k, n_+)$ positive modes. PyG 1-skip keeps sorted columns $1, \ldots, k$ after the first (Appendix E). A complete repeated eigenspace has a basis-invariant projector; splitting its cutoff makes the selected subspace basis-dependent.

2.2 The locality identity

For a connected target C and a node-disjoint D with no cross-edge, $G = C \sqcup D$ supplies a known node correspondence π. Component Locality requires

$$F(C \sqcup D)\big|_{\pi(C)} = F(C). \tag{1}$$

Selected eigensystem. Let U_C contain standalone target modes and $U_{G\to C}$ contain union modes restricted to $\pi(C)$. We compare $K_C = U_C U_C^\top$ and $K_{G\to C} = U_{G\to C} U_{G\to C}^\top$, identifying signs and bases of complete exact eigenspaces.

Encoder and model. Aligned PE, hidden tensors, and predictions are compared directly at tolerances fixed before the independent locality test (Appendix I). The target's nodes, features, and edges stay fixed. Full-model locality also requires component-local spectral aggregation, attention, and readout.

2.3 Direct sums and spectral allocation

The identity $L(C \sqcup D) = L(C) \oplus L(D)$ makes the spectrum a multiset union (Figure 1). Away from cutoff boundaries, the restricted retained kernel projects onto a local positive prefix. With $k_C = \min(k, |C| - 1)$, retained rank r_C, and $d_C = k_C - r_C$,

$$\mathrm{KLE} = \frac{\|K_C - K_{G\to C}\|_F}{\|K_C\|_F} = \sqrt{d_C/k_C}, \qquad \text{Coverage} = 1 - d_C/k_C.$$

Completing an exact cutoff eigenspace preserves basis invariance and the distinction between global and local allocation.

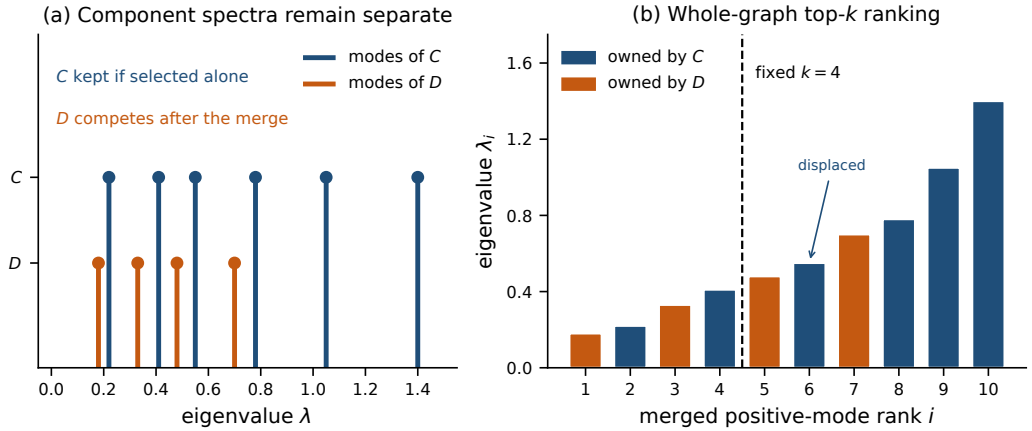

Figure 1: Spectrum merging. A whole-graph prefix ranks the merged component eigenvalues: modes of D can occupy slots used by C alone. Vertical positions distinguish components in (a); bar heights show eigenvalues in (b).

2.4 A SUFFICIENT DIRECT-SUM CRITERION

Direct-sum-compatible selector. Write $S(L) = (U_L, \Lambda_L)$ for selected orthonormal eigenvectors and their diagonal eigenvalue matrix. The selector is direct-sum-compatible if

$$S(L \oplus M) \sim (U_L \oplus U_M, \Lambda_L \oplus \Lambda_M). \tag{2}$$

Here $\sim$ is defined component-wise under the known direct-sum decomposition. It permits eigenpair reordering, sign changes, and rotations within complete exact eigenspaces of the same component, without mixing the L and M blocks. If component identity is discarded, basis-free equivalence additionally requires completing the full coincident eigenspace across blocks. The shared element-wise map assigns equal weights within each exact eigenspace.

Theorem 1 (Global prefixes and Component Locality)**.** *Fix $k \geq 1$. Let S_k^{pref} discard structural zeros, sort the remaining Laplacian eigenpairs by nondecreasing eigenvalue, and retain exactly $\min(k, n_+)$ modes, breaking ties arbitrarily.*

1. *There exist connected C, D whose selected union kernel restricted to C differs from its standalone kernel.*
2. *For every connected C with $k_C \geq 1$, a sufficiently long path D has at least k positive eigenvalues strictly below the* first *positive eigenvalue of C. The union prefix then has $r_C = 0$ and $d_C = k_C$.*
3. *A fixed scalar function applied to the full spectrum satisfies $f(L \oplus M) = f(L) \oplus f(M)$.*
4. *A direct-sum-compatible selector, the shared element-wise φ and component-local ρ of Section 3.1, and component-local attention and readout satisfy Equation* (1)*.*

Appendix A gives the proof and supporting propositions. The theorem establishes a constructive global-prefix counterexample and a sufficient condition for restoring locality.

3 COMPONENT-LOCAL SPECTRAL ENCODING

3.1 THE SELECTOR INTERVENTION

The controlled comparison varies the matrix on which eigenpairs are selected (Figure 2). *GlobalSelect-LocalSPE* selects the first k positive modes of the whole union, then restricts spectral aggregation and all downstream computation to each component. *CL-SPE* solves each non-isolated component separately, discards its structural zero, and retains $k_c = \min(k, |c| - 1)$ positive modes.

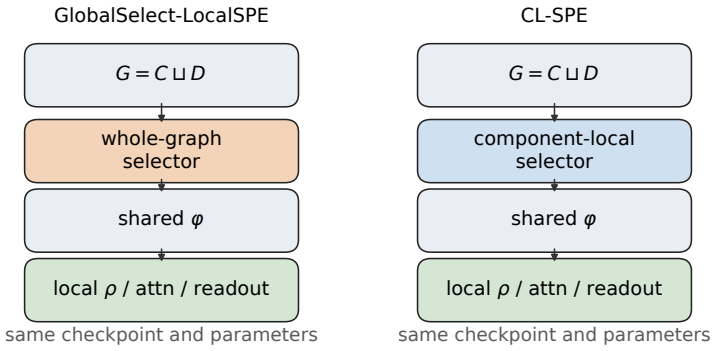

Figure 2: Controlled comparison. Both arms share φ, component-local ρ, component-block attention, target-only readout, parameters, and checkpoint. The intervention changes selector scope.

Isolates receive zero PE with an inactive mode mask. Both arms use the same element-wise eigenvalue map φ, component-local aggregation ρ, and model parameters.

Masked SPE computation. For component c, pad its selected eigenpairs to width K: $U_c \in \mathbb{R}^{n_c \times K}$, eigenvalues $\lambda_c \in \mathbb{R}^K$, and mode mask $m_c \in \{0,1\}^K$. In the Global arm, these vectors are the selected union vectors restricted to c. For each of $a = 1, \ldots, 32$ channels,

$$\mathcal{K}_{c,a} = U_c \,\mathrm{diag}(m_c \odot \varphi_a(\lambda_c))\, U_c^\top, \qquad \mathrm{PE}|_c = \mathbf{1}_{\{\|m_c\|_1 > 0\}}\, \rho(\mathcal{K}_{c,1:32}, b_c)\big|_{\mathrm{valid}}, \tag{3}$$

where the same MLP $\varphi : \mathbb{R} \to \mathbb{R}^{32}$ acts separately on each eigenvalue. The tensor $\mathcal{K}_c$ has shape $n_c \times n_c \times 32$. The node mask b_c removes padding in the two-layer component-local propagation ρ, yielding 32 features per valid node. Component keys scatter these outputs back to union order. Inactive modes contribute zero and isolates return zero PE. Appendix B gives the full recurrence and configuration.

Training for the controlled study uses connected graphs, on which the selectors coincide. The same checkpoint then evaluates the standalone target and its disjoint unions. This design attributes the resulting PE, hidden-state, and prediction differences to spectral selection.

3.2 Budget and Eigenspace Interventions

BudgetMatched. The whole-graph selector retains $K_{\mathrm{total}} = \sum_c k_c$ modes. This matches the total number of component-local modes while allowing the merged eigenvalue order to allocate them among components. Comparing Global-Positive-k, BudgetMatched, and CL separates total budget from per-component allocation.

ClusterComplete. When a cutoff intersects an exact eigenvalue cluster, the selector retains the complete block. Its projector is invariant to rotations within that eigenspace (Proposition 4). The global and component-wise versions retain their respective selector scopes. For distinct nearby eigenvalues, evaluation uses the actual perturbed eigensystem.

3.3 Context and Reaction-Wide Interaction

A contextual-SPE study applies four nested interventions at two dataset–backbone cells: A uses whole selection, global φ, and whole ρ; B localizes ρ; C also localizes φ; and D uses component-wise selection. The sequence $A \to B \to C \to D$ measures a fixed intervention order. The $C \to D$ step holds the contextual encoder fixed while changing selector scope.

The USPTO application retains component-local SPE and enables GraphGPS attention across molecules within a reaction. Reaction boundaries isolate different examples in a batch. Each method is trained with its own selector under a common architecture and training settings. This design measures the behavior of complete training algorithms on natural disconnected inputs. The shared-checkpoint study supplies the causal selector comparison.

4 EXPERIMENTAL DESIGN

Controlled evidence. Synthetic evaluation covers six graph families, random and conditioned pairings, $k \in \{8, 16, 32\}$, and one or four distractors. The main synthetic comparison uses $k = 16$ and one distractor. The main real-graph evaluation uses CLUSTER, MNIST, PascalVOC-SP, and Peptides-struct: seven dataset–backbone combinations, called cells, with five SPE seeds each. PATTERN, Peptides-func, and ZINC provide five additional cells with three seeds each. Models train and validate on connected graphs. Each main cell contains up to 1,000 connected, nonboundary test targets sampled within label-by-size strata; PascalVOC-SP has 999.

Pairing and identity. Each target is paired with four ordinary distractors, shared across selectors and seeds. Metrics first average the four pairings within a target. The main comparison excludes cutoffs that split an eigenvalue cluster in either the target or union. An independent holdout tests locality at previously calibrated tolerances, using the seed-0 checkpoint in each cell. Ordinary, simultaneous four-distractor, and duplicate unions provide additional stress tests. Isolates, tiny components, bridge edges, and near-degenerate cases provide sensitivity analyses. Comparisons use known node correspondences, shared checkpoints, target-only readout, and no cross-edges. All sampled targets are evaluated in the seven main cells.

Metrics and inference. KLE is unsquared relative Frobenius kernel error. Coverage measures overlap with the component-local eigenspace. PE and hidden drift use the same relative norm on aligned target tensors. PredDrift is Jensen–Shannon divergence for classification and scaled absolute deviation for regression. A practical prediction-drift reduction requires mean Global PredDrift $> 10^{-5}$, a CL reduction of at least 80%, a paired 95% interval below zero, at least four of five seeds in the same direction, and at least 95% of target pairs evaluated. Identity tolerances are $\tau_m = \max(\text{preset}, 5 \times \text{replicate-noise } p_{99})$; the PE and prediction presets are 10^{-6}. Appendix C gives metric formulas and bootstrap details.

Natural reaction graphs. The USPTO study uses 40,008 training, 5,001 validation, and 5,007 test reactions from canonicalized USPTO-50K (Schneider et al., 2016). Source molecules define the input graph; product atom correspondences define reactive-atom labels offline. Eligible reactions have both supervised classes. Primary reactions also contain at least two non-isolated source molecules. NoPE, Global16, BudgetMatched (BM), and CL each use five seeds. OGB atom and bond features (Hu et al., 2020) exclude atom-map IDs. The shared GraphGPS backbone has six layers, hidden width 128, and eight attention heads. Training uses reaction-balanced masked binary cross-entropy.
Mechanism analysis uses nonboundary primary reactions. Task improvement is tested on their BM-displaced subset, where BM retains fewer modes than the local quota on at least one reactive component. Overall noninferiority uses all 3,512 primary test reactions, including boundary cases. Paired bootstrap resampling acts on reactions while holding the five seeds fixed. Before test evaluation, we set the noninferiority margin to 0.005 AUPRC, the largest practically acceptable degradation of 0.5 percentage points. Appendix N gives the populations, training settings, statistical comparisons, and leakage sensitivities.

5 RESULTS

5.1 WHOLE-GRAPH SELECTION VIOLATES COMPONENT LOCALITY

At $k = 16$ with one random distractor, cycle, ER, grid, path, and SBM each have displaced, nonboundary cases in 59–64 of their 64 draws (Appendix D). Among 384 ordinary cases, sorting the component spectra predicts the selected-mode allocation at rate 1.0 whenever both cutoffs are non-boundary. Global-Positive-k has mean $\mathrm{KLE} = 0.6887$ and mean target coverage 0.5199. The public PyG 1-skip rule, which skips only the first eigencolumn, has mean $\mathrm{KLE} = 0.7357$ and coverage 0.4852. CL-LapPE-k has mean $\mathrm{KLE} = 1.58 \times 10^{-17}$ and coverage 1.0.

The conditioned pairs are grouped by d_C, k_C, and cutoff boundaries (Figure 3b). At $d_C = 0$ ($n = 80$), Global has unit coverage and numerical-zero KLE. Mean Global KLE rises to 0.470 at mild displacement ($d_C \approx 3.56$) and 0.794 at strong displacement ($d_C \approx 10.09$). CL retains unit coverage throughout these fixed-k, nonboundary strata.

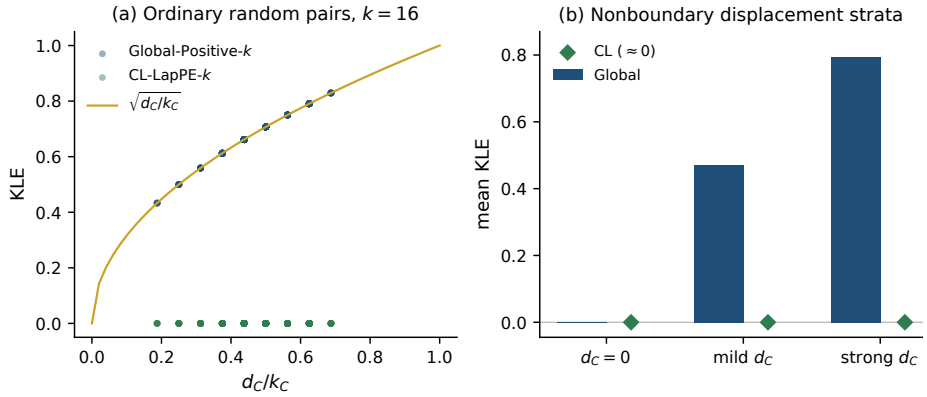

Figure 3: Global kernel error grows with displacement; CL preserves target coverage. (a) Ordinary random pairs at $k = 16$: Global tracks $\sqrt{d_C/k_C}$; CL sits at numerical zero. (b) Displacement strata: Global mean KLE rises from 0 ($d_C = 0$) to 0.47 (mild) and 0.79 (strong); CL KLE remains zero (coverage remains one). Green diamonds mark CL at numerical zero.

5.2 Component-local selection restores encoding and end-to-end locality

On the independent holdout, all seven cells satisfy the calibrated locality tolerances for fresh eigensolves, node permutations, and CL unions, using their seed-0 checkpoints (Appendix I). PE and hidden drift on holdout unions are zero; the largest prediction drift is 1.13×10^{-7}.

The holdout uses 40 previously unused standalone validation graphs and 30 validation union pairs per cell. With the cache disabled, each standalone graph undergoes 10 repeated and 10 node-permuted fresh eigensolves. Each check requires $p_{99} \leq \tau_m$ and maximum drift $\leq 10\tau_m$, with tolerances fixed before holdout evaluation.

Ordinary, four-distractor, and duplicate unions also satisfy these PE, hidden, and prediction tolerances in $7/7$ cells, extending the locality checks across different forms of disconnected input. NoPE and RWSE remain local on the same unions. The five additional benchmark cells each reduce PredDrift by more than 99.99% at a shared checkpoint (Appendix F), and untrained synthetic SPE has nonzero Global PEDrift and zero CL PEDrift (Appendix D).

5.3 Same-checkpoint prediction locality

Table 1: Prediction drift at a shared checkpoint. Each cell averages four distractors per target and five seeds. Pred. indicates the practical reduction defined in Section 4; Id. indicates locality within calibrated tolerances on the independent seed-0 holdout (Appendix I). Floor denotes drift below the practical threshold. Prediction estimates use all five seeds.

Cell	Global	CL	Red.	95% CI (CL−GS)	Pred.	Id.
CLUSTER/GINE	3.69×10^{-10}	0	100%	$[-3.82, -3.56]\times10^{-10}$	floor	yes
CLUSTER/GPS	0.00230	0	100%	$[-0.00257, -0.00204]$	yes	yes
MNIST/GINE	0.00115	0	100%	$[-0.00162, -0.00075]$	yes	yes
PascalVOC-SP/GINE	0.0298	0	100%	$[-0.0305, -0.0290]$	yes	yes
PascalVOC-SP/GPS	0.00484	0	100%	$[-0.00506, -0.00463]$	yes	yes
Peptides-struct/GINE	0.902	0	100%	$[-0.920, -0.885]$	yes	yes
Peptides-struct/GPS	0.710	0	100%	$[-0.730, -0.693]$	yes	yes

CL produces a practical prediction-drift reduction on all four datasets and $6/7$ cells, each with $5/5$ seeds aligned (Table 1). CL PredDrift is numerical zero throughout. CLUSTER/GINE has Global drift 3.69×10^{-10}, below the practical threshold. Absolute loss drift improves practically in $6/7$ cells; signed task loss meets the improvement criterion in $4/7$ (Table 2). Locality restoration therefore provides task-loss stability with cell-specific gains in average loss. At $k \in \{8, 32\}$, CLUSTER/GPS and Peptides-struct/GINE retain zero CL PredDrift in all 12 runs across three seeds per setting; Global drift stays positive (Appendix K).

Table 2: Task-loss stability at the shared checkpoints. Both contrasts are CL minus Global; negative values favor CL. Stars mark effects meeting both statistical and practical criteria; full q values are in Appendix J.

Cell	Signed loss contrast	Absolute loss-drift contrast
CLUSTER/GINE	-8.3×10^{-11}	-5.8×10^{-9}
CLUSTER/GPS	-0.00186	$-0.0328^{\star}$
MNIST/GINE	-0.00396	$-0.0284^{\star}$
PascalVOC-SP/GINE	$-0.0708^{\star}$	$-0.137^{\star}$
PascalVOC-SP/GPS	$-0.0147^{\star}$	$-0.0698^{\star}$
Peptides-struct/GINE	$-0.671^{\star}$	$-0.689^{\star}$
Peptides-struct/GPS	$-0.507^{\star}$	$-0.521^{\star}$

5.4 What causes the effect?

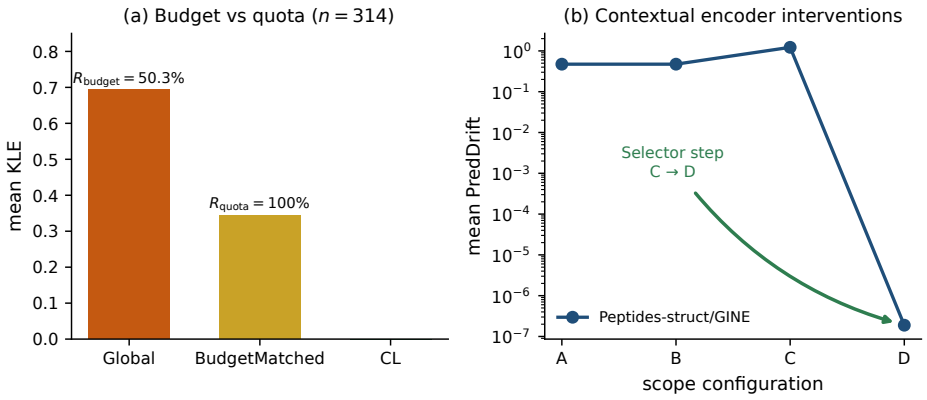

Figure 4: Mechanism split. (a) Nonboundary pairs without isolates ($n = 314$): BudgetMatched halves Global KLE; CL removes the rest in this construction. (b) Nested intervention trace on Peptides-struct/GINE: the selector step $C \to D$ removes remaining PredDrift.

Budget versus quota. On the ordinary, double-nonboundary, non-isolate, paired-complete subset ($n = 314$), mean KLE is 0.693 (Global), 0.344 (BudgetMatched), and 1.60×10^{-17} (CL). Mode counts of BudgetMatched and CL match at rate 1.0. BudgetMatched recovers $R_{\text{budget}} = 50.3\%$ of the Global kernel error; component quotas remove $R_{\text{quota}} = 100\%$ of the remainder (Figure 4a; Appendix C).

Nested intervention trace. Peptides-struct/GINE has $A = B = 0.473$, $C = 1.224$, and $D = 1.89 \times 10^{-7}$ (Figure 4b). The $B \to C$ increase shows the interaction of eigenvalue context and global selection. The selector step $C \to D$ removes the residual: $R_{\text{sel}} = 1.224$, 95% CI $[1.116, 1.346]$, $5/5$ seeds aligned, and a 100% reduction. CLUSTER/GPS has $5/5$ aligned seeds and a 99.8% selector reduction. These contrasts follow the fixed intervention order (Appendix K).

Training with disconnected inputs. On Peptides-struct/GINE, training with 50% union samples reduces Global PredDrift from 0.905 under connected-only training to 0.205, while CL drift is zero under both training conditions. In this ablation, union augmentation reduces the drift and component-local selection removes its remaining cross-component dependence (Appendix K).

Architecture scope. At the same checkpoints, component-block attention keeps CL PredDrift at zero, while Global gives 0.00176 on CLUSTER/GPS and 0.379 on Peptides-struct/GINE. Graph-wide attention or normalization adds a separate cross-component mixing path. The component-local downstream stack therefore isolates the selector effect (Appendix K).

5.5 Degeneracy boundary

Exact cutoff degeneracy. Full-block Haar rotations cover 69 certified cases and 32,620 rotations. The maximum rotation error is 8.31×10^{-16}. ClusterComplete local-versus-union projector error is

at most 5.18×10^{-13}, and CL KLE/PE remain inside tolerance. Fixed-k truncation that splits an exact block is basis-dependent, with p99 KLE up to 0.75 on conditioned duplicates. ClusterComplete restores the projector of the complete eigenspace.

Near-degeneracy. D-only perturbations leave C fixed: CL KLE is 1.46×10^{-17}, with PE, hidden, and prediction drifts inside tolerance on both evaluated cells. C-local perturbations are evaluated as cutoff sensitivity versus $\|\Delta L\|_F$ (Appendix H).

Selector runtime. Timing on 48 synthetic path unions gives median warm selector times of $1.089\,\mathrm{ms}$ for CL and $0.236\,\mathrm{ms}$ for Global: a $4.6\times$ ratio and an absolute CL time of about $1\,\mathrm{ms}$. Global timing starts from an already computed union spectrum; CL includes component extraction, cache lookup, eigensolves on misses, selection, and assembly. The ratio describes these timed paths under the in-memory cache policy (Appendix M).

5.6 Natural reaction graphs preserve the allocation mechanism

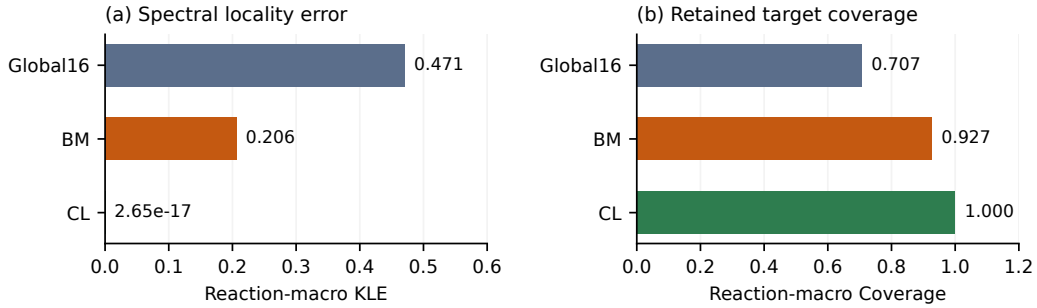

Figure 5: Natural USPTO reaction graphs: all 24,495 nonboundary primary training reactions. Reactive-component metrics are averaged within reactions, then equally across reactions. BM and CL retain the same total number of modes. CL recovers the local kernel and full coverage.

The source graphs already contain separate molecular components. Of the 24,495 nonboundary primary training reactions, 12,044 (49.17%) have positive BM displacement; validation contains 1,517 such reactions. BM has reaction-macro KLE $= 0.206$ and Coverage 0.927. CL reaches numerical-zero KLE and unit coverage (Figure 5). Within the BM-displaced training subset, BM KLE is 0.419 and Coverage is 0.852. BM and CL match their intended total mode counts on all 50,016 reactions.

Table 3: Reactive-atom prediction. Top: reaction-macro average precision on all 5,007 eligible test reactions (mean $\pm$ population SD across five seeds). Bottom: paired CL minus BM effects with reaction-bootstrap 95% intervals. Primary reactions contain at least two non-isolated molecules and both supervised classes; the BM-displaced row excludes cutoff boundaries.

Task scores: all eligible reactions ($n = 5{,}007$)			
NoPE	Global16	BM	CL
0.954726 ± 0.002125	0.957348 ± 0.001679	0.957277 ± 0.001525	0.956115 ± 0.001521
Paired effects: primary reactions			
Population (n)	AP difference	95% CI	Conclusion
All primary (3,512)	+0.000248	$[-0.000669, +0.001145]$	Noninferior
BM-displaced (1,481)	+0.001063	$[-0.000315, +0.002562]$	Gain criterion unmet

Table 3 reports the task scores and the two paired comparisons on their respective populations. Overall noninferiority requires the lower confidence bound to exceed -0.005; CL satisfies this on all 3,512 primary test reactions. Task improvement requires a mean gain of at least 0.005 and a positive lower confidence bound on the nonboundary BM-displaced subset. This improvement threshold is unmet. Thus spectral locality is recovered while overall reactive-atom prediction remains noninferior to BM. Appendix N.3 provides all seed-wise scores and population sensitivities.

Post-hoc exclusion of test reactions sharing a training patent ID leaves 1,040 primary reactions: CL minus BM AP is -0.001294, with 95% CI $[-0.003150, +0.000672]$, retaining noninferiority at the -0.005 margin (Appendix N.6).

Under fresh node permutations of 200 reactions and all 20 checkpoints, every nonboundary case has maximum probability drift below 10^{-5}; the maximum across methods is 5.01×10^{-6}. Appendix N.5 gives boundary-stratified kernel and prediction results, molecule-copy label comparisons, and component-order checks.

6 Related work

Spectral representations. Eigenmaps and spectral clustering (Belkin & Niyogi, 2003; von Luxburg, 2007) motivate Laplacian coordinates in graph transformers (Dwivedi & Bresson, 2021; Kreuzer et al., 2021; Ying et al., 2021; Rampášek et al., 2022; Ma et al., 2023; Shirzad et al., 2023). SignNet and BasisNet address sign and basis ambiguity (Lim et al., 2022); SPE and equivariant encodings study perturbation stability (Huang et al., 2024; Wang et al., 2022; Maskey et al., 2022). EPCF (Sun et al., 2026) removes multiple zero modes from disconnected graphs and uses a whole-graph positive prefix for equivariant collaborative filtering. We analyze how positive modes compete for that finite budget when a component is added.

Operator structure and selection. Davis–Kahan theory studies eigenspace motion under perturbations (Davis & Kahan, 1970); direct sums hold the target block fixed (Chung, 1997). Encoding comparisons study expressivity (Black et al., 2024), and PEARL learns efficient basis-invariant graph-shift representations (Kanatsoulis et al., 2025). We isolate finite-prefix allocation and give a sufficient direct-sum criterion for locality.

Graph computation and reactions. Message passing respects components (Kipf & Welling, 2017; Gilmer et al., 2017; Xu et al., 2019); attention can exchange information across them (Chen et al., 2022; Rampášek et al., 2022). Batching separately precomputed encodings preserves their original scope (Fey & Lenssen, 2019). Molecular GNNs predict reactive centers (Jin et al., 2017) and support retrosynthesis (Dai et al., 2019). Our USPTO study examines spectral allocation within source reaction graphs while retaining reaction-wide interaction.

7 Discussion and future work

Component-wise spectral coordinates can coexist with learned interaction: the controlled study localizes the full downstream path, while USPTO retains reaction-wide attention. This separation supports explicit control over structural coordinates and task-dependent information exchange.

Future work can study patent-family-disjoint evaluation, additional reaction datasets, and chemistry-aware quotas based on molecular size and reactivity. Extending the contextual intervention trace across architectures can clarify interactions among selection, eigenvalue context, and aggregation. Cluster-aware budgets and faster component-wise eigensolvers can jointly support basis invariance, local allocation, and efficient computation.

8 Conclusion

A finite global prefix lets an unrelated component displace target modes. Component-wise quotas restore the direct-sum identity, which a component-local downstream path preserves through encoding and prediction. Independent holdout checks support this identity across seven cells; budget and eigenspace interventions explain its mechanism. Natural USPTO graphs reproduce the allocation effect: CL recovers numerical-zero kernel error and unit coverage, with overall task noninferiority on 3,512 primary test reactions. Component-wise allocation provides a concrete design principle for spectral positional encodings.

AI USE STATEMENT

Generative AI tools assisted with experimental implementation, result checking, figure preparation, literature retrieval, and language editing. Numerical claims are computed from the experiments described here. Responsibility for the final manuscript rests with the authors.

ETHICS STATEMENT

The study uses public graph benchmarks, synthetic graph unions, and public patent-derived reaction records. Model inputs contain molecular atom and bond features. The experiments involve no human-subject intervention.

REPRODUCIBILITY STATEMENT

Appendix B specifies model architectures, training settings, and selector implementations. Appendix C defines the metrics, sampling units, and uncertainty estimates. Appendix N gives reaction populations, supervision rules, and the full USPTO results.

REFERENCES

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

Mitchell Black, Zhengchao Wan, Gal Mishne, Amir Nayyeri, and Yusu Wang. Comparing graph transformers via positional encodings. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24, 2024.

Dexiong Chen, Leslie O'Bray, and Karsten Borgwardt. Structure-aware transformer for graph representation learning. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, 2022.

Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

Hanjun Dai, Chengtao Li, Connor W. Coley, Bo Dai, and Le Song. Retrosynthesis prediction with conditional graph logic network. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

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

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *AAAI Workshop on Deep Learning on Graphs: Methods and Applications*, 2021.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR 2019 Workshop on Representation Learning on Graphs and Manifolds*, 2019. URL https://arxiv.org/abs/1903.02428.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1263–1272, 2017.

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. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Yinan Huang, William Lu, Joshua Robinson, Yu Yang, Muhan Zhang, Stefanie Jegelka, and Pan Li. On the stability of expressive positional encodings for graphs. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=xAqcJ9XoTf.

Wengong Jin, Connor W. Coley, Regina Barzilay, and Tommi Jaakkola. Predicting organic reaction outcomes with Weisfeiler–Lehman network. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Charilaos Kanatsoulis, Evelyn Choi, Stefanie Jegelka, Jure Leskovec, and Alejandro Ribeiro. Learning efficient positional encodings with graph neural networks. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=AWg2tkbydO.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, 2021.

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022. URL https://openreview.net/forum?id=BlM64by6gc.

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 *Proceedings of the 40th International Conference on Machine Learning*, ICML'23, 2023.

Sohir Maskey, Ali Parviz, Maximilian Thiessen, Hannes Stärk, Ylli Sadikaj, and Haggai Maron. Generalized laplacian positional encoding for graph representation learning, 2022. URL https://arxiv.org/abs/2210.15956.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22. Curran Associates Inc., 2022.

Nadine Schneider, Nikolaus Stiefl, and Gregory A. Landrum. What's what: The (nearly) definitive guide to reaction role assignment. *Journal of Chemical Information and Modeling*, 2016. doi: 10.1021/acs.jcim.6b00564.

Hamed Shirzad, Ameya Velingker, Balaji Venkatachalam, Danica J. Sutherland, and Ali Kemal Sinop. Exphormer: sparse transformers for graphs. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23, 2023.

Xin Sun, Jishen Sun, Li Pang, Guiling Wang, Zhizhong Liu, Xin Liu, and Jian Yu. Epcf: An equivariant positional propagation enhanced graph neural network for collaborative filtering. *Information*, 17(7), 2026. doi: 10.3390/info17070644.

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17, 2007. doi: 10.1007/s11222-007-9033-z.

Haorui Wang, Haoteng Yin, Muhan Zhang, and Pan Li. Equivariant and stable positional encoding for more powerful graph neural networks. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=e95i1IHcWj.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

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*, 2021.

A FORMAL PROPOSITIONS

Throughout, $G = (V, E)$ is undirected with nonnegative weights, C is a connected target, and D is node-disjoint from C with no cross-edge, so $G = C \sqcup D$. The Laplacian is $L_{\mathrm{sym}} = I - D^{-1/2} A D^{-1/2}$. Isolated-node blocks follow the PyG fill-value-1 convention unless noted.

Proof of Theorem 1. The direct-sum identities in Propositions 1 and 2 give the merged component spectrum.

(ii), hence (i). Let $\lambda_1^+(C) > 0$ be the first positive target eigenvalue. For the normalized Laplacian of the path P_n, the positive eigenvalues are

$$\lambda_j(P_n) = 1 - \cos\left(\frac{j\pi}{n-1}\right), \qquad j = 1, \ldots, n-1.$$

For fixed k, choose $n > k+1$ large enough that $1 - \cos(k\pi/(n-1)) < \lambda_1^+(C)$. The first k positive path modes then precede every positive target mode in the merged spectrum. Thus $r_C = 0$, whereas the standalone target kernel has rank $k_C \geq 1$. Their kernels differ independently of eigenvector signs or basis choices.

(iii). Diagonalizing each block separately applies the same scalar function to its eigenvalues, so functional calculus preserves the direct sum. Heat kernels e^{-tL} and resolvents $(I + L)^{-1}$ are examples.

(iv). Equation (2) supplies the standalone selected eigensystem on each block, up to its stated equivalences. The element-wise φ assigns a common weight to equal eigenvalues. Thus signs, order, and rotations within a retained exact eigenspace leave each channel of $\mathcal{K}_c$ unchanged. Every component-local downstream operation receives the same target tensors and edges. Composing those operations establishes Equation (1). □

Proposition 1 (Block-diagonal Laplacian)**.** *If C and D share no edge, then in the node order that lists C then D,*

$$L(C \sqcup D) = L(C) \oplus L(D).$$

Proof. No cross-edge implies $A(G) = A(C) \oplus A(D)$. Degrees are computed inside each component, so $D_G^{-1/2} = D_C^{-1/2} \oplus D_D^{-1/2}$. Products of block-diagonal matrices remain block-diagonal, hence so does $D_G^{-1/2} A_G D_G^{-1/2}$. The identity term is diagonal in the same partition. The PyG isolate fill and the zero-diagonal isolate fill each depend only on the node's own degree, so both conventions preserve the direct sum. □

Proposition 2 (Multiset spectrum)**.** $\mathrm{spec} L(G)$ *is the multiset union of the component spectra. The number of structural zeros is $q_{\mathrm{PyG}} = \#\{C_c : |E_c| > 0\}$ under the PyG convention and $q_{\text{0-diag}} = \#\{C_c\}$ under zero-diagonal isolates.*

Proof. A block-diagonal matrix has spectrum equal to the multiset union of its blocks, with eigenvectors extended by zeros (Proposition 1). For a connected component with at least one positive edge, $\ker L_{\mathrm{sym}}$ is one-dimensional, spanned by $D^{1/2}\mathbf{1}$. A PyG isolate block is the 1×1 matrix 1 and contributes eigenvalue 1, not 0. A zero-diagonal isolate block is the 1×1 matrix 0 and contributes one structural zero. □

Proposition 3 (Rank displacement)**.** *Assume the target-local and merged positive-prefix cutoffs do not split an eigenvalue cluster. Let $S_k(G)$ retain the first $\min(k, n_+(G))$ positive modes, and let $r_C = \operatorname{rank} K_{G \to C}$. Then $0 \leq r_C \leq k_C$, and $K_{G \to C} = K_C$ exactly when $r_C = k_C$. If $N_D(< \lambda_{k_C}^C) > k - k_C$, where N_D counts positive distractor modes below the k_C-th positive target eigenvalue, then $r_C < k_C$. With $d_C = k_C - r_C$, $\mathrm{KLE} = \sqrt{d_C / k_C}$ and $\mathrm{Coverage} = 1 - d_C / k_C$.*

Proof. Complete retained eigenspaces admit a component-supported eigenbasis without changing their projector. Restriction to C therefore leaves a prefix of its positive eigenspaces, of rank r_C. The standalone prefix contains that subspace and has rank k_C, so their kernels agree exactly at equal rank. If more than $k - k_C$ distractor modes lie below the last standalone target eigenvalue, a global

budget of k cannot retain all k_C target modes. The excluded target projector has rank d_C; hence $\|K_C - K_{G\to C}\|_F = \sqrt{d_C}$, $\|K_C\|_F = \sqrt{k_C}$, and the overlap fraction is r_C/k_C. □

Proposition 4 (Exact cutoff degeneracy)**.** *Suppose the cutoff eigenvalue $\lambda_{(k)}$ has multiplicity $m > 1$ in the merged spectrum and only $0 < r < m$ vectors from that block are kept. The retained r-plane is not determined by the spectrum: if U_λ is any orthonormal basis of the block and $Q \in O(m)$, then $U_\lambda Q$ is an equally legitimate solver output, and the truncated kernel depends on Q. If instead $r = m$, the projector $U_\lambda U_\lambda^\top$ is Haar-invariant.*

Proof. The columns of $U_\lambda Q$ remain eigenvectors for λ. Keeping the first $r < m$ columns keeps $\mathrm{span}(U_\lambda Q e_1, \ldots, U_\lambda Q e_r)$, which varies with Q. If $r = m$, then

$$(U_\lambda Q)(U_\lambda Q)^\top = U_\lambda Q Q^\top U_\lambda^\top = U_\lambda U_\lambda^\top.$$

ClusterComplete sets $r = m$ and therefore returns a basis-free projector. It does not change the fact that modes of D may still occupy other slots. □

Near-degenerate but distinct eigenvalues have no continuous Haar freedom, which is why arbitrary Q-rotations do not preserve that eigensystem.

Proposition 5 (Component-local selector)**.** *Let Φ compute, on every non-isolated component c, the spectrum of $L(c)$, discard its unique structural zero, and keep $\min(k, |c| - 1)$ positive modes, returning a zero PE on isolates. If φ, ρ, attention, and the target readout are component-local, then*

$$\Phi(C \sqcup D)\big|_C = \Phi(C)$$

for every D that is disjoint from C and shares no cross-edge.

Proof. By Proposition 1, the induced Laplacian of C inside G is exactly $L(C)$. The selector on C is a deterministic function of $L(C)$ and does not read D. Component-local φ, ρ, attention, and readout therefore receive the same input on C in the union as they do on the standalone target. If the cutoff splits an exact cluster of C itself, a deterministic solver still sees the same matrix and the same node order in both evaluations, so the identity continues to hold; ClusterComplete additionally makes that identity basis-free. □

B Model architecture and training

Laplacian and isolates. Primary results use PyG L_{sym} with isolate blocks equal to the scalar 1. Positive-mode selection discards structural zeros and keeps k eigenvalues, or $\min(k, n-1)$ on a component. Component keys are connected-component identifiers computed on the unweighted support.

Selectors. Global-Positive-k ranks the merged positive spectrum. CL-LapPE-k ranks each component independently. BudgetMatched spends $\sum_c \min(k, |c|-1)$ whole-graph modes before restriction. ClusterComplete completes any exact cluster that meets the cutoff. $\mathrm{AllocationRatio}_{\mathrm{nominal}}$ uses denominator $k_C = \min(k, |C| - 1)$. EnergyPerSelectedRank uses the target-side rank actually retained in the standalone reference after cluster completion.

Same checkpoint in the controlled study. SPE models train on connected graphs. Union evaluation uses the same validation-selected checkpoint for Global and CL. Hidden and prediction drifts are aligned before the norm is taken. The locality test uses fresh eigensolves with cache disabled.

Padding, scatter, and keys. Variable-length component PEs are written into a padded buffer with a Boolean mask, then scattered back to union node order by connected-component keys on the unweighted support. Target-only readout gathers nodes of C after that scatter. NoPE and RWSE use the same scatter path.

Pairings. Ordinary distractors are fixed per target. Primary pairs require non-boundary cutoffs on both the target-local and whole-union spectra. Four ordinary distractors are averaged before seed aggregation.

B.1 CONTROLLED MODEL AND TRAINING CONFIGURATION

Table 4 specifies the model architecture and training settings for the controlled study. Training uses connected graphs and validation selects a connected-graph checkpoint. Graph-task readout applies the prediction head to each node and then averages only the target nodes; node-task losses use the target mask. Classification uses cross-entropy, graph multilabel prediction uses masked binary cross-entropy with logits, and regression uses masked L1 loss.

Table 4: Complete controlled-study configuration. All primary normalization is local to individual nodes or edges. The SPE recurrence is in Appendix B.2.

Setting	GINE	GPS
Backbone layers	5	8
Hidden / edge width	128	
Node feature width	96	128
PE integration	Concatenate $96 + 32$	Project $32 \to 128$, then add
Attention	—	4 heads; component blocks
Backbone normalization	Post-residual LayerNorm	Pre-sublayer LayerNorm
Attribute normalization	LayerNorm per node / edge	
Message / update widths	$128 \to 128$; $128 \to 256 \to 128$	
GPS feed-forward width	—	256
Backbone dropout	0.1	
Primary spectral rank / PE width	$k = 16$, $d_{\mathrm{PE}} = 32$	
Element-wise φ	$1 \to 64 \to 64 \to 32$; hidden ReLUs	
Component-local ρ	2 layers, width 64; per-node LayerNorm	
SPE channel mixer / output	$2048 \to 64$; $64 \to 64 \to 32$	
SPE masks	Mode mask, node mask, inactive-component output mask	
Prediction head	LayerNorm; $128 \to 128 \to d_{\mathrm{task}}$, ReLU, dropout 0.1	
Graph readout	Mean of target-node predictions	
Optimizer / weight decay	AdamW / 10^{-5}	
Initial learning rate	10^{-3}	2×10^{-4}
Maximum epochs / early stopping	200 / 30 epochs without validation improvement	
Learning-rate schedule	10-epoch linear warmup; cosine decay to zero	
Gradient norm clip	1.0	
Checkpoint criterion	Validation MAE minimum; accuracy or AP maximum	
Train / validation inputs	Connected graphs; seeded shuffled training batches	
Precision	Float64 eigensolve; float32 training	
Additional / main SPE seeds	$\{0, 1, 2\}$ / $\{0, 1, 2, 3, 4\}$	
Control seeds / master seed	$\{0, 1, 2\}$ / 20260824	
Batch size: ZINC	128	—
Batch size: PATTERN	32	16
Batch size: Peptides-func	32	16
Batch size: MNIST	64	—
Batch size: CLUSTER	24	12
Batch size: PascalVOC-SP	16	8
Batch size: Peptides-struct	32	16

B.2 COMPLETE MASKED SPE RECURRENCE

Let $b_c \in \{0, 1\}^{n_c}$ mark valid nodes in the padded component buffer, and let $\mathcal{K}_{c,a}$ be defined by Equation (3). We evaluate matrix products through their eigenpair factors, without materializing the full three-way kernel tensor. Starting from $H_c^0 = b_c \mathbf{1}_{64}^\top$, each of two layers computes

$$M_{c,a}^\ell = \mathcal{K}_{c,a} H_c^\ell, \qquad a = 1, \dots, 32, \tag{4}$$

$$H_c^{\ell+1} = b_c \odot \mathrm{LN}_{64}\big[\mathrm{MLP}_\ell\big((1 + \epsilon_\ell) H_c^\ell + B_\ell[M_{c,1}^\ell \| \cdots \| M_{c,32}^\ell]\big)\big], \quad \ell = 0, 1, \tag{5}$$

$$\mathrm{PE}_c = \mathbf{1}_{\{\|m_c\|_1 > 0\}}\ \mathrm{MLP}_{\mathrm{out}}(H_c^2)\big|_{\{i : b_{c,i} = 1\}}. \tag{6}$$

Here $B_\ell : \mathbb{R}^{2048} \to \mathbb{R}^{64}$ is affine, ϵ_ℓ is learned from zero, and each update MLP has widths $64 \to 64 \to 64$ with a ReLU between its linear layers. The output MLP has widths $64 \to 64 \to 32$, also with an intermediate ReLU. The element-wise eigenvalue MLP has widths $1 \to 64 \to 64 \to 32$ with ReLU after each hidden layer. LayerNorm acts on the 64 channels of each node. Masks remove padded modes before propagation and padded nodes after every layer; an inactive component is zeroed after the output MLP, including its bias. A stable sort by component key builds the dense buffer, and its inverse restores the original node order. All learned maps are shared across components.

C METRIC DEFINITIONS

Let $\varepsilon = 10^{-15}$ and write $\|\cdot\|_F$ for the unnormalized Frobenius norm. Restrictions to C use the known index set $\pi(C)$ from the union construction.

Spectral. Let U_C contain the orthonormal standalone reference modes, including eigenspace completion for ClusterComplete, and define

$$r_C^{\mathrm{ref}} = \mathrm{rank}(U_C), \qquad k_C^{\mathrm{nom}} = \min(k, |C| - 1).$$

With $K_C = U_C U_C^\top$ and $K_{G\to C} = U_{G\to C} U_{G\to C}^\top$,

$$\mathrm{KLE} = \frac{\|K_C - K_{G\to C}\|_F}{\|K_C\|_F + \varepsilon}, \qquad \mathrm{Coverage} = \frac{\|U_C^\top U_{G\to C}\|_F^2}{r_C^{\mathrm{ref}}}, \qquad \mathrm{TotalEnergy} = \|U_{G\to C}\|_F^2.$$

$\mathrm{AllocationRatio}_{\mathrm{nominal}} = \mathrm{TotalEnergy}/k_C^{\mathrm{nom}}$. For fixed-$k$, nonboundary cases, $r_C^{\mathrm{ref}} = k_C^{\mathrm{nom}} = k_C$, recovering Proposition 3. ClusterComplete uses its completed reference rank for Coverage while keeping the nominal quota in $\mathrm{AllocationRatio}_{\mathrm{nominal}}$. Heat-weighted KLE replaces K by the heat kernel at a fixed time t.

Subspace consistency error (SCE). Let $Q_C = \mathrm{orth}(U_C)$ and $Q_{G\to C} = \mathrm{orth}(U_{G\to C})$, where orth keeps left singular vectors above the threshold $\epsilon_{\mathrm{mach}} \max(n, p)\sigma_{\max}$ for an $n \times p$ input. Then

$$\mathrm{SCE} = \frac{\|Q_C Q_C^\top - Q_{G\to C} Q_{G\to C}^\top\|_F}{\|Q_C Q_C^\top\|_F + \varepsilon}.$$

SCE compares orthonormalized subspace projectors; KLE retains the energy of the original restricted factors. The rank gap is $\mathrm{rank}(U_C) - \mathrm{rank}(U_{G\to C})$.

Budget, quota, and selector contrasts. Let $\bar{e}_{\mathrm{G}}, \bar{e}_{\mathrm{BM}}, \bar{e}_{\mathrm{CL}}$ be the paired-complete mean KLE values on the same 314 ordinary, double-nonboundary, non-isolate cases. The reported reductions are

$$R_{\mathrm{budget}} = 1 - \frac{\bar{e}_{\mathrm{BM}}}{\bar{e}_{\mathrm{G}}}, \qquad R_{\mathrm{quota}} = 1 - \frac{\bar{e}_{\mathrm{CL}}}{\bar{e}_{\mathrm{BM}}}.$$

For the fixed contextual intervention order $A \to B \to C \to D$, let $\bar{d}_C, \bar{d}_D$ be the paired mean PredDrift after the respective steps. Then $R_{\mathrm{sel}} = \bar{d}_C - \bar{d}_D$, in PredDrift units. Its relative reduction is $1 - \bar{d}_D/\bar{d}_C$.

PE and hidden. PEDrift and HiddenDrift are $\|T_C - T_{G\to C}\|_F / (\|T_C\|_F + \varepsilon)$ on the target-restricted SPE or hidden tensor after π. No additional permutation, sign, or rotation is fitted.

Prediction. Let y_C and $y_{G\to C}$ be the task outputs on $\pi(C)$. For graph or node multiclass, PredDrift is the mean Jensen–Shannon divergence of the softmax probabilities; node tasks average over nodes of C. For graph multilabel, it is the mean Bernoulli JS of the sigmoid probabilities. For regression, it is the mean absolute deviation divided by a fixed per-coordinate training scale. Logits enter only through softmax or sigmoid; the reported PredDrift is on probabilities or scaled scalars, not raw logits.

Statistical criteria. A practical prediction-drift reduction requires mean Global PredDrift above 10^{-5}, a CL reduction of at least 80%, a 95% CI of CL − Global with upper bound below zero, at least four of five seeds in the same direction, and pairing completion of at least 95%. Identity tolerances are $\tau_m = \max(\text{preset}, 5 \times \text{replicate-noise } p_{99})$ with preset 1.000×10^{-6} for PE and prediction. AbsLossDrift is practical when the effect is below -10^{-4} and the CI lies below zero.

Targets. Targets in the main comparison are a stratified sample of at most 1000 connected non-boundary test graphs, using fixed label×size strata and inclusion weights. Realized counts are 1000, 1000, 1000, 999, 999, 1000, 1000 for the seven cells in Table 1.

Paired bootstrap. The controlled-study intervals use 2,000 stratified paired target resamples. The four distractors are averaged within each target, and the seed ensemble is averaged within that target before resampling. Targets are sampled inside the fixed label-by-size strata and retain their inclusion weights. Prediction drift uses the percentile CI; signed loss and absolute loss drift use separate seven-cell BH families for their one-sided plus-one tests (Appendix J). The USPTO study resamples reactions within each reported population while keeping the five seeds fixed (Appendix N.3).

D SYNTHETIC FULL RESULTS

Table 5 summarizes ordinary random pairs at $k = 16$, count $= 1$. CL-ClusterComplete has mean reference rank 17.5. Its per-case Coverage is 1, while mean AllocationRatio$_{\mathrm{nominal}} = 17.5/16 = 1.09375$; EnergyPerSelectedRank also remains 1. BudgetMatched uses nominal $k_C^{\mathrm{nom}} = 16$ and AllocationRatio $0.9445 = 15.112/16$.

Table 5: Ordinary random pairs, $k = 16$, count $= 1$ ($n = 384$ per method). SCE is subspace consistency error (Appendix C).

Method	mean KLE	SCE	Coverage	TotalEnergy	Alloc.$_{\mathrm{nom}}$
Global-Positive-k	0.6887	0.6887	0.5199	8.32	0.5199
PyG 1-skip	0.7357	0.7357	0.4852	8.26	0.5164
BudgetMatched	0.3994	0.3994	0.8707	15.11	0.9445
CL-LapPE-k	1.58×10^{-17}	9.82×10^{-16}	1	16	1
CL-ClusterComplete	1.67×10^{-17}	1.05×10^{-15}	1	17.5	1.0938

Table 6: Family prevalence on ordinary random unions ($k = 16$, one distractor). The rate is the number of double-nonboundary displaced draws divided by all 64 draws per family. The last column is the raw Global-Positive-k mean KLE over all 64 draws, including boundary cases.

Family	double-NB	displaced	rate	raw mean KLE
cycle	63	63	0.984	0.720
ER	64	64	1.000	0.595
grid	59	59	0.922	0.735
path	64	64	1.000	0.728
SBM	64	64	1.000	0.693
barbell	0	0	0.000	0.661

At nonboundary cutoffs, component assignments from the merged spectrum match the selected modes at rate 1.0. Heat-weighted Global KLE at $t = 1$ and $t = 10$ is 0.571 and 0.128 on the same $k = 16$ ordinary pairs; CL remains 0. With untrained SPE and component-local aggregation, global selection gives mean PEDrift 0.00220, while component-local selection gives 0.

Table 7: Global-Positive-k versus CL-LapPE-k mean KLE by rank and distractor count.

	count $= 1$, $n = 384$			count $= 4$, $n = 96$		
k	Global	CL	Cov.$_{\mathrm{G}}$	Global	CL	Cov.$_{\mathrm{G}}$
8	0.669	1.20×10^{-17}	0.530	0.842	1.15×10^{-17}	0.273
16	0.689	1.58×10^{-17}	0.520	0.876	1.55×10^{-17}	0.228
32	0.731	2.04×10^{-17}	0.466	0.884	2.05×10^{-17}	0.217

E LAPLACIAN ENCODING IN PYG

Standard GraphGPS and PyG training computes Laplacian positional encodings for each graph before mini-batch collation (Fey & Lenssen, 2019; Rampášek et al., 2022). Each eigensolve therefore acts on one graph, so batching connected graphs preserves the component scope.

The Laplacian positional-encoding transform in PyG 2.8.0.post1 computes L_{sym}, sorts eigenpairs, keeps columns $1, \ldots, k$, and applies a random sign per column. It skips exactly one column, so any additional structural-zero modes remain. On a disconnected input, leftover kernel modes of other components can occupy the retained prefix.

This rule encounters a merged spectrum on naturally disconnected graphs, on a mini-batch eigen-decomposed as one matrix, and on constructed disjoint unions. Table 5 reports PyG 1-skip on constructed unions: mean $\mathrm{KLE} = 0.7357$ versus 0.6887 for Global-Positive-k. The locality failure is present under the public 1-skip rule.

F Additional benchmark results

Table 8: Prediction drift on additional benchmarks (three SPE seeds). Each comparison uses the same checkpoint for Global and CL.

Cell	Global	CL	Reduction	Drift reduced
PATTERN/GINE	0.00378	5.57×10^{-9}	$> 99.99\%$	yes
PATTERN/GPS	0.00404	4.89×10^{-9}	$> 99.99\%$	yes
Peptides-func/GINE	0.0448	3.72×10^{-9}	$> 99.99\%$	yes
Peptides-func/GPS	0.0517	1.88×10^{-9}	$> 99.99\%$	yes
ZINC/GINE	0.263	8.79×10^{-8}	$> 99.99\%$	yes

NoPE and RWSE use the same union scatter as SPE. Mean PredDrift is 0 on every additional and main NoPE/RWSE cell.

Table 9: Connected-graph validation of SPE checkpoints versus the task-specific trivial baseline. Classification cells report accuracy; Peptides-func reports AP; ZINC and Peptides-struct report MAE (lower is better). Every seed improves on the task-specific trivial baseline.

Cell	n seeds	val. mean$\pm$sd	trivial	metric
CLUSTER/GINE	5	0.171 ± 0.000	0.167	acc.
CLUSTER/GPS	5	0.770 ± 0.001	0.167	acc.
MNIST/GINE	5	0.975 ± 0.001	0.112	acc.
PascalVOC-SP/GINE	5	0.722 ± 0.002	0.699	acc.
PascalVOC-SP/GPS	5	0.709 ± 0.002	0.699	acc.
Peptides-struct/GINE	5	0.268 ± 0.001	0.642	MAE
Peptides-struct/GPS	5	0.260 ± 0.001	0.642	MAE
PATTERN/GINE	3	0.905 ± 0.001	0.824	acc.
PATTERN/GPS	3	0.909 ± 0.001	0.824	acc.
Peptides-func/GINE	3	0.598 ± 0.007	0.165	AP
Peptides-func/GPS	3	0.649 ± 0.013	0.165	AP
ZINC/GINE	3	0.233 ± 0.003	1.472	MAE

G Target pairings and stress tests

All sampled targets are evaluated in the seven main cells: 1000, 1000, 1000, 999, 999, 1000, and 1000 targets, respectively, giving 100% completion. Ordinary, four-distractor, and duplicate unions satisfy the PE, hidden, and prediction locality tolerances in $7/7$ cells. Unions with isolates or tiny components satisfy PE and prediction tolerances in $7/7$ cells and hidden tolerances in $6/7$. For Peptides-struct/GINE, the maximum hidden drift is 1.09×10^{-5}, compared with $10\tau = 1.00 \times 10^{-5}$; its p99 remains within τ. All conditions use the same tolerances as the independent holdout.

Table 10: Per-seed CL$-$Global PredDrift (five SPE seeds). Cells with practical drift reduction have $5/5$ seeds in the same direction. CLUSTER/GINE is the numerical-floor cell.

Cell	s0	s1	s2	s3	s4	dir.
CLUSTER/GINE	0	0	-6.67×10^{-10}	-7.58×10^{-10}	-4.19×10^{-10}	floor
CLUSTER/GPS	-0.00187	-0.00238	-0.00195	-0.00110	-0.00418	$5/5$
MNIST/GINE	-0.00123	-0.00109	-0.00078	-0.00061	-0.00203	$5/5$
PascalVOC-SP/GINE	-0.0272	-0.0272	-0.0291	-0.0267	-0.0387	$5/5$
PascalVOC-SP/GPS	-0.00302	-0.00502	-0.00636	-0.00563	-0.00418	$5/5$
Peptides-struct/GINE	-0.918	-0.816	-0.985	-0.932	-0.859	$5/5$
Peptides-struct/GPS	-0.680	-0.735	-0.703	-0.698	-0.735	$5/5$

Table 11: Exact repeated eigenspaces. Haar error is full-block rotation error; projector error is ClusterComplete local-versus-union.

Kind	n	max projector	max Haar	fixed-k p99 KLE
analytic cycle	16	0	8.31×10^{-16}	5.8×10^{-15}
duplicate cycle	16	1.84×10^{-13}	8.31×10^{-16}	0.707
duplicate conditioned	37	5.18×10^{-13}	8.31×10^{-16}	0.75

H EXACT AND NEAR DEGENERACY

D-only CL KLE is 1.46×10^{-17} across $\varepsilon \in [10^{-8}, 10^{-2}]$. C-local mean KLE is about 0.129 on that range and is read as cutoff sensitivity, not as a locality identity.

I LOCALITY TOLERANCES AND INDEPENDENT HOLDOUT

Locality is tested on an independent holdout at $\tau_m = \max(\text{preset}, 5 \times \text{replicate-noise } p_{99})$. The tolerances were calibrated using 10 standalone and 10 union cases after an initial test-union inspection. They were then fixed before testing 40 previously unused standalone validation graphs and 30 validation-target/ validation-distractor union pairs per cell. The seed-0 SPE checkpoint was used in each of the seven cells, with fresh eigensolves and the spectrum cache disabled. Each standalone graph underwent 10 repeated eigensolves and 10 node-permuted fresh eigensolves. No test target entered this holdout.

Each check requires $p_{99} \le \tau_m$ and maximum drift $\le 10\tau_m$. All 56 cell/metric/condition comparisons satisfy these tolerances, including PE, hidden, and prediction drift on holdout unions in Table 12. The thresholds were unchanged. These results support the identity claim in Section 5.2 and the Id. column of Table 1; that column uses seed 0, while its PredDrift estimates use five seeds.

Table 12: Independent holdout identity test: seed 0, 30 union pairs per cell. Columns show maximum aligned drift and fixed thresholds. PE threshold is 10^{-6} for every cell. Standalone evaluations satisfy the same tolerances.

Cell	max PE	max hidden	max Pred	τ_{hidden}	τ_{Pred}	Id.
CLUSTER/GINE	0	0	0	1.00×10^{-6}	1.00×10^{-6}	yes
CLUSTER/GPS	0	0	0	1.87×10^{-6}	1.00×10^{-6}	yes
MNIST/GINE	0	0	2.21×10^{-8}	1.00×10^{-6}	1.00×10^{-6}	yes
PascalVOC-SP/GINE	0	0	0	1.00×10^{-6}	1.00×10^{-6}	yes
PascalVOC-SP/GPS	0	0	0	1.99×10^{-6}	1.00×10^{-6}	yes
Peptides-struct/GINE	0	0	9.69×10^{-8}	1.00×10^{-6}	1.38×10^{-6}	yes
Peptides-struct/GPS	0	0	1.13×10^{-7}	1.75×10^{-6}	1.13×10^{-6}	yes

Additional stress tests use ordinary, four-distractor, and duplicate unions at the same fixed thresholds. All satisfy PE, hidden, and prediction locality in $7/7$ cells. Unions with isolates or tiny components satisfy hidden locality in $6/7$, as detailed in Appendix G.

J TASK-LOSS STABILITY

Signed $\Delta\ell$ and AbsLossDrift are two independent 7-cell BH families. One-sided plus-one p has floor $1/2001$. After BH, signed loss has $q < 0.05$ on $4/7$ cells (practical on the same four). AbsLossDrift has $q = 4.998 \times 10^{-4}$ on all seven cells and is practical on six. Prediction-drift comparisons use paired confidence intervals.

Starred entries meet the practical threshold.

K SUPPORTING ABLATIONS

k sensitivity. On CLUSTER/GPS and Peptides-struct/GINE, $k \in \{8, 32\}$, 3 seeds, CL PredDrift is 0 in all 12 runs. Global PredDrift remains positive.

Table 13: Signed task-loss and absolute loss-drift contrasts, with independent Benjamini–Hochberg correction for each metric.

Cell	signed $\Delta\ell$	q	AbsLossDrift	q
CLUSTER/GINE	-8.3×10^{-11}	0.408	-5.8×10^{-9}	5.0×10^{-4}
CLUSTER/GPS	-0.00186	0.148	$-0.0328^{\star}$	5.0×10^{-4}
MNIST/GINE	-0.00396	0.126	$-0.0284^{\star}$	5.0×10^{-4}
PascalVOC-SP/GINE	$-0.0708^{\star}$	8.7×10^{-4}	$-0.137^{\star}$	5.0×10^{-4}
PascalVOC-SP/GPS	$-0.0147^{\star}$	8.7×10^{-4}	$-0.0698^{\star}$	5.0×10^{-4}
Peptides-struct/GINE	$-0.671^{\star}$	8.7×10^{-4}	$-0.689^{\star}$	5.0×10^{-4}
Peptides-struct/GPS	$-0.507^{\star}$	8.7×10^{-4}	$-0.521^{\star}$	5.0×10^{-4}

Union augmentation. Peptides-struct/GINE trained with 50% unions: Global PredDrift 0.205 versus 0.905 under connected-only training; CL is 0 in both.

Contextual interventions. CLUSTER/GPS: $R_{\text{sel}} = 2.01 \times 10^{-6}$, CI $[1.28, 3.16] \times 10^{-6}$, reduction 99.8%, $5/5$ seeds. Peptides-struct/GINE: $R_{\text{sel}} = 1.224$, CI $[1.116, 1.346]$, reduction 100%, with the paired interval entirely above zero.

Table 14: Contextual SPE mean PredDrift by nested scope (5 seeds, completion 1.0).

Cell	A	B	C	D
CLUSTER/GPS	0.00304	0.00304	2.02×10^{-6}	4.73×10^{-9}
Peptides-struct/GINE	0.473	0.473	1.224	1.89×10^{-7}

Architecture scope. Evaluation-only ablations hold the checkpoint and change one mixing channel. Component-block attention keeps CL PredDrift at 0 (CLUSTER/GPS Global 0.00176; Peptides-struct/GINE Global 0.379). Graph-wide normalization and graph-wide attention add a second mixing path even under a local selector, which is why the main comparison freezes a component-local downstream stack.

Bridge edges. At bridge weight 0 the pair remains a disjoint union and all drifts are 0. Any positive weight leaves the Component Locality identity setting and is read as a continuous perturbation.

L Graph-level screening and reaction-graph evidence

Screening ZINC, Peptides-func, Peptides-struct, and MNIST found a small share of disconnected graphs with components large enough for the spectral comparison (quota-eligible share $\leq 0.39\%$). Their task labels describe a whole molecule or image. The controlled study therefore places real graphs in disjoint unions while holding the target C fixed. The USPTO study in Appendix N adds naturally disconnected source graphs with atom-level supervision.

Table 15: Training-split screening for intrinsically disconnected inputs. The quota share counts graphs with components large enough for the spectral comparison. Task labels describe the entire graph.

Dataset	graphs	disc.+edge	quota share	Label scope
ZINC	10000	0	0	Graph
Peptides-func	10873	88	0.0028	Graph
Peptides-struct	10873	88	0.0028	Graph
MNIST	55000	0	0	Graph

M Runtime, memory, and reproducibility

Peak GPU memory in the memory benchmark is $3.83\,\text{GB}$. The runtime comparison covers 48 synthetic path unions with 86–439 nodes. Median cold whole-union eigensolve time is $2.219\,\text{ms}$. Median warm selector times are $0.236\,\text{ms}$ for Global and $1.089\,\text{ms}$ for CL, a ratio of 4.60. The Global timer receives an already computed union spectrum. The CL timer includes component extraction,

in-memory component-spectrum cache lookups (with eigensolves on misses), selection, and assembly. Cache hits and misses are not timed separately. The reported ratio describes these timed paths under their cache policy. Amortizing the cold eigensolve over two selectors gives $1.109\,\mathrm{ms}$ per selector. Controlled training settings are in Table 4; strict identity evaluation uses component-wise execution, float64 spectral/message reductions, float32 learned layers, and disabled dropout.

N USPTO NATURAL REACTION-GRAPH SUPPLEMENT

N.1 POPULATIONS AND SOURCE-SIDE SUPERVISION

The study uses the canonicalized USPTO-50K split. All 50,016 reactions parse successfully. An eligible reaction has at least one positive and one negative supervised source atom. A primary reaction additionally contains at least two non-isolated source molecules. Table 16 gives the resulting populations. Table 17 reports displacement counts and eigensolver completion. Displaced training reactions have BM KLE 0.418979 and Coverage 0.851830; eigensolver failures are zero in every split.

Reaction SMILES follow the `reactants>reagents>products` format. Only the left source field supplies nodes and bonds. Every molecule present there is included. A reagent or catalyst represented in that field is therefore part of the input. The separate middle reagent field is empty in all 50,016 reactions. Products are used offline for supervision.

A mapped source atom is supervised when its atom-map ID occurs in the product. Its label is positive when the mapped bond environment, formal charge, hydrogen count, or chirality changes. Mapped atoms absent from the product are excluded from the loss and task metrics. Unmapped atoms remain supervised negatives. Product coordinates and atom-map numbers are absent from the OGB atom and bond feature tensors. Reaction class is also absent from model features. Atom-map values are used only to construct labels and verify atom correspondences.

Table 16: Reaction populations. Primary reactions contain at least two non-isolated source molecules and both supervised classes.

Split	Raw/parsed	Eligible	Primary	Isolate	Primary isolate
train	40,008	40,004	28,002	366	4
val	5,001	5,001	3,506	47	0
test	5,007	5,007	3,512	53	0

Table 17: Reaction displacement and eigensolver completion. Displacement is evaluated on primary reactions with no BM or local cutoff boundary; the failure rate covers all training reactions.

Quantity	Population	Value
Displaced reactions	Training, nonboundary primary	12,044
Displaced reactions	Validation, nonboundary primary	1,517
Displaced share	Training, nonboundary primary	49.17%
BM KLE	Training, displaced subset	0.418979
Eigensolver failure rate	All training reactions	0%

N.2 SPECTRAL ALLOCATION AND ANALYSIS POPULATIONS

Each non-isolated component has quota $k_c = \min(16, |c| - 1)$; isolates have $k_c = 0$. BM selects $K_{\mathrm{total}} = \sum_c k_c$ whole-reaction modes. CL selects each component's quota separately. Global16 retains a positive prefix of length at most 16. No eigensolve spans different reactions in a mini-batch.

Let $\mathcal{C}_R^+$ denote components with a positive supervised atom, and let r_c^{BM} count the selected BM modes assigned to component c. The primary displacement statistic is

$$q_R^{\mathrm{BM}} = \frac{\sum_{c \in \mathcal{C}_R^+} \max(k_c - r_c^{\mathrm{BM}}, 0)}{\sum_{c \in \mathcal{C}_R^+} k_c}.$$

The fixed-budget statistic q_R^{fixed} uses the corresponding Global16 quotas. Nonboundary primary reactions exclude either a BM cutoff boundary or a component-local cutoff boundary. Displacement

prevalence uses this fixed population. Table 19 also reports all primary reactions; Table 20 supplies the boundary frequencies for both full and primary populations.

Both BM and CL match their intended total selected-mode count in every reaction (Table 18). The maximum K_{total} is 32. BM represents whole-graph eigenvectors with 32 columns. CL reuses columns over disjoint component supports, with separate eigenvalue and mode-mask rows for each component. Each local spectrum uses at most 16 columns. Its reported total sums selected modes across components.

Table 18: Actual selected-mode count versus the sum of local budgets. The CL count sums modes across components.

Split	Selector	n	Mismatches	Max. modes	Storage width
train	BM	40,008	0	32	32
train	CL	40,008	0	32	32
val	BM	5,001	0	32	32
val	CL	5,001	0	32	32
test	BM	5,007	0	32	32
test	CL	5,007	0	32	32
all	BM	50,016	0	32	32
all	CL	50,016	0	32	32

Table 19: Fixed-budget and BM displacement prevalence on common populations. NB excludes BM or local cutoff boundaries.

Split	Population	n	Fixed count	Fixed rate	BM count	BM rate
train	All primary	28,002	24,178	86.34%	15,243	54.44%
train	NB primary	24,495	20,674	84.40%	12,044	49.17%
val	All primary	3,506	3,074	87.68%	1,900	54.19%
val	NB primary	3,077	2,645	85.96%	1,517	49.30%
test	All primary	3,512	2,992	85.19%	1,876	53.42%
test	NB primary	3,083	2,563	83.13%	1,481	48.04%

Table 20: Cutoff-boundary counts and frequencies. The last column is the union of BM and local flags.

Split	Population	Fixed	BM	Local	BM/local union
train	All	7767 (19.41%)	3609 (9.02%)	3562 (8.90%)	5164 (12.91%)
train	Primary	6110 (21.82%)	1952 (6.97%)	1919 (6.85%)	3507 (12.52%)
val	All	969 (19.38%)	449 (8.98%)	441 (8.82%)	634 (12.68%)
val	Primary	764 (21.79%)	244 (6.96%)	238 (6.79%)	429 (12.24%)
test	All	971 (19.39%)	412 (8.23%)	418 (8.35%)	610 (12.18%)
test	Primary	791 (22.52%)	232 (6.61%)	240 (6.83%)	429 (12.22%)

N.3 Training and task estimands

Each of NoPE, Global16, BM, and CL is trained with five random seeds, 0–4. The common GraphGPS model has six layers, hidden width 128, eight attention heads, feed-forward width 256, and dropout 0.1. SPE uses 32 output channels. Attention separates reactions by the batch index and allows all source molecules within a reaction to interact. Component-local spectral aggregation is preserved.

Training uses AdamW with learning rate 10^{-3}, weight decay 10^{-5}, gradient clipping at 1, and batch size 128. The limit is 100 epochs with early-stopping patience 20. Validation reaction-macro AUPRC selects each checkpoint. Positive-class weighting is computed from supervised training atoms. The masked binary cross-entropy first averages supervised atoms within each reaction and then averages reactions in the batch. Tables 21 and 24 report all trained methods and individual seeds.

For reaction r, seed s, and method m, let $a^m_{r,s}$ be average precision over that reaction's supervised atoms. The paired statistic on population $\mathcal{P}$ is

$$\widehat{\Delta}_{\mathcal{P}} = \frac{1}{|\mathcal{P}|} \sum_{r \in \mathcal{P}} \left[\frac{1}{5} \sum_{s=0}^{4} \left(a^{\mathrm{CL}}_{r,s} - a^{\mathrm{BM}}_{r,s} \right) \right].$$

Thus the average is taken over per-seed AP differences. The bootstrap samples reactions with replacement, preserves the same five seed values for every sampled reaction, and uses 2,000 resamples with random seed 20260824. The 95% interval is the percentile interval of those paired reaction means. Table 23 gives the effect for each seed.

We define task improvement as $\widehat{\Delta} \geq 0.005$ and a positive interval lower bound on nonboundary BM-displaced primary reactions. That subset contains 1,481 test reactions, with $\widehat{\Delta} = +0.001063$ and 95% CI $[-0.000315, +0.002562]$; this task-improvement threshold is unmet. Overall noninferiority uses all 3,512 primary reactions, including boundaries, and requires the interval lower bound to exceed -0.005. The observed interval $[-0.000669, +0.001145]$ satisfies that criterion. The 3,083 nonboundary primary reactions provide the mechanism-consistency analysis. Table 22 reports each population separately, including boundary and zero-displacement subsets.

The interpretation is natural spectral-locality recovery with overall task noninferiority. The USPTO comparison estimates complete training-algorithm differences. The controlled union experiments in the main study supply the same-checkpoint selector intervention.

Table 21: Task results on all 5,007 eligible test reactions. Mean and population SD summarize the five fixed seeds; micro-AP and macro-F1 are secondary metrics.

Method	Reaction-macro AP	Seed SD	Micro-AP	Macro-F1
NoPE	0.954726	0.002125	0.970411	0.910247
Global16	0.957348	0.001679	0.972161	0.912599
BM	0.957277	0.001525	0.972221	0.912207
CL	0.956115	0.001521	0.971419	0.911690

Table 22: Paired CL minus BM effects in reaction-macro AP. Confidence intervals resample reactions while keeping the five-seed ensemble fixed.

Population	n	Δ	95% CI
All primary (noninferiority)	3,512	+0.000248	$[-0.000669, +0.001145]$
Nonboundary primary	3,083	+0.000159	$[-0.000900, +0.001175]$
Nonboundary BM-displaced	1,481	+0.001063	$[-0.000315, +0.002562]$
BM-displaced incl. boundary	1,876	+0.001101	$[-0.000096, +0.002333]$
Nonboundary $q_R^{\mathrm{BM}} = 0$	1,602	-0.000676	$[-0.002218, +0.000687]$

Table 23: Seed-wise CL minus BM reaction-macro AP effects.

Population	s0	s1	s2	s3	s4
All primary (noninferiority)	+0.001617	+0.000691	-0.001334	-0.000764	+0.001031
Nonboundary primary	+0.001138	+0.000894	-0.001448	-0.000597	+0.000808
Nonboundary BM-displaced	+0.000990	-0.000194	+0.000167	+0.001099	+0.003251
BM-displaced incl. boundary	+0.001939	-0.000124	-0.000031	+0.000330	+0.003390
Nonboundary $q_R^{\mathrm{BM}} = 0$	+0.001275	+0.001900	-0.002941	-0.002165	-0.001450

N.4 Reaction-macro mechanism and selector sensitivities

For each reaction, KLE, Coverage, and TotalEnergy are first averaged over its reactive components with positive local budget. Reactions then receive equal weight. Table 25 reports both all 40,008 training reactions and the 24,495 nonboundary primary training reactions. The latter population is the one used in Figure 5.

Global-ClusterComplete extends the positive-16 cutoff to include its entire eigenvalue cluster. Table 26 reports its selected-mode counts and spectral metrics alongside Global16, BM, and CL. It is an untrained spectral sensitivity selector.

AllocationRatio uses $\mathrm{TotalEnergy}_c/k_c$. EnergyPerSelectedRank uses $\mathrm{TotalEnergy}_c/r_c$, where r_c is the selected-mode allocation to component c. Components with $r_c = 0$ have an undefined second ratio and are omitted from that within-reaction mean; their counts appear in Table 27. Reaction means are then averaged equally. At cutoff boundaries, these allocations describe the selected basis.

Table 24: Validation and test performance for all four methods and five seeds.

Method	Seed	Val. macro-AP	Test macro-AP	Micro-AP	Macro-F1
NoPE	0	0.955432	0.955351	0.969732	0.910636
NoPE	1	0.956038	0.952151	0.969945	0.905694
NoPE	2	0.955418	0.952505	0.970015	0.908996
NoPE	3	0.955321	0.955834	0.970859	0.910993
NoPE	4	0.954131	0.957791	0.971505	0.914919
Global16	0	0.956003	0.956064	0.973041	0.909877
Global16	1	0.955327	0.956001	0.971045	0.911953
Global16	2	0.957928	0.958263	0.971821	0.913405
Global16	3	0.955717	0.956164	0.972041	0.908434
Global16	4	0.958422	0.960246	0.972859	0.919324
BM	0	0.956699	0.958430	0.973253	0.916377
BM	1	0.957121	0.959167	0.972645	0.911449
BM	2	0.957482	0.954881	0.970942	0.909976
BM	3	0.958037	0.956334	0.971547	0.911290
BM	4	0.957640	0.957572	0.972719	0.911942
CL	0	0.957716	0.956537	0.973828	0.914462
CL	1	0.958224	0.958146	0.972395	0.913385
CL	2	0.955920	0.953463	0.968654	0.907158
CL	3	0.956476	0.956543	0.972278	0.911877
CL	4	0.957451	0.955888	0.969940	0.911566

Table 25: Reactive-component metrics with equal reaction weights: average components within each reaction, then average reactions.

Population	Selector	n	KLE	Coverage	TotalEnergy
All train	BM	40,008	1.72×10^{-1}	0.938065	12.425494
All train	CL	40,008	1.82×10^{-17}	1.000000	12.433367
All train	Global16	40,008	3.46×10^{-1}	0.783213	10.094115
All train	IsolateMatched	40,008	1.69×10^{-1}	0.938655	12.433367
All train	PyG 1-skip	40,008	4.25×10^{-1}	0.756197	10.132107
NB primary train	BM	24,495	2.06×10^{-1}	0.927146	11.010921
NB primary train	CL	24,495	2.65×10^{-17}	1.000000	11.010941
NB primary train	Global16	24,495	4.71×10^{-1}	0.707171	7.793386
NB primary train	IsolateMatched	24,495	2.06×10^{-1}	0.927146	11.010941
NB primary train	PyG 1-skip	24,495	5.88×10^{-1}	0.668229	7.855399

Table 26: Global-ClusterComplete extends the positive-16 cutoff to the complete eigenvalue cluster. All rows use reaction-macro spectral metrics.

Population	Selector	n	Mean modes	KLE	Coverage
All train	Global16	40,008	15.584	3.46×10^{-1}	0.783213
All train	BM	40,008	20.252	1.72×10^{-1}	0.938065
All train	CL	40,008	20.252	1.82×10^{-17}	1.000000
All train	ClusterComplete	40,008	16.023	3.43×10^{-1}	0.797758
NB primary train	Global16	24,495	15.601	4.71×10^{-1}	0.707171
NB primary train	BM	24,495	22.040	2.06×10^{-1}	0.927146
NB primary train	CL	24,495	22.040	2.65×10^{-17}	1.000000
NB primary train	ClusterComplete	24,495	16.035	4.51×10^{-1}	0.725146
NB primary test	Global16	3,083	15.588	4.65×10^{-1}	0.709889
NB primary test	BM	3,083	22.068	2.01×10^{-1}	0.930774
NB primary test	CL	3,083	22.068	2.71×10^{-17}	1.000000
NB primary test	ClusterComplete	3,083	16.040	4.45×10^{-1}	0.728123

Table 27: Reaction-macro allocation ratios. Zero-owner-quota components are excluded from EnergyPerSelectedRank and counted explicitly.

Population	Selector	AllocationRatio	Energy/rank	Zero-quota components
All train	BM	0.962565	1.000000	1,068
All train	CL	1.000000	1.000000	0
All train	Global16	0.783846	1.000000	1,325
All train	IsolateMatched	0.963129	1.000000	1,068
All train	PyG 1-skip	0.806343	1.000000	566
NB primary	BM	0.955185	1.000000	846
NB primary	CL	1.000000	1.000000	0
NB primary	Global16	0.707171	1.000000	959
NB primary	IsolateMatched	0.955186	1.000000	846
NB primary	PyG 1-skip	0.739755	1.000000	386

N.5 Label consistency and permutation equivariance

Label consistency and molecule copies. We independently reconstruct labels, supervision masks, and atom-map correspondences from the reaction SMILES for 800 reactions per split. All agree with the labels and masks used for training and evaluation. The molecule-copy comparison identifies exact attributed graph isomorphisms among molecules within a reaction. Matching uses the model's map-free OGB atom and bond attributes. Pinned-node automorphism tests determine atom orbits, and a copy isomorphism aligns them. We compare supervision masks across each aligned orbit and reactive labels among its supervised atoms.

No within-reaction molecule copies occur in the 50,016 reaction graphs (Table 28). The copy-related ambiguity count is therefore zero, including in all 3,512 primary test reactions. Excluding ambiguous reactions leaves the task-gain and overall-noninferiority results exactly unchanged.

Spectral permutation. The kernel permutation test uses 200 primary training reactions with one node permutation per reaction, shared across selectors. A case is classified as boundary when the selector has a cutoff boundary in either the original or permuted graph. For CL, the 180 nonboundary cases have mean KLE 3.10×10^{-15}; the 20 boundary cases have mean KLE 0.247370. Table 29 also reports IsolateMatched. Completing the cutoff eigenspace retains all modes at its eigenvalue; fixed-rank selection can split that eigenspace.

Final predictions. For each method, models trained with all five seeds evaluate the same 200 permutations in evaluation mode. Eigenpairs are recomputed on each permuted source graph, and predictions are inverse-aligned to the original node order. We compare logits and sigmoid probabilities on all nodes and separately on supervised atoms. Table 30 reports mean and maximum absolute differences, stratified by each selector's boundary flag. NoPE has no spectral boundary. Table 31 reports all 20 trained models.

The diagnostic float32 probability tolerance is 10^{-5}. All nonboundary reaction–model pairs satisfy it. Fixed-rank boundary cases are reported separately. A control applies the node permutation directly to the unpermuted graph's PE, separating model equivariance from spectral reselection; its maximum probability drift across all methods and cases is 7.39×10^{-6}.

Component ordering. Before dense SPE packing, nodes are stably sorted by connected component, and an inverse sort restores their outputs to the input order. This handles permutations that interleave components. On the 200 unpermuted graphs, whose nodes are already grouped by component, this ordering preserves logits exactly for all 20 models.

N.6 Leakage sensitivity and isolate policy

The fixed test split contains 50 exact source/product SMILES pairs also present in training (0.9986%) and 3,466 reactions sharing a patent ID with training (69.2231%). We assess their effect by excluding overlapping reactions from the test population, using the same trained models and paired reaction bootstrap.

Table 28: Label consistency for isomorphic molecule copies. Supervision masks and labels are compared over aligned atom automorphism orbits using map-free atom and bond attributes.

Split	Population	n	With copies	Ambiguous
train	All	40,008	0	0
train	Primary	28,002	0	0
val	All	5,001	0	0
val	Primary	3,506	0	0
test	All	5,007	0	0
test	Primary	3,512	0	0

Table 29: Kernel equivariance under node permutations on 200 primary training reactions, stratified by each selector's cutoff boundary.

Selector	Stratum	n	Mean KLE	Max. KLE
CL	nonboundary	180	3.10×10^{-15}	1.30×10^{-14}
CL	boundary	20	2.47×10^{-1}	4.12×10^{-1}
IsolateMatched	nonboundary	190	3.39×10^{-15}	2.20×10^{-14}
IsolateMatched	boundary	10	1.72×10^{-1}	3.34×10^{-1}

Table 30: Prediction equivariance under node permutations: all-node probability and logit differences after inverse alignment. Means weight node comparisons equally across models trained with five seeds.

Method	Stratum	Reactions	Cases	Mean prob.	Max. prob.	Max. logit
NoPE	NB	200	1000	3.78×10^{-9}	1.64×10^{-6}	9.54×10^{-6}
Global16	NB	151	755	8.08×10^{-9}	3.73×10^{-6}	3.86×10^{-5}
Global16	Boundary	49	245	4.52×10^{-4}	3.76×10^{-1}	4.50×10^{0}
BM	NB	190	950	6.29×10^{-9}	2.92×10^{-6}	2.48×10^{-5}
BM	Boundary	10	50	2.51×10^{-4}	6.99×10^{-2}	2.40×10^{0}
CL	NB	180	900	7.25×10^{-9}	5.01×10^{-6}	3.05×10^{-5}
CL	Boundary	20	100	5.45×10^{-4}	1.53×10^{-1}	1.93×10^{0}

Table 31: Probability drift for each method and seed. NB denotes reactions without a cutoff boundary for that selector.

Method	Seed	NB reactions	NB mean	NB max.	Boundary max.
NoPE	0	200	3.46×10^{-9}	1.19×10^{-6}	–
NoPE	1	200	3.03×10^{-9}	8.94×10^{-7}	–
NoPE	2	200	4.86×10^{-9}	1.13×10^{-6}	–
NoPE	3	200	3.81×10^{-9}	1.64×10^{-6}	–
NoPE	4	200	3.73×10^{-9}	1.13×10^{-6}	–
Global16	0	151	7.88×10^{-9}	1.56×10^{-6}	3.76×10^{-1}
Global16	1	151	7.42×10^{-9}	3.46×10^{-6}	5.24×10^{-2}
Global16	2	151	9.65×10^{-9}	3.73×10^{-6}	2.85×10^{-1}
Global16	3	151	7.41×10^{-9}	2.98×10^{-6}	1.31×10^{-1}
Global16	4	151	8.04×10^{-9}	3.16×10^{-6}	7.42×10^{-3}
BM	0	190	6.83×10^{-9}	2.92×10^{-6}	4.69×10^{-2}
BM	1	190	8.37×10^{-9}	2.24×10^{-6}	5.95×10^{-2}
BM	2	190	3.84×10^{-9}	1.10×10^{-6}	1.13×10^{-2}
BM	3	190	6.83×10^{-9}	1.31×10^{-6}	1.53×10^{-2}
BM	4	190	5.61×10^{-9}	2.32×10^{-6}	6.99×10^{-2}
CL	0	180	8.87×10^{-9}	2.50×10^{-6}	1.16×10^{-1}
CL	1	180	8.17×10^{-9}	2.15×10^{-6}	9.11×10^{-2}
CL	2	180	7.35×10^{-9}	1.13×10^{-6}	4.71×10^{-2}
CL	3	180	5.89×10^{-9}	5.01×10^{-6}	1.53×10^{-1}
CL	4	180	5.95×10^{-9}	1.49×10^{-6}	5.61×10^{-2}

After exact-repeat exclusion, 3,480 primary reactions remain, with $\Delta = +0.000257$ and CI $[-0.000700, +0.001205]$. After shared-patent exclusion, 1,040 primary reactions remain, with $\Delta = -0.001294$ and CI $[-0.003150, +0.000672]$. Both all-primary subsets meet overall noninferiority. Table 32 also reports the corresponding nonboundary primary and BM-displaced subsets. The task-improvement threshold remains unmet in both exclusions.

IsolateMatched is an untrained sensitivity selector: it applies BudgetMatched to the graph after removing isolated nodes, preventing isolates from taking spectral modes. Training, validation, and test contain 366, 47, and 53 isolate reactions. Primary isolate counts are 4, 0, and 0. Training isolate prevalence is $366/40{,}008 = 0.9148\%$. Table 33 reports BM minus IsolateMatched spectral differences. All mode counts match; primary validation and test spectral statistics coincide. The application comparison retains this complete training-algorithm interpretation.

Table 32: Post-hoc sensitivity to excluding training-set overlaps, using the same trained models. Noninferiority is evaluated on all primary reactions in each group.

Test inclusion	Population	n	$\Delta_{\mathrm{CL-BM}}$	95% CI
Original split	BM-displaced	1,481	+0.001063	$[-0.000315, +0.002562]$
Original split	NB primary	3,083	+0.000159	$[-0.000900, +0.001175]$
Original split	All primary	3,512	+0.000248	$[-0.000669, +0.001145]$
Exclude exact repeats	BM-displaced	1,478	+0.001065	$[-0.000313, +0.002481]$
Exclude exact repeats	NB primary	3,053	+0.000168	$[-0.000821, +0.001229]$
Exclude exact repeats	All primary	3,480	+0.000257	$[-0.000700, +0.001205]$
Exclude shared patents	BM-displaced	363	+0.000619	$[-0.002452, +0.003397]$
Exclude shared patents	NB primary	939	-0.001229	$[-0.003331, +0.000673]$
Exclude shared patents	All primary	1,040	-0.001294	$[-0.003150, +0.000672]$

Table 33: BM minus IsolateMatched spectral differences. Both selectors retain the same total number of modes in every reaction.

Split	Population	n	ΔKLE	ΔCoverage	ΔEnergy
train	All	40,008	2.22×10^{-3}	-5.90×10^{-4}	-7.87×10^{-3}
train	Primary	28,002	1.09×10^{-5}	-4.43×10^{-6}	-7.14×10^{-5}
val	All	5,001	2.20×10^{-3}	-5.73×10^{-4}	-7.60×10^{-3}
val	Primary	3,506	0	0	0
test	All	5,007	2.55×10^{-3}	-6.56×10^{-4}	-9.39×10^{-3}
test	Primary	3,512	0	0	0