# SoLIFT: Source–Local Feature Interfaces for Graph Foundation Models

**Anonymous authors**
Paper under double-blind review

## Abstract

Graph foundation models (GFMs) aim to transfer knowledge across graphs, requiring shared encoders to accommodate heterogeneous node attributes. A prominent class of GFM methods maps node attributes to a fixed input width through graph-local spectral transformations. These select feature directions and coordinates independently within each graph. Consequently, recurring source directions may be discarded if they are not locally dominant. We introduce SoLIFT (**So**urce–**L**ocal **I**nterfaces **F**or **T**ransfer), a framework that uses source evidence to guide feature selection and coordinate choice. Under an explicit positional feature convention, it combines frozen source subspace evidence with local variation through two variants. SoLIFT-Split reserves common capacity and fills the remaining width with residual PCA, while SoLIFT-Fusion jointly selects directions and orients the selected basis toward a fixed source reference. Our theoretical analysis characterizes how source evidence affects feature selection and shows that coordinate orientation preserves the selected representations' within-graph geometry. We evaluate our framework with seven representative GFM pipelines for transfer from source graphs to unseen target graphs. Our framework improves overall transfer performance over the tested graph-local preprocessing baselines.

## 1 Introduction

Graph foundation models (GFMs) aim to learn reusable representations from diverse graphs and support learning across tasks and domains (Liu et al., 2025; Mao et al., 2024). A prominent class of these methods learns from source graphs and transfers the resulting model to unseen target graphs for downstream tasks through a model-specific adaptation procedure (Xia & Huang, 2026; Wang et al., 2025). This workflow requires a shared encoder to process node attributes that can differ in both dimension and meaning across graphs. Methods with a fixed input width therefore preprocess each graph's attributes into compatible node features, both for source training and for target use. Graph-local spectral transformations provide one such interface. AnyGraph uses graph-wise feature SVD (Xia & Huang, 2026), while MDGFM uses PCA (Wang et al., 2025). Although these transformations make input dimensions compatible, they also determine which attribute information reaches the model. Graph-local PCA and feature SVD select directions according to variation within each graph, without directly accounting for whether those directions recur across sources. A direction supported by several source graphs need not dominate any one of them and can be discarded under a limited feature budget. Shared training then operates on representations shaped by these independent selection decisions. For a fixed interface, optimizing the downstream encoder does not revise the retained directions, although graph structure may supply additional information. This motivates using source evidence for feature selection as well as model training (Figure 1).

Source evidence must nevertheless complement graph-local variation. Retaining only recurring directions can exclude variation specific to a new graph. The interface must decide how to allocate its limited width between directions supported across sources and variation specific to the current graph. To combine evidence from graphs with different node counts and no node correspondence, we compare feature subspaces. This comparison uses an explicit coordinate convention: native feature columns are matched by position and shorter vectors are extended with zeros. The resulting numerical overlap does not establish semantic correspondence and can change under independent column permutations. Within this convention, source recurrence supplies evidence beyond local variance. Subspace selection leaves a second decision: which coordinates should represent the retained directions to the shared encoder? Different orthonormal bases preserve the selected representations'



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.

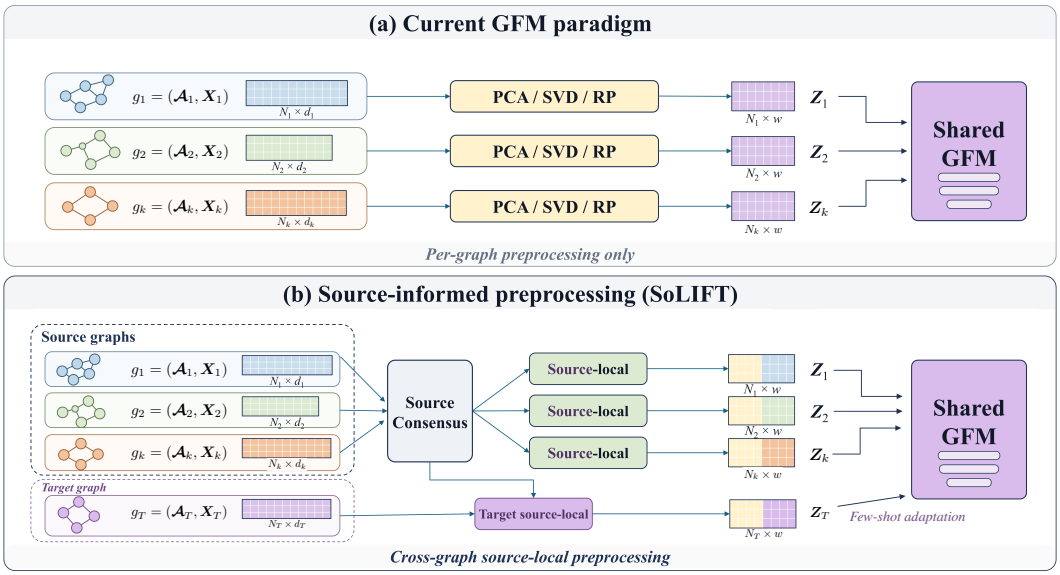



Figure 1: **Let source graphs inform the feature bottleneck.** (a) A class of GFMs preprocesses each graph independently. (b) SoLIFT combines recurring source geometry with graph-specific evidence.

within-graph inner products and distances, but change the inputs seen by the same model parameters. For a shared linear input map, different graph-specific rotations generally cannot be absorbed by one parameter change (Proposition 9). Preserving each graph's geometry therefore need not make coordinate choices equivalent for the shared model. Source information can guide direction selection and provide a reference for choosing their coordinates. We then ask: *can source evidence improve feature selection and coordinate choice at fixed width under the same GFM architecture and training/adaptation policy?*

We introduce **SoLIFT** (**So**urce–**L**ocal **I**nterfaces **F**or **T**ransfer) to bring source evidence into the feature interface while retaining graph-local variation. Building on shared/private subspace learning (Shi & Al Kontar, 2024) and projector aggregation (Fan et al., 2019), it aggregates source PCA projectors with equal graph weights and reuses this source state without target updates. Within this framework, **SoLIFT-SPLIT** (SPLIT) explicitly reserves part of the feature budget for source-supported directions and uses residual PCA to fill the remaining width with variation outside that common block. **SoLIFT-FUSION** (FUSION) instead uses spectral fusion to select all retained directions jointly by local and source subspace agreement. It then orients the selected basis toward a full-source reference, preserving the selected representations' within-graph inner products and distances. Our analysis relates changes in selection to how source evidence couples directions inside and outside the local subspace under stated rank and gap conditions (Theorem 5). Predictive effects of orientation depend on downstream parameterization and are evaluated empirically. Our contributions:

- **A source-informed GFM interface.** We introduce SoLIFT to incorporate source evidence into feature selection and coordinate choice at a fixed-width GFM interface. With a shared state fitted only on sources and an explicit positional convention, SPLIT and FUSION provide distinct ways to combine source and local subspaces.
- **Theoretical analysis.** We give sufficient conditions for estimating a common source subspace. For FUSION's exact spectral selection, we bound departure from local PCA under stated rank, coupling, and spectral-gap conditions. The analysis distinguishes subspace changes from coordinate orientation, which preserves the selected representation's geometry.
- **Empirical evaluation.** We evaluate transfer from source graphs to unseen target graphs using seven representative GFM pipelines. SoLIFT achieves better overall transfer performance than the tested graph-local preprocessing baselines.

## 2 RELATED WORK

**Feature interfaces for GFMs.** GFM input interfaces accommodate heterogeneous attributes through learned mappings or graph-local spectral transformations. Learned interfaces include

GraphFM's dataset-specific MLPs (Lachi et al., 2025) and FUG's dimension encoder (Zhao et al., 2024b). OFA embeds node and edge descriptions with a language model (Liu et al., 2024), while GraphAlign combines language-model features, graphwise centering, and learned feature experts (Hou et al., 2024). Both exploit textual information beyond SoLIFT's positional coordinate convention. Spectral reduction can also precede learned alignment. MDGFM combines PCA with learned alignment (Wang et al., 2025), and GCOPE uses SVD by default before learning cross-graph coordinators (Zhao et al., 2024a). AnyGraph combines feature and structure SVD and periodically refreshes its spectral representations during training (Xia & Huang, 2026). Domain-specific tokens and target prompts can adjust message aggregation (Yu et al., 2025). LEDA aligns latent distributions to guide learned domain projections, coupling input mapping to multi-domain training (Shan et al., 2026). Other representations support transfer through statistical dependencies under invariance assumptions (STAGE, Shen et al., 2025), or covariance operators after random projection (ALL-IN, Eliasof et al., 2026). For ALL-IN, feature-permutation invariance holds in distribution and orthogonal invariance concerns the expected operator. SoLIFT places source reuse in the preprocessing stage, fitting its spectral source state separately from GFM training and reusing it for unseen targets. This stage uses features alone, with graph structure and support labels entering the subsequent GFM stages. Matching attribute columns by position defines a coordinate convention, but neither establishes semantic correspondence nor guarantees invariance to independent attribute permutations across graphs.

**Shared and private subspace learning.** Shared/private methods separate common and individual variation. JIVE decomposes blocks measured on the same objects (Lock et al., 2013), while Personalized PCA jointly optimizes global and local components under orthogonality constraints to minimize reconstruction error (Shi & Al Kontar, 2024). SPLIT instead fits a common frame before conditionally optimizing residual variation. Distributed PCA averages local projectors for communication-efficient eigenspace estimation (Fan et al., 2019), while optimal subspace averaging formalizes agreement through projector distances (Santamaría et al., 2016). Closest to SPLIT, multi-study PCA aggregates weighted target and informative-source projectors, then uses residual PCA in the estimated common space's orthogonal complement for any remaining target-private directions (Li et al., 2024). When source relevance is unknown, it also provides a target-dependent procedure for selecting informative sources through overlap with an iteratively estimated shared subspace (Li et al., 2024). SPLIT instead fits an equally weighted consensus on sources alone and reuses it across unseen targets within a prescribed GFM input width. This reuse avoids refitting shared evidence for each target, but does not selectively exclude sources according to their relevance to that target. SPLIT adapts established aggregation and residual constructions rather than introducing a new general-purpose shared/private decomposition. FUSION also uses a weighted projector barycenter, assigning equal coefficients to local and average external-source projector agreement to select the full width without a prescribed common/local split. Once the subspace is selected, its basis determines the coordinates supplied to a shared encoder. Classical orthogonal Procrustes (Schönemann, 1966) orients FUSION's selected basis toward a fixed reference constructed from all source graphs, preserving the selected subspace and its node representations' within-graph inner products and distances.

# 3 METHODOLOGY

## 3.1 OVERVIEW

We study fixed-width GFM interfaces for source graphs $g_1, \ldots, g_k$ and an unseen target $g_T$. For $i \in \{1, \ldots, k, T\}$, let $g_i = (\boldsymbol{\mathcal{A}}_i, \boldsymbol{X}_i)$ have adjacency $\boldsymbol{\mathcal{A}}_i \in \mathbb{R}^{N_i \times N_i}$ and features $\boldsymbol{X}_i \in \mathbb{R}^{N_i \times d_i}$. Node counts, feature widths, and label spaces may differ. At positive integer width $w$, we fit a reusable source state $\mathcal{S}$ and transform each graph:

$$\mathcal{S} = \mathrm{Fit}(\{\boldsymbol{X}_j\}_{j=1}^{k}), \qquad \boldsymbol{Z}_i = \mathcal{T}_i(\boldsymbol{X}_i; \mathcal{S}) \in \mathbb{R}^{N_i \times w}. \tag{1}$$

Each interface uses fresh source training with the same downstream architecture and training/adaptation policy. Preprocessing uses features alone. Target transforms keep $\mathcal{S}$ fixed but use unlabeled features, including query features, for local evidence. This transductive access precedes GFM adaptation with target structure and a small labeled support set. Query labels serve only for scoring.

SoLIFT aggregates local feature subspaces into source evidence (Figure 2). SPLIT reserves common directions and fills the remaining width with local variation outside their span. FUSION jointly selects all directions by agreement with local and source subspaces, then orients the selected basis toward a

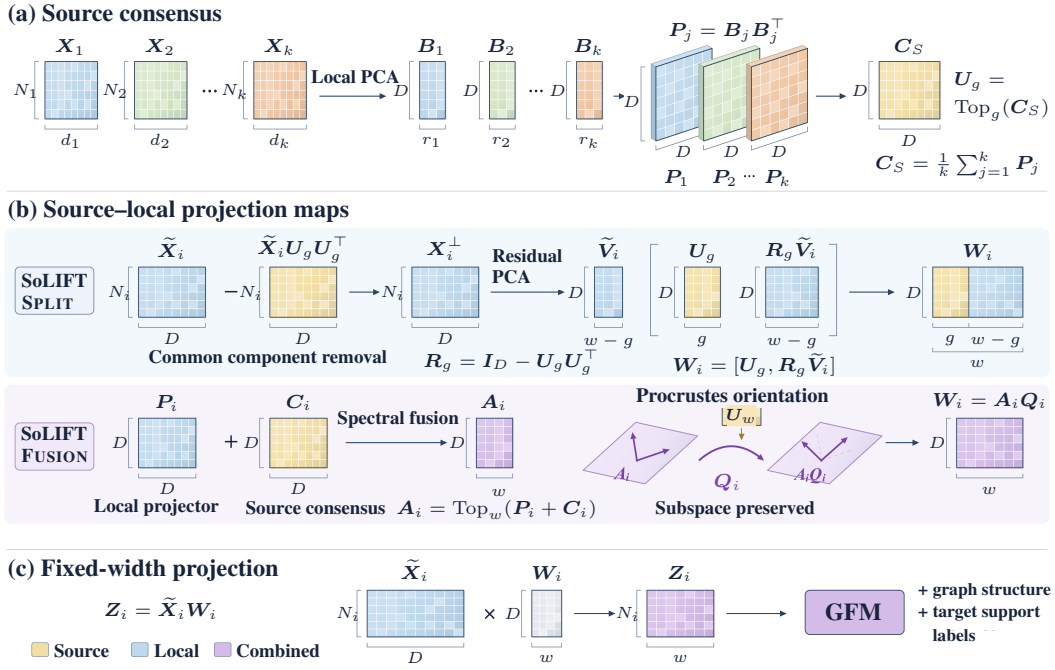



Figure 2: **SoLIFT combines source consensus with local subspaces.** (a) Source consensus estimation. (b) SPLIT allocates common and residual capacity, while FUSION jointly selects a subspace and orients its basis. (c) Fixed-width features support GFM training and adaptation.

fixed source reference. Selection thus determines which directions reach the shared encoder, while orientation determines the coordinates of their node scores.

### 3.2 SOURCE SUBSPACE CONSENSUS

Graphs can have different node counts and no node correspondence, but we can compare directions in their feature spaces. Centering isolates variation around each graph's mean, and local PCA summarizes that variation in a small subspace. We match native columns by position in $\mathbb{R}^D$, where $D = \max\{w, d_1, \ldots, d_k, d_T\}$: Fit uses source widths alone, and a wider target only zero-extends stored state. This numerical convention does not establish semantic correspondence, and independently permuting native attribute columns can change subspace comparisons. Let $\boldsymbol{E}_i \in \mathbb{R}^{D\times d_i}$ append zero coordinates, $\mathbf{1} \in \mathbb{R}^{N_i}$ be the all-ones vector, and $\boldsymbol{H}_i = \boldsymbol{I}_{N_i} - N_i^{-1}\mathbf{1}\mathbf{1}^\top$. The centered native and extended features are

$$\boldsymbol{X}_i^c = \boldsymbol{H}_i \boldsymbol{X}_i \in \mathbb{R}^{N_i\times d_i}, \qquad \widetilde{\boldsymbol{X}}_i = \boldsymbol{X}_i^c \boldsymbol{E}_i^\top \in \mathbb{R}^{N_i\times D}. \tag{2}$$

Right singular vectors describe feature directions, while multiplying the centered features by these vectors gives node scores. The thin SVD has $m_i = \min(N_i, d_i)$, $\boldsymbol{L}_i^x \in \mathbb{R}^{N_i\times m_i}$ and $\boldsymbol{V}_i^x \in \mathbb{R}^{d_i\times m_i}$. Order $\sigma_{i1} \geq \cdots \geq \sigma_{im_i} \geq 0$ and retain $r_i = \min(w, N_i, d_i)$ right singular vectors with the largest singular values:

$$\boldsymbol{X}_i^c = \boldsymbol{L}_i^x \,\mathrm{diag}(\sigma_{i1}, \ldots, \sigma_{im_i})(\boldsymbol{V}_i^x)^\top, \qquad \boldsymbol{B}_i^{\mathrm{ret}} = [v_{i1}^x, \ldots, v_{ir_i}^x] \in \mathbb{R}^{d_i\times r_i}. \tag{3}$$

A subspace has no intrinsic choice of basis. A direction and its negative represent the same variation but would cancel if averaged directly. We therefore compare the retained subspaces through projectors:

$$\boldsymbol{B}_i = \boldsymbol{E}_i \boldsymbol{B}_i^{\mathrm{ret}} \in \mathbb{R}^{D\times r_i}, \qquad \boldsymbol{P}_i = \boldsymbol{B}_i \boldsymbol{B}_i^\top \in \mathbb{R}^{D\times D}. \tag{4}$$

Rotating or reflecting a basis leaves its projector unchanged: $\boldsymbol{B}_i \boldsymbol{O}_i$ gives the same $\boldsymbol{P}_i$ whenever $\boldsymbol{O}_i \in \mathbb{R}^{r_i\times r_i}$ satisfies $\boldsymbol{O}_i^\top \boldsymbol{O}_i = \boldsymbol{I}_{r_i}$. Source aggregation therefore compares subspaces independently of their chosen bases.

For a unit direction $u \in \mathbb{R}^D$, source support $\|\boldsymbol{B}_j^\top u\|_2^2$ is zero for directions orthogonal to the source subspace and one for directions within it. We seek a subspace maximizing this support averaged across sources. Represent a candidate by $\boldsymbol{V} \in \mathbb{R}^{D\times r}$, $r \in \{0, \ldots, D\}$, with orthonormal columns, also called an orthonormal frame. The corresponding consensus operator is $\boldsymbol{C}_S = \frac{1}{k}\sum_{j=1}^k \boldsymbol{P}_j \in \mathbb{R}^{D\times D}$. Then $u^\top \boldsymbol{C}_S u = k^{-1}\sum_{j=1}^k \|\boldsymbol{B}_j^\top u\|_2^2$, and $\mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{C}_S \boldsymbol{V})$ sums support over the frame. This

aggregates retained directions rather than their variance. Equal coefficients avoid weighting by node counts or singular-value magnitudes, but do not equalize traces when returned ranks differ. Decompose $\boldsymbol{C}_S$, order $\lambda_1 \geq \cdots \geq \lambda_D$, and select:

$$\boldsymbol{C}_S = \boldsymbol{U}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_D)\boldsymbol{U}^\top, \qquad \boldsymbol{U}_r = [u_1, \ldots, u_r] \in \mathbb{R}^{D\times r}. \tag{5}$$

Each eigenvalue measures its direction's average source support. A direction represented in several local subspaces can therefore score highly without dominating any one graph. We abbreviate this ordered eigendecomposition and leading-column selection by $\mathrm{Top}_r(\boldsymbol{M})$ for a symmetric $D \times D$ matrix $\boldsymbol{M}$. Classically, this support criterion also minimizes projector distance (Santamaría et al., 2016), as formalized in the proposition below.

**Proposition 1** (Consensus characterization, proof in Appendix A.3)**.** *For orthogonal source projectors $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_k$, the frame $\boldsymbol{U}_r$ in Equation* (5) *maximizes the average source support below:*

$$\boldsymbol{U}_r \in \operatorname*{arg\,max}_{\boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}_r} \mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{C}_S \boldsymbol{V}). \tag{6}$$

*Equivalently, its projector minimizes the mean squared Frobenius distance to the source projectors among rank-$r$ orthogonal projectors. If $0 < r < D$ and $\lambda_r > \lambda_{r+1}$, its projector is unique.*

Consensus expresses collective support rather than an exact intersection of source subspaces. Interpreting it as an estimate of a population common subspace requires additional conditions. The next theorem assumes that common strength exceeds average private geometry and that errors preserve this separation. For an orthonormal basis matrix $\boldsymbol{V}$, write $\boldsymbol{\Pi}_{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{V}^\top$ for its subspace projector.

**Theorem 2** (Conditional consensus estimation, proof in Appendix A.4)**.** *Let $\boldsymbol{U}_\star \in \mathbb{R}^{D\times q}$ have orthonormal columns. Suppose ideal population projectors $\boldsymbol{\Pi}_i \in \mathbb{R}^{D\times D}$, $i = 1, \ldots, k$, admit*

$$\boldsymbol{\Pi}_i = \boldsymbol{U}_\star \boldsymbol{\Omega}_i \boldsymbol{U}_\star^\top + \boldsymbol{L}_i + \boldsymbol{\Xi}_i, \qquad \boldsymbol{\Omega}_i \succeq a_i \boldsymbol{I}_q,$$

*where $\boldsymbol{\Omega}_i \in \mathbb{R}^{q\times q}$ and $\boldsymbol{L}_i, \boldsymbol{\Xi}_i \in \mathbb{R}^{D\times D}$ satisfy $\boldsymbol{L}_i\boldsymbol{U}_\star = 0$, $0 \preceq \boldsymbol{L}_i \preceq \boldsymbol{I}_D$, and $\|\boldsymbol{\Xi}_i\|_{\mathrm{op}} \leq h_i$. Let $\bar{a} = k^{-1}\sum_i a_i$, $\bar{\boldsymbol{L}} = k^{-1}\sum_i \boldsymbol{L}_i$, and assume $\delta = \bar{a} - \|\bar{\boldsymbol{L}}\|_{\mathrm{op}} > 0$. Define the total error*

$$\epsilon = \frac{1}{k}\sum_i h_i + \left\| \frac{1}{k}\sum_i (\boldsymbol{P}_i - \boldsymbol{\Pi}_i) \right\|_{\mathrm{op}}. \tag{7}$$

*If $\epsilon < \delta/2$, then*

$$\|\boldsymbol{\Pi}_{\mathrm{Top}_q(\boldsymbol{C}_S)} - \boldsymbol{U}_\star\boldsymbol{U}_\star^\top\|_{\mathrm{op}} \leq \frac{2\epsilon}{\delta}.$$

Here $\boldsymbol{\Omega}_i$ describes common strength, $\boldsymbol{L}_i$ private geometry orthogonal to it, and $\boldsymbol{\Xi}_i$ population misspecification. Private geometry may persist after averaging, provided its strength remains below the common component. The error $\epsilon$ includes misspecification and population-to-returned projector discrepancies from empirical, numerical, or rank-policy effects. The bound is small when this total error is small relative to $\delta$. These sufficient conditions concern geometric recovery, not predictive relevance.

The empirical consensus supplies bases at the widths needed by each interface. SPLIT reserves $g \in \{0, \ldots, w\}$ common coordinates using $\boldsymbol{U}_g$. FUSION uses the full relevant consensus for selection and $\boldsymbol{U}_w$ for orientation. Neither design width need equal the population rank $q$. We now explain how each design combines source and local evidence.

### 3.3 SOLIFT-SPLIT: EXPLICIT CAPACITY ALLOCATION

SPLIT makes source evidence explicit by reserving $g$ coordinates for $\boldsymbol{U}_g$. These directions are retained even when they do not dominate the current graph's variation. The remaining $w - g$ coordinates should capture variation outside this prescribed common subspace. To avoid selecting directions already in the common block, we first remove that component with $\boldsymbol{R}_g = \boldsymbol{I}_D - \boldsymbol{U}_g\boldsymbol{U}_g^\top$:

$$\boldsymbol{X}_i^\perp = \widetilde{\boldsymbol{X}}_i\boldsymbol{R}_g = \widetilde{\boldsymbol{X}}_i - (\widetilde{\boldsymbol{X}}_i\boldsymbol{U}_g)\boldsymbol{U}_g^\top \in \mathbb{R}^{N_i\times D}. \tag{8}$$

Within this complement, retained variation is the mean squared energy of centered node scores. For an orthonormal $\boldsymbol{V} \in \mathbb{R}^{D\times(w-g)}$ in $\boldsymbol{U}_g^\perp$, this equals $N_i^{-1}\|\widetilde{\boldsymbol{X}}_i\boldsymbol{V}\|_F^2 = \mathrm{tr}(\boldsymbol{V}^\top\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V})$, where $\widehat{\boldsymbol{\Sigma}}_i = N_i^{-1}\widetilde{\boldsymbol{X}}_i^\top\widetilde{\boldsymbol{X}}_i$. We obtain the residual directions by a thin SVD with $m_i^\perp = \min(N_i, D)$, ordered singular values $\tau_{i1} \geq \cdots \geq \tau_{im_i^\perp} \geq 0$. When at least $w - g$ columns are available, we retain the first $w - g$ right singular vectors:

$$\boldsymbol{X}_i^\perp = \boldsymbol{L}_i^\perp\,\mathrm{diag}(\tau_{i1}, \ldots, \tau_{im_i^\perp})(\boldsymbol{V}_i^\perp)^\top, \qquad \widetilde{\boldsymbol{V}}_i = [v_{i1}^\perp, \ldots, v_{i,w-g}^\perp] \in \mathbb{R}^{D\times(w-g)}. \tag{9}$$

The combined map $\boldsymbol{W}_i^{\text{split}} = [\boldsymbol{U}_g, \boldsymbol{R}_g\widetilde{\boldsymbol{V}}_i]$ yields the node features

$$\boldsymbol{Z}_i^{\text{split}} = [\widetilde{\boldsymbol{X}}_i\boldsymbol{U}_g,\ \boldsymbol{X}_i^{\perp}\widetilde{\boldsymbol{V}}_i]. \tag{10}$$

The common block is shared across graphs, while the residual block adapts locally. The next proposition characterizes this allocation's best residual variance when the common block is fixed.

**Proposition 3** (Conditional residual optimality, proof in Appendix A.5)**.** *With $\boldsymbol{U}_g$ fixed, exact PCA within $\boldsymbol{U}_g^{\perp}$ produces an orthonormal residual frame $\boldsymbol{V} \in \mathbb{R}^{D\times(w-g)}$ solving*

$$\max_{\boldsymbol{V}}\ \operatorname{tr}(\boldsymbol{V}^\top\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V}) \quad \text{subject to} \quad \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_{w-g}, \quad \boldsymbol{U}_g^\top\boldsymbol{V} = 0. \tag{11}$$

This optimum is conditional on the common block and its capacity. Connecting the ideal orthonormal solution to the actual map requires keeping $\boldsymbol{R}_g$. Positive residual directions satisfy $\boldsymbol{R}_g v = v$, but zero singular values can yield completion directions with common components. Applying $\boldsymbol{R}_g$ removes those components without changing the residual scores, although the projected completion columns need not be orthonormal. With enough returned columns, the map attains the same optimal residual score energy under exact SVD. Randomized residual PCA approximates this reference, and orthogonal blocks need not have independent scores. The allocation $g$ remains a design choice. At $g = 0$, SPLIT is local PCA under its solver convention. At $g = w$, only the common block remains.

### 3.4 SOLIFT-FUSION: JOINT SELECTION AND COORDINATE ORIENTATION

FUSION jointly selects all $w$ directions by agreement with source and local subspaces. It changes both the capacity rule and the criterion, replacing residual variance with subspace overlap. No prescribed common direction is guaranteed retention. Here source consensus provides external evidence, so we exclude the current source graph to avoid counting its projector twice. The unseen target uses all sources. Use $\mathcal{J}_i = \{1,\dots,k\}\setminus\{i\}$ for sources ($k \geq 2$) and $\mathcal{J}_T = \{1,\dots,k\}$ for the unseen target. The relevant consensus is

$$\boldsymbol{C}_i = \frac{1}{|\mathcal{J}_i|}\sum_{j\in\mathcal{J}_i}\boldsymbol{P}_j. \tag{12}$$

Thus $\boldsymbol{C}_i = \boldsymbol{C}_{-i}$ for sources and $\boldsymbol{C}_T = \boldsymbol{C}_S$, with $0 \preceq \boldsymbol{C}_i \preceq \boldsymbol{I}_D$. For a unit direction $u$, $u^\top\boldsymbol{P}_i u$ measures overlap with the graph's own subspace, and $u^\top\boldsymbol{C}_i u$ measures average support from the relevant sources. Both lie in $[0,1]$. We choose equal coefficients on this comparable directional scale, although this choice does not equalize total projector mass. Summing these two scores over the selected directions gives the joint objective:

$$J_i(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{P}_i\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{C}_i\boldsymbol{A}), \qquad \boldsymbol{A}_i \in \operatorname*{arg\,max}_{\boldsymbol{A}^\top\boldsymbol{A}=\boldsymbol{I}_w} J_i(\boldsymbol{A}). \tag{13}$$

Since $J_i(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^\top(\boldsymbol{P}_i+\boldsymbol{C}_i)\boldsymbol{A})$, form $\boldsymbol{M}_i = \boldsymbol{P}_i + \boldsymbol{C}_i$, order its eigenvalues $\mu_{i1} \geq \cdots \geq \mu_{iD}$, and retain the first $w$ eigenvectors to solve the joint maximization:

$$\boldsymbol{M}_i = \boldsymbol{T}_i\operatorname{diag}(\mu_{i1},\dots,\mu_{iD})\boldsymbol{T}_i^\top, \qquad \boldsymbol{A}_i = [t_{i1},\dots,t_{iw}] \in \mathbb{R}^{D\times w}. \tag{14}$$

When the local basis returns $w$ columns, it is a feasible baseline with maximal local overlap. The next proposition bounds how much of that overlap joint selection can trade for greater source agreement.

**Proposition 4** (Source–local compensation, proof in Appendix A.6)**.** *Let $\boldsymbol{P}$ be a rank-$w$ local projector, $0 \preceq \boldsymbol{C} \preceq \boldsymbol{I}$, and $\boldsymbol{A}$ an exact leading orthonormal width-$w$ frame of $\boldsymbol{P}+\boldsymbol{C}$. Then*

$$0 \leq w - \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{P}\boldsymbol{A}) \leq \operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{C}\boldsymbol{A}) - \operatorname{tr}(\boldsymbol{P}\boldsymbol{C}). \tag{15}$$

Denote the middle term in Equation (15) by $\ell$, the loss of local overlap, and the rightmost term by $a$, the gain in source agreement relative to the local baseline. The source gain compensates for the local loss in subspace affinity. Since directions are not weighted by variance, this does not bound reconstruction error or predictive loss. Its returned-rank condition does not require $w$ positive-variance directions.

We next ask when source evidence changes the local subspace. For FUSION's equal-coefficient operator, we compare the coupling between this subspace and its orthogonal complement with their spectral separation. Let $\boldsymbol{P} \in \mathbb{R}^{D\times D}$ be a rank-$w$ orthogonal projector with $0 < w < D$, let $\boldsymbol{C} \in \mathbb{R}^{D\times D}$ satisfy $0 \preceq \boldsymbol{C} \preceq \boldsymbol{I}_D$, and let $\boldsymbol{A} \in \mathbb{R}^{D\times w}$ contain exact leading orthonormal eigenvectors of $\boldsymbol{P}+\boldsymbol{C}$. The coupling $b$ measures how strongly source evidence mixes local directions with directions outside the local subspace. The block gap $\delta_0$ compares the weakest fused score inside

with the strongest outside when this mixing is removed:

$$b = \|(\boldsymbol{I}_D - \boldsymbol{P})\boldsymbol{C}\boldsymbol{P}\|_{\mathrm{op}}, \qquad \delta_0 = 1 + \min_{\substack{\|u\|_2=1 \\ \boldsymbol{P}u=u}} u^\top \boldsymbol{C}u - \max_{\substack{\|v\|_2=1 \\ \boldsymbol{P}v=0}} v^\top \boldsymbol{C}v. \tag{16}$$

This block gap generally differs from the eigengap of $\boldsymbol{P} + \boldsymbol{C}$.

**Theorem 5** (Stability under source coupling, proof in Appendix A.7)**.** *Under the above setup, if $\delta_0 > 0$ and $b < \delta_0/2$, the selected projector and its local overlap loss satisfy*

$$\|\boldsymbol{\Pi}_{\boldsymbol{A}} - \boldsymbol{P}\|_{\mathrm{op}} \le \frac{2b}{\delta_0} \text{ and } \ell \le w\left(\frac{2b}{\delta_0}\right)^2. \tag{17}$$

*If $b = 0$ and $\boldsymbol{P} + \boldsymbol{C}$ has no eigenvalue tie at the selection boundary, then $\boldsymbol{\Pi}_{\boldsymbol{A}} = \boldsymbol{P}$.*

Weak coupling relative to the block gap keeps exact selection close to local PCA. With zero coupling and no boundary tie, even nonzero source support leaves the local span unchanged. FUSION therefore need not insert new directions, whereas SPLIT reserves a prescribed common block. These conclusions concern the rank-$w$ returned projector, including any null completion, and quantify geometry rather than retained feature energy or predictive benefit.

The selected subspace still admits different orthonormal bases, which can give different node scores to the shared encoder. We choose coordinates by orienting $\boldsymbol{A}_i$ toward the full-source reference $\boldsymbol{U}_w$, which remains fixed under source exclusion. Assuming both bases return $w$ columns, orthogonal Procrustes (Schönemann, 1966) finds a closest basis to this reference within the selected span:

$$\boldsymbol{Q}_i \in \operatorname*{arg\,min}_{\boldsymbol{Q}\in\mathbb{R}^{w\times w},\, \boldsymbol{Q}^\top\boldsymbol{Q}=\boldsymbol{I}_w} \|\boldsymbol{A}_i\boldsymbol{Q} - \boldsymbol{U}_w\|_F^2. \tag{18}$$

The cross-basis SVD, with nonnegative diagonal $\boldsymbol{\Gamma}_i$, gives the rotation and the resulting node features:

$$\boldsymbol{A}_i^\top \boldsymbol{U}_w = \boldsymbol{L}_i^Q \boldsymbol{\Gamma}_i (\boldsymbol{R}_i^Q)^\top, \qquad \boldsymbol{Q}_i = \boldsymbol{L}_i^Q (\boldsymbol{R}_i^Q)^\top, \tag{19}$$

$$\boldsymbol{W}_i^{\mathrm{fuse}} = \boldsymbol{A}_i\boldsymbol{Q}_i \in \mathbb{R}^{D\times w}, \qquad \boldsymbol{Z}_i^{\mathrm{fuse}} = \widetilde{\boldsymbol{X}}_i \boldsymbol{W}_i^{\mathrm{fuse}}. \tag{20}$$

The following result characterizes this classical solution and separates its coordinate change from any change in the selected geometry.

**Proposition 6** (Procrustes orientation, proof in Appendix A.8)**.** *Let $\boldsymbol{A}_i, \boldsymbol{U}_w \in \mathbb{R}^{D\times w}$ have orthonormal columns. The factor $\boldsymbol{Q}_i$ defined above minimizes Equation* (18)*. For any orthogonal $\boldsymbol{Q}_i \in \mathbb{R}^{w\times w}$, the resulting maps satisfy*

$$\boldsymbol{W}_i^{\mathrm{fuse}} (\boldsymbol{W}_i^{\mathrm{fuse}})^\top = \boldsymbol{A}_i \boldsymbol{A}_i^\top, \qquad \boldsymbol{Z}_i^{\mathrm{fuse}} (\boldsymbol{Z}_i^{\mathrm{fuse}})^\top = \widetilde{\boldsymbol{X}}_i \boldsymbol{A}_i \boldsymbol{A}_i^\top \widetilde{\boldsymbol{X}}_i^\top. \tag{21}$$

*Thus orientation preserves the selected representations' within-graph inner products and distances, even with nonunique minimizers from a rank-deficient cross-frame matrix.*

The predictive effect of these coordinates depends on the shared encoder's parameterization and training. Both variants use thin factors to avoid constructing dense $D \times D$ operators. Local and residual PCA use randomized solvers (Halko et al., 2011), and FUSION truncates small source-factor singular values. Exact guarantees require their stated rank, optimality, and gap conditions for the operator actually optimized. Returned width can exceed positive-score rank, and zero-padding does not recover missing directions. Partial orientation preserves the geometric invariants but need not solve the full-width Procrustes problem. Appendices A.1, A.2 and B.3 give the full computational paths.

## 4 EXPERIMENTS

**Datasets and baselines.** We train on nine source graphs (Table 4) and evaluate node classification on five unseen target graphs: Cora (Yang et al., 2016), BlogCatalog (Yang et al., 2020), Squirrel (Rozemberczki et al., 2021), Computers, and Physics (Shchur et al., 2018). Support budgets are 1, 3, 5, 10, or 20 labels per class from each target's training pool. We evaluate AnyGraph (Xia & Huang, 2026), MF-GIA (Zhuo & Luo, 2026), MDGFM (Wang et al., 2025), GraphGlue (Sun et al., 2026), BRIDGE (Yuan et al., 2025), GraphLoRA (Yang et al., 2025), and GraphFM (Lachi et al., 2025) through adapted or generalist benchmark implementations (Appendix B.4). Each pipeline compares random projection (RP), randomized PCA, randomized SVD (RSVD), exact centered SVD (SVD), SPLIT G48/L16, and FUSION, all at width 64. Each interface receives fresh source training under matched training and adaptation policies, with paired target episodes. Macro-F1 averages weight targets and support budgets equally, and *avg* also averages pipelines. SD measures variation across three seed means. The exploratory 95% target-bootstrap intervals summarize variation across five

targets, distinct from seed SDs (Appendix B.5). The macro-F1 emphasis and capacity sweep are retrospective.

**Performance across GFM pipelines.** FUSION improves on exact centered SVD in all seven pipeline averages, raising mean macro-F1 from 45.75% to 47.72% (Table 1). The 1.97 percentage point (pp) gain is over the strongest tested graph-local baseline in aggregate, supporting the use of source evidence in the feature interface. Accuracy also improves by 2.55 pp, with gains in all seven pipelines (Appendix C.1). The two designs have different empirical strengths. FUSION leads on five pipelines, while SPLIT leads on MF-GIA and GraphGlue but trails SVD on BRIDGE and GraphLoRA. These results assess complete interfaces, including their effects on source training and target adaptation. The gains are obtained at the same input dimensionality, supporting source-informed preprocessing as a way to improve transfer within an existing encoder's feature budget.

Table 1: Macro-F1 (%) across seven GFM pipelines, reported as mean ± SD. Bold and underline mark the best and second-best results in each column.

| Method | AnyGraph Xia & Huang (2026) | MF-GIA Zhuo & Luo (2026) | MDGFM Wang et al. (2025) | GraphGlue Sun et al. (2026) | BRIDGE Yuan et al. (2025) | GraphLoRA Yang et al. (2025) | GraphFM Lachi et al. (2025) | avg |
|---|---|---|---|---|---|---|---|---|
| RP | 33.41±1.64 | 27.60±0.63 | 39.34±0.54 | 28.24±1.56 | 43.99±1.70 | 51.36±1.05 | 11.33±0.80 | 33.61±0.66 |
| PCA | 44.27±1.95 | 43.86±1.06 | 50.24±1.59 | 33.07±1.52 | 53.18±3.27 | 56.19±1.07 | 31.73±3.09 | 44.65±0.74 |
| RSVD | 42.82±0.95 | 43.86±1.06 | 49.89±2.05 | 34.54±3.28 | 53.20±3.31 | 56.23±1.18 | 31.73±3.09 | 44.61±0.16 |
| SVD | 45.99±1.09 | 44.44±0.62 | 50.04±2.55 | 37.04±1.23 | <u>53.91</u>±3.05 | <u>56.57</u>±0.70 | 32.26±2.87 | 45.75±1.03 |
| SPLIT | <u>47.65</u>±0.52 | **45.91**±0.94 | <u>50.44</u>±0.19 | **39.74**±1.77 | 52.77±3.38 | 55.68±1.00 | <u>32.37</u>±3.16 | <u>46.37</u>±0.78 |
| FUSION | **51.02**±0.78 | <u>45.64</u>±1.24 | **53.45**±0.75 | <u>37.47</u>±2.75 | **55.07**±1.83 | **57.21**±0.68 | **34.14**±3.02 | **47.72**±0.42 |

**Source–local design choices.** Local variation remains important even when source consensus is available. The consensus-only endpoint is the weakest of the five allocations in every pipeline (Table 2a), averaging 32.51% macro-F1 against 45.79% for the local endpoint and 46.37% for G48/L16. Yet the best allocation varies. BRIDGE and GraphLoRA favor the local endpoint, whereas AnyGraph and GraphGlue favor G48/L16. Thus source evidence can complement local variation, but assigning it more capacity does not consistently help. This motivates examining joint selection without a prescribed split.

Table 2: Source–local design analysis. (a) SPLIT capacity allocations and macro-F1 (%). (b) Ablation effects (variant minus FUSION, pp). Bold marks column bests, and † the main setting.

| (a) Common/local | AnyGraph | MF-GIA | MDGFM | GraphGlue | BRIDGE | GraphLoRA | GraphFM | avg |
|---|---|---|---|---|---|---|---|---|
| 0/64 | 46.58 | 44.16 | 50.67 | 36.93 | **53.59** | **56.90** | 31.73 | 45.79 |
| 16/48 | 44.62 | **46.30** | 50.84 | 35.74 | 52.24 | 56.82 | 32.21 | 45.54 |
| 32/32 | 46.75 | 45.20 | **51.37** | 36.96 | 51.43 | 56.20 | **32.62** | 45.79 |
| 48/16† | **47.65** | 45.91 | 50.44 | **39.74** | 52.77 | 55.68 | 32.37 | **46.37** |
| 64/0 | 35.97 | 24.58 | 39.41 | 27.67 | 38.55 | 48.83 | 12.54 | 32.51 |

| (b) Variant | AnyGraph | GraphLoRA | GraphFM | avg | 95% target CI |
|---|---|---|---|---|---|
| Without orientation | +0.09 | −1.16 | +0.62 | −0.15 | [−0.95, +0.73] |
| Local + orientation | −0.58 | −0.68 | +0.05 | −0.40 | [−0.67, −0.19] |
| Consensus only | −15.23 | −9.04 | −21.13 | −15.13 | [−26.22, −5.94] |

The ablations on AnyGraph, GraphLoRA, and GraphFM probe the roles of selection and orientation (Table 2b and Appendix C.3). FUSION exceeds the oriented local-selection variant by 0.40 pp on average, although that variant slightly outperforms it on GraphFM. Consensus-only processing loses 15.13 pp, reinforcing the importance of local evidence. Joint selection therefore adds a modest average benefit on this subset. Orientation has a less consistent predictive effect. Omitting it helps AnyGraph and GraphFM but hurts GraphLoRA, and its aggregate target-bootstrap interval includes zero. Preserving the selected geometry (Proposition 6) does not guarantee better predictions. These retrained-pipeline contrasts do not assign independent shares of the full interface gain to selection and orientation.

**Transfer across targets and label budgets.** The gains across pipeline averages do not extend uniformly to individual targets. FUSION gains 7.04 pp on Physics but loses 1.88 pp on Computers, with the same respective signs in every pipeline (Table 3 and Figure 4). Its overall gain remains positive when any single target is omitted, so it is not attributable solely to one favorable target (Appendix C.4). Nevertheless, target-bootstrap intervals for both designs include zero. The evidence supports an average improvement on the evaluated collection, while five targets do not establish a general advantage across graph domains.

The aggregate FUSION benefit persists beyond the smallest label budget. Averaged across pipelines and targets, it exceeds SVD at all five support sizes by 1.31 to 2.53 pp (Table 3b). More labels do not eliminate the aggregate advantage within this range, although its magnitude is not monotonic. SPLIT does not share this pattern and slightly trails SVD at 20 labels per class.

Table 3: Transfer across targets and label budgets. (a) Macro-F1 differences from SVD (pp). (b) Macro-F1 (%) by support labels per class, reported as mean ± SD.

| (a) Method | Cora | Computers | Physics | BlogCatalog | Squirrel | avg | 95% target CI |
|---|---|---|---|---|---|---|---|
| SPLIT | +0.32 | −1.53 | +4.31 | −0.01 | −0.01 | +0.62 | [−0.86, +2.59] |
| FUSION | +3.84 | −1.88 | +7.04 | +0.84 | −0.01 | +1.97 | [−0.59, +4.99] |

| (b) Method | 1 | 3 | 5 | 10 | 20 |
|---|---|---|---|---|---|
| SVD | 30.93±1.66 | 40.83±2.16 | 47.25±0.32 | 53.31±0.42 | 56.43±0.81 |
| SPLIT | 32.04±1.07 | 42.15±2.09 | 47.80±1.79 | 53.49±0.42 | 56.36±0.16 |
| FUSION | 33.24±1.27 | 43.36±1.37 | 49.02±1.50 | 54.63±0.66 | 58.32±1.70 |

**Local consistency after adaptation.** The retrospectively selected BRIDGE–Cora case illustrates local class consistency after complete adaptation (Figure 3). In the original normalized space, class-balanced ten-neighbor agreement averages 65.9% for FUSION versus 62.1% for SVD over three seeds (Appendix C.5). This agreement is higher than SVD's in each seed, although the gain varies across seeds. The t-SNE panels show the same query nodes, with class overlap remaining. They are fitted independently, so positions and global cluster scales are not comparable. The neighborhood statistic measures local consistency rather than accuracy. This complete-pipeline case neither isolates component effects nor validates the input-geometry theory.

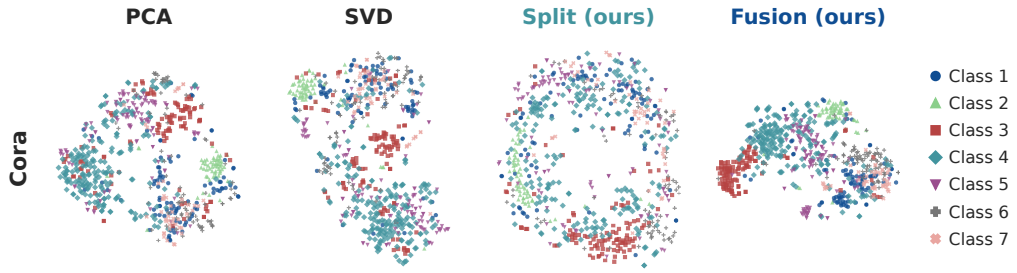



Figure 3: t-SNE visualization of a GFM's (BRIDGE) representations on Cora.

## 5 CONCLUSION

Source evidence can inform which feature directions reach a shared GFM encoder, as well as the model learned from those features. SoLIFT makes this connection through a reusable source subspace state, combining recurring source geometry with graph-local variation under an explicit positional feature convention. The distinction between selection and coordinate choice is central: selection changes the retained subspace, whereas orientation preserves its node representations' within-graph geometry while changing the coordinates seen by the encoder. Our theory characterizes exact subspace selection under stated rank and spectral-gap conditions, relating departures from local PCA to source coupling. Across seven representative GFM pipelines, the resulting interfaces achieve better overall transfer performance than the tested graph-local preprocessing baselines on unseen targets. These results support treating the fixed-width feature interface as part of transfer design. Source recurrence and local variation guide feature selection, while coordinate choice determines how the selected node representations reach the shared encoder during source training and target adaptation.

## AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on research methodology, refine mathematical statements, assist in writing proofs, and interpret archived experimental results. Additionally, we used these tools for literature retrieval, manuscript drafting and revision, and preparation of analysis code, tables, and figures. Citation metadata were checked against archived primary-source records, and numerical summaries were recomputed from retained experimental records. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

## REPRODUCIBILITY STATEMENT

We provide the theoretical assumptions and proofs in Appendix A and document the data, feature transforms, and benchmark protocols in Appendix B. To support verification of the reported results, we include archived experimental records, provenance manifests, and analysis scripts with the manuscript source. These scripts reconstruct numerical summaries, uncertainty estimates, and experimental plots from retained outcomes and stored visualization coordinates. The package supports reanalysis without rerunning model training and does not include a complete training release. Appendix B.5 describes the analysis workflow and the additional benchmark code, pinned implementations, data, and software environments needed to reproduce model training.

## ETHICS STATEMENT

This study uses existing graph benchmarks without collecting new data from human participants. The experiments assess predictive performance and do not evaluate fairness, privacy protection, or suitability for consequential decisions about people. Use and redistribution of the benchmark data remain subject to their access terms, applicable permissions, and privacy requirements.

## REFERENCES

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1–2):151–175, 2010. doi: 10.1007/s10994-009-5152-4. URL https://doi.org/10.1007/s10994-009-5152-4.

Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 245–250. ACM, 2001. doi: 10.1145/502512.502546. URL https://doi.org/10.1145/502512.502546.

Moshe Eliasof, Krishna Sri Ipsit Mantri, Beatrice Bevilacqua, Bruno Ribeiro, and Carola-Bibiane Schönlieb. Bridging input feature spaces towards graph foundation models. In *International Conference on Learning Representations*, pp. 157197–157229, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/feaf23597f3c93803d1b4230aadc8f44-Abstract-Conference.html.

Jianqing Fan, Dong Wang, Kaizheng Wang, and Ziwei Zhu. Distributed estimation of principal eigenspaces. *Annals of Statistics*, 47(6):3009–3031, 2019. doi: 10.1214/18-AOS1713. URL https://doi.org/10.1214/18-aos1713.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011. doi: 10.1137/090771806. URL https://epubs.siam.org/doi/10.1137/090771806.

Zhenyu Hou, Haozhan Li, Yukuo Cen, Jie Tang, and Yuxiao Dong. GraphAlign: Pretraining one graph neural network on multiple graphs via feature alignment. arXiv preprint arXiv:2406.02953, 2024. URL https://arxiv.org/abs/2406.02953.

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, pp. 22118–22133. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065):20150202, 2016. doi: 10.1098/rsta.2015.0202. URL https://doi.org/10.1098/rsta.2015.0202.

Divyansha Lachi, Mehdi Azabou, Vinam Arora, and Eva L. Dyer. GraphFM: A generalist graph transformer that learns transferable representations across diverse domains. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=sZTpRfRUtR.

Zeyu Li, Kangxiang Qin, Yong He, Wang Zhou, and Xinsheng Zhang. Knowledge transfer across multiple principal component analysis studies. arXiv preprint arXiv:2403.07431, 2024. URL https://arxiv.org/abs/2403.07431.

Hao Liu, Jiarui Feng, Lecheng Kong, Ningyue Liang, Dacheng Tao, Yixin Chen, and Muhan Zhang. One for all: Towards training one graph model for all classification tasks. In *International Conference on Learning Representations*, pp. 20188–20210, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/57faf5642eb06e0602b95f6aa989b38a-Abstract-Conference.html.

Jiawei Liu, Cheng Yang, Zhiyuan Lu, Junze Chen, Yibo Li, Mengmei Zhang, Ting Bai, Yuan Fang, Lichao Sun, Philip S. Yu, and Chuan Shi. Graph foundation models: Concepts, opportunities and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):5023–5044, 2025. doi: 10.1109/TPAMI.2025.3548729. URL https://doi.org/10.1109/TPAMI.2025.3548729.

Eric F. Lock, Katherine A. Hoadley, J. S. Marron, and Andrew B. Nobel. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Annals of Applied Statistics*, 7(1):523–542, 2013. doi: 10.1214/12-AOAS597. URL https://doi.org/10.1214/12-aoas597.

Haitao Mao, Zhikai Chen, Wenzhuo Tang, Jianan Zhao, Yao Ma, Tong Zhao, Neil Shah, Mikhail Galkin, and Jiliang Tang. Position: Graph foundation models are already here. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34670–34692, 2024. URL https://proceedings.mlr.press/v235/mao24a.html.

Péter Mernyei and Cătălina Cangea. Wiki-CS: A wikipedia-based benchmark for graph neural networks. In *Graph Representation Learning and Beyond Workshop, ICML*, 2020. URL https://grlplus.github.io/papers/32.pdf.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1e2agrFvS.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of GNNs under heterophily: Are we really making progress? In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tJbbQfw-5wv.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(2):cnab014, 2021. doi: 10.1093/comnet/cnab014. URL https://doi.org/10.1093/comnet/cnab014.

Ignacio Santamaría, Louis L. Scharf, Chris Peterson, Michael Kirby, and Joseph M. Francos. An order fitting rule for optimal subspace averaging. In *IEEE Statistical Signal Processing Workshop*, pp. 1–4, 2016. doi: 10.1109/SSP.2016.7551843. URL https://doi.org/10.1109/SSP.2016.7551843.

Peter H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi: 10.1007/BF02289451. URL https://doi.org/10.1007/BF02289451.

Lianze Shan, Jitao Zhao, Dongxiao He, Siqi Liu, Jiaxu Cui, and Weixiong Zhang. LEDA: Latent semantic distribution alignment for multi-domain graph pre-training. In *Proceedings of the ACM Web Conference 2026*, pp. 614–625, 2026. doi: 10.1145/3774904.3792126. URL https://doi.org/10.1145/3774904.3792126.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. In *Relational Representation Learning Workshop, NeurIPS*, 2018. URL https://www.cs.cit.tum.de/daml/gnn-benchmark/.

Yangyi Shen, Jincheng Zhou, Beatrice Bevilacqua, Joshua Robinson, Charilaos Kanatsoulis, Jure Leskovec, and Bruno Ribeiro. Zero-shot generalization of GNNs over distinct attribute domains. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 54701–54731, 2025. URL https://proceedings.mlr.press/v267/shen25p.html.

Naichen Shi and Raed Al Kontar. Personalized PCA: Decoupling shared and unique features. *Journal of Machine Learning Research*, 25(41):1–82, 2024. URL https://jmlr.org/papers/v25/22-0810.html.

Li Sun, Zhenhao Huang, Silei Chen, Lanxu Yang, Junda Ye, Sen Su, and Philip S. Yu. Multi-domain riemannian graph gluing for building graph foundation models. In *International Conference on Learning Representations*, pp. 37129–37169, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/3e9dfea4bb53006684177ba49ac11ec7-Abstract-Conference.html.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z. URL https://doi.org/10.1007/s10208-011-9099-z.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL https://jmlr.org/papers/v9/vandermaaten08a.html.

Shuo Wang, Bokui Wang, Zhixiang Shen, Boyan Deng, and Zhao Kang. Multi-domain graph foundation models: Robust knowledge transfer via topology alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64806–64821, 2025. URL https://proceedings.mlr.press/v267/wang25dj.html.

Lianghao Xia and Chao Huang. AnyGraph: Graph foundation model in the wild. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 882–896, 2026. doi: 10.18653/v1/2026.findings-acl.44. URL https://aclanthology.org/2026.findings-acl.44/.

Renchi Yang, Jieming Shi, Xiaokui Xiao, Yin Yang, Juncheng Liu, and Sourav S. Bhowmick. Scaling attributed network embedding to massive graphs. *Proceedings of the VLDB Endowment*, 14(1):37–49, 2020. doi: 10.14778/3421424.3421430. URL https://doi.org/10.14778/3421424.3421430.

Zhe-Rui Yang, Jindong Han, Chang-Dong Wang, and Hao Liu. GraphLoRA: Structure-aware contrastive low-rank adaptation for cross-graph transfer learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, pp. 1785–1796, 2025. doi: 10.1145/3690624.3709186. URL https://doi.org/10.1145/3690624.3709186.

Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 40–48. PMLR, 2016. URL https://proceedings.mlr.press/v48/yanga16.html.

Xingtong Yu, Zechuan Gong, Chang Zhou, Yuan Fang, and Hui Zhang. SAMGPT: Text-free graph foundation model for multi-domain pre-training and cross-domain adaptation. In *Proceedings of the ACM on Web Conference 2025*, pp. 1142–1153, 2025. doi: 10.1145/3696410.3714828. URL https://doi.org/10.1145/3696410.3714828.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008. URL https://doi.org/10.1093/biomet/asv008.

Haonan Yuan, Qingyun Sun, Junhua Shi, Xingcheng Fu, Bryan Hooi, Jianxin Li, and Philip S. Yu. How much can transfer? BRIDGE: Bounded multi-domain graph foundation model with generalization guarantees. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 73604–73644, 2025. URL https://proceedings.mlr.press/v267/yuan25h.html.

Haihong Zhao, Aochuan Chen, Xiangguo Sun, Hong Cheng, and Jia Li. All in one and one for all: A simple yet effective method towards cross-domain graph pretraining. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4443–4454, 2024a. doi: 10.1145/3637528.3671913. URL https://doi.org/10.1145/3637528.3671913.

Jitao Zhao, Di Jin, Meng Ge, Lianze Shan, Xin Wang, Dongxiao He, and Zhiyong Feng. FUG: Feature-universal graph contrastive pre-training for graphs with diverse node features. In *Advances in Neural Information Processing Systems*, volume 37, pp. 4003–4034, 2024b. doi: 10.52202/079017-0131. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/075b7d4bd7fc32d9cf468a7b67c38d15-Abstract-Conference.html.

Wei Zhuo and Siqiang Luo. Modality-free graph in-context alignment. In *International Conference on Learning Representations*, pp. 21310–21336, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/23fcc63005ac1a6e460ec4e209d17607-Abstract-Conference.html.

# A PROOFS AND ADDITIONAL THEORETICAL STATEMENTS

After specifying coordinates, thin-factor computation, and numerical rank conventions, this appendix proves the six results in Section 3 in their order of appearance. Additional analyses cover independent-source concentration, shared linear maps, and the limits of geometric guarantees. Throughout, $\mathbf{\Pi}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{U}^\top$ for an orthonormal frame $\boldsymbol{U}$. Exact spectral objectives are the reference for the randomized numerical implementation. These operator-level results do not order downstream predictors or guarantee target accuracy. Source graphs are $g_1, \ldots, g_k$, and $g_T$ denotes the target graph. Unless stated otherwise, sums over source indices run from 1 to $k$.

## A.1 COORDINATES, THIN FACTORS, AND NUMERICAL CONVENTIONS

**Coordinates and local rank.** The embedded features and returned local projectors are defined in Equations (2) and (4). The analytical ambient dimension $D = \max\{w, d_1, \ldots, d_k, d_T\}$ is a bookkeeping convention. The implementation fits source factors at $D_S = \max_{1\le j\le k} d_j$ and transforms graph $i$ at $D_i^{\mathrm{work}} = \max(D_S, d_i)$. A wider target zero-extends the stored source state without refitting, leaving target-only columns available locally. Equal-index native columns define the shared coordinates, but padding does not infer semantic correspondence between attributes.

The native centered PCA frame (Jolliffe & Cadima, 2016) has requested rank $r_i = \min\{w, N_i, d_i\}$. Its zero-extension $\boldsymbol{B}_i$ supplies coordinates, while $\boldsymbol{P}_i = \boldsymbol{B}_i\boldsymbol{B}_i^\top$ describes the returned subspace independently of basis orientation. Its nominal rank is $r_i$ (equal to $w$ in the regular case), which need not equal the rank of the score matrix $\widetilde{\boldsymbol{X}_i}\boldsymbol{B}_i$. Centering can leave fewer than $r_i$ positive singular values, so the requested rank is deliberately distinguished from this data-carrying rank.

**Exact thin-factor computation.** The spectral constructions in Equations (5) and (14) admit exact, untruncated thin-factor paths. With $M = \sum_{j=1}^k r_j$, collect the source frames in $\boldsymbol{F}_S \in \mathbb{R}^{D\times M}$. Both SVDs below use diagonal singular-value matrices ordered nonincreasingly before selecting columns:

$$\begin{aligned} \boldsymbol{F}_S &= k^{-1/2}[\boldsymbol{B}_1, \ldots, \boldsymbol{B}_k], \qquad \boldsymbol{F}_S = \boldsymbol{U}_S^F \boldsymbol{S}_S^F (\boldsymbol{V}_S^F)^\top, \\ \boldsymbol{C}_S &= \boldsymbol{F}_S\boldsymbol{F}_S^\top = \boldsymbol{U}_S^F (\boldsymbol{S}_S^F)^2 (\boldsymbol{U}_S^F)^\top, \qquad \boldsymbol{U}_r = [u_{S1}^F, \ldots, u_{Sr}^F]. \end{aligned} \tag{22}$$

For any nonempty source index set $\mathcal{J}$, define $\boldsymbol{F}_{\mathcal{J}} = |\mathcal{J}|^{-1/2}[\boldsymbol{B}_j]_{j\in\mathcal{J}}$ and $\boldsymbol{C}_{\mathcal{J}} = |\mathcal{J}|^{-1}\sum_{j\in\mathcal{J}} \boldsymbol{P}_j$. Using $\mathcal{J}_i$ from Equation (12), FUSION combines this factor with the current graph's frame:

$$\begin{aligned} \boldsymbol{F}_i^{\mathrm{joint}} &= [\boldsymbol{B}_i, \boldsymbol{F}_{\mathcal{J}_i}], \qquad \boldsymbol{F}_i^{\mathrm{joint}} = \boldsymbol{L}_i^{\mathrm{joint}} \boldsymbol{S}_i^{\mathrm{joint}} (\boldsymbol{V}_i^{\mathrm{joint}})^\top, \\ \boldsymbol{P}_i + \boldsymbol{C}_i &= \boldsymbol{F}_i^{\mathrm{joint}} (\boldsymbol{F}_i^{\mathrm{joint}})^\top, \qquad \boldsymbol{A}_i = [l_{i1}^{\mathrm{joint}}, \ldots, l_{iw}^{\mathrm{joint}}]. \end{aligned} \tag{23}$$

**Lemma 7** (Exact thin-factor spectral constructions)**.** *The exact, untruncated source and joint factors in Equations* (22) *and* (23) *have Gram operators $\boldsymbol{C}_S$ and $\boldsymbol{P}_i + \boldsymbol{C}_i$, respectively. Their ordered left singular vectors therefore provide leading spectral frames for the corresponding operators, provided the requested columns are available. A thin SVD of a $D \times m$ factor supplies $\min(D, m)$ columns. Selection beyond its positive rank uses null completions. At a tied boundary, different decompositions may select different leading projectors, and no unique basis is asserted.*

*Proof of Lemma 7.* Write the untruncated SVD as $\boldsymbol{F}_{\mathcal{J}} = \boldsymbol{U}_{\mathcal{J}}^F \boldsymbol{S}_{\mathcal{J}}^F (\boldsymbol{V}_{\mathcal{J}}^F)^\top$. Orthonormality of the right singular vectors gives

$$\boldsymbol{C}_{\mathcal{J}} = \boldsymbol{F}_{\mathcal{J}}\boldsymbol{F}_{\mathcal{J}}^\top = (\boldsymbol{U}_{\mathcal{J}}^F \boldsymbol{S}_{\mathcal{J}}^F)(\boldsymbol{U}_{\mathcal{J}}^F \boldsymbol{S}_{\mathcal{J}}^F)^\top.$$

Concatenating the local frame adds $\boldsymbol{P}_i$ to this Gram matrix, proving Equation (23). Reversing the concatenation order or replacing a source factor by its untruncated weighted left factor preserves the operator. Positive singular values square to its nonzero eigenvalues. Dropping positive source singular values is an approximation. Tied boundaries and null-space completions need not yield identical frames. □

**Stored state and numerical widths.** SPLIT stores its common frame $\boldsymbol{U}_g$, while FUSION stores weighted source factors and the full-source reference $\boldsymbol{U}_w$. A source factor requires $O(DM)$ storage, and caching all $k$ excluded-source factors can require $O(kDM)$. The weighted factors determine selection, whereas the truncated reference only sets coordinates. Substituting that reference changes

the fusion objective. The numerical rules below specify returned widths and distinguish exact properties from implementation.

In SPLIT, the source-factor thin SVD returns $g_{\mathrm{ret}} = \min\{g, D_S, \sum_j r_j\}$ common columns, without filtering its singular values by positivity. At $g = 0$, the common fit is skipped. Residual PCA uses the actual returned common frame and requests $w - g$ columns, returning $r_i^{\mathrm{loc}} = \min\{w - g, N_i, D_i^{\mathrm{work}}\}$. Missing common columns are not reassigned to the local branch: the $g_{\mathrm{ret}} + r_i^{\mathrm{loc}}$ score columns are concatenated and zeros are appended to reach width $w$.

For FUSION, consider either the full or excluded source set $\mathcal{J}$ and let $m_{\mathcal{J}} = \sum_{j \in \mathcal{J}} r_j$. The thin SVD of $\boldsymbol{F}_{\mathcal{J}}$ retains singular values satisfying

$$s_a > \varepsilon_{\mathrm{mach}} \max(D_S, m_{\mathcal{J}}) s_1,$$

where $s_1$ is the largest singular value and $\varepsilon_{\mathrm{mach}}$ is machine precision. An empty spectrum retains no columns. Write the retained factor as $\bar{\boldsymbol{U}}_{\mathcal{J}} \bar{\boldsymbol{S}}_{\mathcal{J}}$ with $h_{\mathcal{J}}$ columns, defining the numerical source operator

$$\boldsymbol{C}_{\mathcal{J}}^{\mathrm{num}} = (\bar{\boldsymbol{U}}_{\mathcal{J}} \bar{\boldsymbol{S}}_{\mathcal{J}})(\bar{\boldsymbol{U}}_{\mathcal{J}} \bar{\boldsymbol{S}}_{\mathcal{J}})^{\top}.$$

FUSION uses this entire weighted factor, zero-extended as needed, rather than its leading width-$w$ reference. Its fused operator is therefore $\boldsymbol{P}_i + \boldsymbol{C}_{\mathcal{J}_i}^{\mathrm{num}}$. The thin SVD of $[\bar{\boldsymbol{U}}_{\mathcal{J}_i} \bar{\boldsymbol{S}}_{\mathcal{J}_i}, \boldsymbol{B}_i]$ returns $f_i = \min\{w, D_i^{\mathrm{work}}, h_{\mathcal{J}_i} + r_i\}$ selected columns, and no further positivity threshold is applied to this fused spectrum.

**Partial coordinate orientation.** The full-source reference has $\rho = \min(w, h_S)$ columns. Set $q_i = \min(f_i, \rho)$, and partition the returned fused frame as $\boldsymbol{A}_i = [\boldsymbol{A}_{i,1}, \boldsymbol{A}_{i,2}]$, with $q_i$ columns in $\boldsymbol{A}_{i,1}$. For $q_i > 0$, the partial alignment is

$$\boldsymbol{A}_{i,1}^{\top} \boldsymbol{U}_{\mathrm{ref},1} = \boldsymbol{L} \boldsymbol{\Gamma} \boldsymbol{R}^{\top}, \qquad \boldsymbol{Q} = \boldsymbol{L} \boldsymbol{R}^{\top}, \qquad \boldsymbol{W}_i^{(0)} = [\boldsymbol{A}_{i,1} \boldsymbol{Q}, \boldsymbol{A}_{i,2}],$$

where $\boldsymbol{U}_{\mathrm{ref},1}$ contains the first $q_i$ full-source reference columns. If $q_i = 0$, the frame is retained without rotation. Trailing columns are signed so their largest absolute loading is positive. Aligned columns whose absolute overlap with their corresponding reference column is at most $\varepsilon_{\mathrm{mach}} \max(D_i^{\mathrm{work}}, q_i)$ use the same convention as well.

Blockwise rotations and column sign changes preserve the returned projector and node-score Gram matrix in exact arithmetic. Partial alignment and the near-zero sign fallback do not in general assert an exact minimizer of the full width-$w$ Procrustes objective. The resulting $f_i$ score columns are padded with zeros to width $w$, without asserting recovery of a rank-$w$ subspace.

### A.2 RETURNED RANK AND ACTIVE DIRECTIONS

For a symmetric matrix, eigenvalues are ordered nonincreasingly. $\mathrm{Top}_r(\boldsymbol{M})$ denotes an orthonormal frame spanning a leading rank-$r$ subspace, while $\boldsymbol{\Pi}_{\mathrm{Top}_r(\boldsymbol{M})}$ is its projector. At a boundary tie the frame and, potentially, the leading projector are set-valued. All rank-$r$ variational statements assume the ambient dimension admits such a frame. A zero-padded output matrix of width $w$ is not necessarily the score matrix of a recovered rank-$w$ subspace.

To distinguish nominal and data-carrying directions, the exact decomposition $\boldsymbol{P}_i = \boldsymbol{P}_i^{+} + \boldsymbol{P}_i^{0}$ uses the positive and zero singular values of $\widetilde{\boldsymbol{X}}_i \boldsymbol{B}_i$. Here $\boldsymbol{P}_i^{+}$ contains directions with positive score energy, whereas $\boldsymbol{P}_i^{0}$ records a completion chosen only to meet the nominal output rank. In numerical diagnostics, an explicitly reported tolerance replaces exact zero, so near-zero scores are only approximately annihilated. Changing basis inside the returned frame diagonalizes its score covariance without changing $\boldsymbol{P}_i$.

For a nonempty exact null completion, $\|\boldsymbol{P}_i^{0}\|_{\mathrm{op}} = 1$ but $\widetilde{\boldsymbol{X}}_i \boldsymbol{P}_i^{0} = 0$. In the extreme case $\widetilde{\boldsymbol{X}}_i = 0$, every representation has zero feature energy even if a solver returns $w$ orthonormal columns. Hence a nominal projector objective cannot generally be equated with retained data information.

### A.3 CONSENSUS CHARACTERIZATION

The shared frame maximizes average agreement with the source subspaces. The same optimization defines a barycenter of their projectors, as stated in Proposition 1.

*Proof of Proposition 1.* Since $r_i = \operatorname{rank}(\boldsymbol{P}_i)$, $\boldsymbol{P}_i^2 = \boldsymbol{P}_i$, and $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_r$,

$$\|\boldsymbol{P}_i - \boldsymbol{U}\boldsymbol{U}^\top\|_F^2 = r_i + r - 2\operatorname{tr}(\boldsymbol{U}^\top \boldsymbol{P}_i \boldsymbol{U}).$$

Averaging over $i$ leaves a constant minus $2\operatorname{tr}(\boldsymbol{U}^\top \boldsymbol{C}_S \boldsymbol{U})$. Ky Fan's maximum principle identifies every leading rank-$r$ frame as a minimizer. A strict boundary eigengap identifies its projector. Unequal returned ranks change only the constant term. □

### A.4 CONDITIONAL CONSENSUS ESTIMATION

The variational characterization identifies an optimizer of empirical subspace agreement. Theorem 2 gives sufficient conditions for estimating a specified common population subspace: shared directions dominate average private geometry, and misspecification plus estimation error is less than half the resulting separation. These conditions do not assert predictive relevance.

*Proof of Theorem 2.* Define $\boldsymbol{M}_0 = \boldsymbol{U}_\star \bar{\boldsymbol{\Omega}} \boldsymbol{U}_\star^\top + \bar{\boldsymbol{L}}$ with $\bar{\boldsymbol{\Omega}} = k^{-1}\sum_i \boldsymbol{\Omega}_i$. The range of $\boldsymbol{U}_\star$ and its orthogonal complement reduce $\boldsymbol{M}_0$: the eigenvalues on $\boldsymbol{U}_\star$ are at least $\bar{a}$, and those on the complement are at most $\|\bar{\boldsymbol{L}}\|_{\mathrm{op}}$. Its leading rank-$q$ projector is therefore $\boldsymbol{U}_\star \boldsymbol{U}_\star^\top$, separated by at least $\delta$.

Write

$$\boldsymbol{C}_S - \boldsymbol{M}_0 = k^{-1}\sum_i \boldsymbol{\Xi}_i + k^{-1}\sum_i (\boldsymbol{P}_i - \boldsymbol{\Pi}_i).$$

Its norm is at most $\epsilon$. Weyl's inequality gives separation between the perturbed leading cluster and the unperturbed complementary cluster of at least $\delta - \epsilon$. The Davis–Kahan sin-angle bound gives

$$\|\boldsymbol{\Pi}_{\mathrm{Top}_q(\boldsymbol{C}_S)} - \boldsymbol{U}_\star \boldsymbol{U}_\star^\top\|_{\mathrm{op}} \le \frac{\epsilon}{\delta - \epsilon} \le \frac{2\epsilon}{\delta}$$

under $\epsilon < \delta/2$ (Yu et al., 2015). This is a conditional operator statement and makes no identification of $g = 48$ with $q$. □

**Sampling and numerical error.** The theorem's empirical error term combines statistical and algorithmic effects with different behavior as the number of sources grows. For a specified analysis rank, let $\widetilde{\boldsymbol{P}}_i$ be an exact empirical PCA projector, $\boldsymbol{\Pi}_i$ its ideal population counterpart, and $\boldsymbol{\mathcal{P}}_i = \mathbb{E}\widetilde{\boldsymbol{P}}_i$. Choose a measurable tie convention if needed. Then

$$\boldsymbol{P}_i - \boldsymbol{\Pi}_i = (\boldsymbol{\mathcal{P}}_i - \boldsymbol{\Pi}_i) + (\widetilde{\boldsymbol{P}}_i - \boldsymbol{\mathcal{P}}_i) + (\boldsymbol{P}_i - \widetilde{\boldsymbol{P}}_i).$$

Define $\boldsymbol{\mathcal{B}}_i = \boldsymbol{\mathcal{P}}_i - \boldsymbol{\Pi}_i$, $\boldsymbol{Y}_i = \widetilde{\boldsymbol{P}}_i - \boldsymbol{\mathcal{P}}_i$, and $\boldsymbol{F}_i = \boldsymbol{P}_i - \widetilde{\boldsymbol{P}}_i$. The last term contains numerical and rank-policy mismatch, and a nominal null completion can prevent it from being small. Let

$$\bar{h} = k^{-1}\sum_i h_i, \qquad \bar{b} = k^{-1}\sum_i \|\boldsymbol{\mathcal{B}}_i\|_{\mathrm{op}},$$

$$\bar{f} = k^{-1}\sum_i \|\boldsymbol{F}_i\|_{\mathrm{op}}, \qquad v = \left\| k^{-2}\sum_i \mathbb{E}\boldsymbol{Y}_i^2 \right\|_{\mathrm{op}}.$$

**Corollary 8** (Independent-source concentration)**.** *Assume the source sampling errors $\boldsymbol{Y}_i$ are independent. For $0 < \eta < 1$, with probability at least $1 - \eta$ their average has norm at most*

$$t_\eta = \sqrt{2v\log(2D/\eta)} + \frac{2\log(2D/\eta)}{3k}.$$

*On this sampling event, every realized set of numerical outputs for which $e = \bar{h} + \bar{b} + \bar{f} + t_\eta < \delta/2$ obeys $\|\boldsymbol{\Pi}_{\mathrm{Top}_q(\boldsymbol{C}_S)} - \boldsymbol{U}_\star \boldsymbol{U}_\star^\top\|_{\mathrm{op}} \le 2e/\delta$.*

*Proof of Corollary 8.* Both $\widetilde{\boldsymbol{P}}_i$ and $\boldsymbol{\mathcal{P}}_i$ are positive-semidefinite contractions, so $-\boldsymbol{I} \preceq \boldsymbol{Y}_i \preceq \boldsymbol{I}$, and $\mathbb{E}\boldsymbol{Y}_i = 0$. Self-adjoint matrix Bernstein with summands $\boldsymbol{Y}_i/k$, norm bound $1/k$, and variance parameter $v$ gives a two-sided tail bounded by $2D\exp[-t^2/(2(v + t/(3k)))]$. Substitution of $t_\eta$ makes this at most $\eta$ (Tropp, 2012). The triangle inequality and Theorem 2 complete the proof. The event concerns sampling errors, while the realized numerical error remains in the small-error condition. □

Independent sources do not imply independent nodes within a graph. If $v = O(k^{-1})$, centered operator fluctuation has an $O(k^{-1/2})$ term (up to logarithms). Neither this observation nor the corollary removes bias, coordinate mismatch, or a solver-error floor. We do not apply an independent-row covariance rate to the real graph nodes.

### A.5 Conditional residual optimality

With the common frame and its allocated width fixed, exact residual PCA maximizes the remaining score energy. The proof of Proposition 3 also explains why the implementation retains the residual projector even when zero-score completion columns are returned.

*Proof of Proposition 3.* Choose an orthonormal complement $\boldsymbol{V}_g^{\mathrm{comp}} \in \mathbb{R}^{D\times(D-g)}$ of $\boldsymbol{U}_g$. Every feasible residual frame is $\boldsymbol{V} = \boldsymbol{V}_g^{\mathrm{comp}}\boldsymbol{W}$, where $\boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}_{w-g}$. The objective becomes

$$\mathrm{tr}\big(\boldsymbol{W}^\top(\boldsymbol{V}_g^{\mathrm{comp}})^\top\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V}_g^{\mathrm{comp}}\boldsymbol{W}\big).$$

It is maximized by a leading rank-$(w-g)$ frame of the covariance in these complementary coordinates:

$$(\boldsymbol{V}_g^{\mathrm{comp}})^\top\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{V}_g^{\mathrm{comp}}.$$

Positive residual eigenvalues identify active directions, with a boundary gap required for a uniquely selected leading active subspace. If its positive rank is below $w-g$, any orthonormal null completion inside the complement is optimal but unidentifiable from the data.

For a positive residual singular value $\tau$, its right singular vector $v$ satisfies

$$\boldsymbol{R}_g\widetilde{\boldsymbol{X}}_i^\top\widetilde{\boldsymbol{X}}_i\boldsymbol{R}_g v = \tau^2 v, \qquad \boldsymbol{R}_g v = v.$$

The second equality follows by applying the idempotent $\boldsymbol{R}_g$ to the first. Null residual directions instead satisfy $\widetilde{\boldsymbol{X}}_i\boldsymbol{R}_g v = 0$. When the thin SVD supplies the requested $w-g$ columns,

$$\frac{1}{N_i}\|\widetilde{\boldsymbol{X}}_i\boldsymbol{R}_g\widetilde{\boldsymbol{V}}_i\|_F^2 = \frac{1}{N_i}\sum_{a=1}^{w-g}\tau_{ia}^2.$$

This is the optimal residual score energy, even when projected null completion columns are not orthonormal. Keeping $\boldsymbol{R}_g$ in the mapping defined in Equation (10) annihilates any common component of these columns. Randomized PCA approximates this exact variational reference. □

### A.6 Source–local compensation

Comparing FUSION's joint optimum with the feasible local frame quantifies the tradeoff between local and source agreement in Proposition 4.

*Proof of Proposition 4.* The argument holds for any $0 \preceq \boldsymbol{C} \preceq \boldsymbol{I}$. Define $J(\boldsymbol{V}) = \mathrm{tr}(\boldsymbol{V}^\top(\boldsymbol{P}+\boldsymbol{C})\boldsymbol{V})$. A rank-$w$ local projector admits an orthonormal frame $\boldsymbol{B} \in \mathbb{R}^{D\times w}$ with $\boldsymbol{B}\boldsymbol{B}^\top = \boldsymbol{P}$. This frame is feasible for the joint objective, so

$$J(\boldsymbol{B}) = w + \mathrm{tr}(\boldsymbol{P}\boldsymbol{C}), \qquad J(\boldsymbol{A}) \geq J(\boldsymbol{B}), \qquad a - \ell = J(\boldsymbol{A}) - J(\boldsymbol{B}) \geq 0.$$

Since $\boldsymbol{P}$ is a contraction and $\boldsymbol{A}$ has $w$ orthonormal columns, $\mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{P}\boldsymbol{A}) \leq w$, hence $0 \leq \ell \leq a$. The feasible baseline requires returned rank $w$ but need not have $w$ positive-variance directions. Lower-rank outputs require a separate comparison.

For SoLIFT, take the nonempty source set $\mathcal{J}_i$ from Equation (12), covering both source exclusion and all-source target use. With $\boldsymbol{C}_i = |\mathcal{J}_i|^{-1}\sum_{j\in\mathcal{J}_i}\boldsymbol{P}_j$, the same objective is equivalent to minimizing

$$\|\boldsymbol{P}_i - \boldsymbol{A}\boldsymbol{A}^\top\|_F^2 + \frac{1}{|\mathcal{J}_i|}\sum_{j\in\mathcal{J}_i}\|\boldsymbol{P}_j - \boldsymbol{A}\boldsymbol{A}^\top\|_F^2.$$

Expansion yields a constant minus $2J_i(\boldsymbol{A})$, so Ky Fan's principle gives the leading width-$w$ frame of $\boldsymbol{P}_i + \boldsymbol{C}_i$ in Equation (14). □

**Interpretation and numerical scope.** With thresholded source factors, the same comparison uses $\boldsymbol{C} = \boldsymbol{C}^{\mathrm{num}}_{\mathcal{J}_i}$ and an exact width-$w$ maximizer of $\boldsymbol{P}_i + \boldsymbol{C}^{\mathrm{num}}_{\mathcal{J}_i}$, provided the proposition's hypotheses hold. Its source agreement term then concerns the truncated operator and is not automatically the agreement with $\boldsymbol{C}_i$.

The compensation result compares unweighted projector overlap rather than reconstruction fidelity. For a graph with $N$ nodes and embedded features $\widetilde{\boldsymbol{X}}$, write $\widehat{\boldsymbol{\Sigma}} = N^{-1}\widetilde{\boldsymbol{X}}^\top\widetilde{\boldsymbol{X}}$. For any orthonormal width-$w$ frame,

$$L^{\mathrm{data}}(\boldsymbol{A}) = \frac{1}{N}\|\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{X}}\boldsymbol{A}\boldsymbol{A}^\top\|_F^2 = \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}) - \mathrm{tr}(\boldsymbol{A}^\top\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}).$$

Its optimum is $L^{\mathrm{opt}}_w = \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}) - \sum_{j=1}^{w}\lambda_j(\widehat{\boldsymbol{\Sigma}})$. Ky Fan proves $L^{\mathrm{data}}(\boldsymbol{A}) - L^{\mathrm{opt}}_w \geq 0$. Unlike Equation (15), this weights directions by score variance. Scaling $\widehat{\boldsymbol{\Sigma}}$ changes the data loss but does not change $\boldsymbol{P}$ or the compensation terms, already ruling out a universal unscaled identification of those quantities.

### A.7 Stability under source coupling

Theorem 5 analyzes the equal-coefficient operator $\boldsymbol{P} + \boldsymbol{C}$ using the coupling $b$ and block gap $\delta_0$ in Equation (16). It bounds the movement of its exact leading rank-$w$ projector and the local affinity loss $\ell$ from Equation (15). It does not analyze a separately tuned source weight.

*Proof of Theorem 5.* Decompose $\boldsymbol{C}$ into diagonal and off-diagonal blocks relative to $\boldsymbol{P}$:

$$\boldsymbol{C}_{\mathrm{diag}} = \boldsymbol{P}\boldsymbol{C}\boldsymbol{P} + (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{C}(\boldsymbol{I}-\boldsymbol{P}),$$
$$\boldsymbol{C}_{\mathrm{off}} = \boldsymbol{P}\boldsymbol{C}(\boldsymbol{I}-\boldsymbol{P}) + (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{C}\boldsymbol{P}.$$

Write $[\boldsymbol{P},\boldsymbol{C}] = \boldsymbol{P}\boldsymbol{C} - \boldsymbol{C}\boldsymbol{P}$. In coordinates adapted to $\boldsymbol{P}$, the off-diagonal block and commutator are

$$\boldsymbol{C}_{\mathrm{off}} \equiv \begin{bmatrix} 0 & \boldsymbol{B}^\top \\ \boldsymbol{B} & 0 \end{bmatrix}, \qquad [\boldsymbol{P},\boldsymbol{C}] \equiv \begin{bmatrix} 0 & \boldsymbol{B}^\top \\ -\boldsymbol{B} & 0 \end{bmatrix}.$$

Here $\boldsymbol{B} \in \mathbb{R}^{(D-w)\times w}$ represents coupling between the two subspaces, and both norms equal $\|\boldsymbol{B}\|_{\mathrm{op}} = b$. By the Rayleigh characterization, the minimum and maximum in Equation (16) are the extreme eigenvalues of the source operator compressed to the two subspaces. When $\delta_0 > 0$, the block-diagonal matrix $\boldsymbol{K}_0 = \boldsymbol{P} + \boldsymbol{C}_{\mathrm{diag}}$ has leading projector $\boldsymbol{P}$ and gap $\delta_0$. Because $\boldsymbol{P} + \boldsymbol{C} = \boldsymbol{K}_0 + \boldsymbol{C}_{\mathrm{off}}$, the same perturbation argument as in Theorem 2 yields

$$\|\boldsymbol{\Pi}_{\boldsymbol{A}} - \boldsymbol{P}\|_{\mathrm{op}} \leq \frac{b}{\delta_0 - b} \leq \frac{2b}{\delta_0}$$

when $b < \delta_0/2$. For principal angles $\theta_1, \ldots, \theta_w$ between the two subspaces, the second bound is

$$\ell = \sum_{a=1}^{w}\sin^2\theta_a \leq w\|\boldsymbol{\Pi}_{\boldsymbol{A}} - \boldsymbol{P}\|_{\mathrm{op}}^2.$$

If $b = 0$, then $[\boldsymbol{P},\boldsymbol{C}] = 0$, so the two symmetric matrices share an eigenbasis. In $\mathrm{range}(\boldsymbol{P})$ the fused eigenvalues are $1 + c_j \geq 1$, whereas outside they are $c_j \leq 1$ because $\boldsymbol{C}$ is a contraction. Without a boundary tie the leading $w$ directions are precisely those of $\boldsymbol{P}$. This proves the no-insertion statement, including its nominal-rank qualification. □

**Scope of the bound.** Weak source–local coupling relative to the block gap keeps FUSION close to local PCA. The conditions are sufficient rather than necessary, and a boundary tie can make the selected projector nonunique. In rank-deficient data, this result concerns a rank-$w$ returned nominal projector, including its null completion, and does not assert the same bound for a lower-rank returned projector. Movement is an operator diagnostic rather than a measure of retained feature energy. For an exact maximizer using $\boldsymbol{C}^{\mathrm{num}}_{\mathcal{J}_i}$, the coupling and gap must be computed for that operator and satisfy the same conditions. A bound computed for the untruncated $\boldsymbol{C}_i$ does not automatically apply.

**Bounds using the fused eigengap.** The movement theorem uses the block gap $\delta_0$, which depends on the local projector. A related bound can be stated through the eigengap of the fused operator. Let $\Delta = \lambda_w(\boldsymbol{P}+\boldsymbol{C}) - \lambda_{w+1}(\boldsymbol{P}+\boldsymbol{C})$. Weyl gives $|\Delta - \delta_0| \le 2b$, so $\Delta > 4b$ implies

$$\|\boldsymbol{\Pi}_{\boldsymbol{A}} - \boldsymbol{P}\|_{\mathrm{op}} \le \frac{2b}{\Delta - 2b}.$$

If the leading projector is unique, it commutes with $\boldsymbol{P}+\boldsymbol{C}$. Therefore

$$\begin{aligned} b &= \|[\boldsymbol{P}+\boldsymbol{C}-\boldsymbol{I}, \boldsymbol{P}-\boldsymbol{\Pi}_{\boldsymbol{A}}]\|_{\mathrm{op}} \\ &\le 2\|\boldsymbol{P}+\boldsymbol{C}-\boldsymbol{I}\|_{\mathrm{op}}\|\boldsymbol{P}-\boldsymbol{\Pi}_{\boldsymbol{A}}\|_{\mathrm{op}} \\ &\le 2\|\boldsymbol{P}-\boldsymbol{\Pi}_{\boldsymbol{A}}\|_{\mathrm{op}}, \end{aligned}$$

since $0 \preceq \boldsymbol{P}+\boldsymbol{C} \preceq 2\boldsymbol{I}$. Thus $b/2 \le \|\boldsymbol{P}-\boldsymbol{\Pi}_{\boldsymbol{A}}\|_{\mathrm{op}}$. These bounds quantify subspace movement. They do not establish whether the selected directions are relevant to target labels.

### A.8 Source exclusion and Procrustes orientation

**Source exclusion.** Source exclusion and orientation serve different roles. Exclusion prevents a source graph from reinforcing its own projector during subspace selection, whereas Procrustes chooses coordinates after that subspace has been selected. For $k \ge 2$, the exact, untruncated projector means obey

$$\boldsymbol{C}_S - \boldsymbol{C}_{-i} = (\boldsymbol{P}_i - \boldsymbol{C}_S)/(k-1), \qquad \|\boldsymbol{C}_S - \boldsymbol{C}_{-i}\|_{\mathrm{op}} \le 1/(k-1),$$

because both $\boldsymbol{P}_i$ and $\boldsymbol{C}_S$ are contractions. If a positive gap of at least $\Delta_*$ persists along the segment between the corresponding full-source and excluded-source fusion operators, subdividing that segment and applying the perturbation bound yields a projector-distance bound $\min\{1, 2/((k-1)\Delta_*)\}$. Independently thresholding the two source factors need not preserve the exclusion identity, so this comparison does not automatically bound the resulting numerical operators. The shared full-source orientation reference is not part of this exclusion identity.

*Proof of Proposition 6.* For orthonormal frames $\boldsymbol{A}, \boldsymbol{U} \in \mathbb{R}^{D\times w}$ and orthogonal $\boldsymbol{Q} \in \mathbb{R}^{w\times w}$, the objective in Equation (18) expands to

$$\|\boldsymbol{A}\boldsymbol{Q} - \boldsymbol{U}\|_F^2 = 2w - 2\,\mathrm{tr}(\boldsymbol{Q}^\top \boldsymbol{A}^\top \boldsymbol{U}).$$

Compute $\boldsymbol{A}^\top\boldsymbol{U} = \boldsymbol{L}\boldsymbol{\Gamma}\boldsymbol{R}^\top$. Von Neumann's trace inequality bounds the trace by $\mathrm{tr}(\boldsymbol{\Gamma})$, attained at $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{R}^\top$, as used in Proposition 6. Rank deficiency can make this minimizer nonunique. Every orthogonal $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}_w$, giving

$$(\boldsymbol{A}\boldsymbol{Q})(\boldsymbol{A}\boldsymbol{Q})^\top = \boldsymbol{A}\boldsymbol{A}^\top, \qquad (\widetilde{\boldsymbol{X}}\boldsymbol{A}\boldsymbol{Q})(\widetilde{\boldsymbol{X}}\boldsymbol{A}\boldsymbol{Q})^\top = \widetilde{\boldsymbol{X}}\boldsymbol{A}\boldsymbol{A}^\top\widetilde{\boldsymbol{X}}^\top.$$

These are the projector and node-Gram invariants in Proposition 6, which preserve the inner products and distances of the already selected node representations. □

**Shared linear maps.** The following result separates coordinate invariance of an unrestricted linear map from the simultaneous constraints imposed by sharing that map across graphs.

**Proposition 9** (Absorption of graph-specific coordinate changes)**.** *For each graph $i \in \{1,\dots,k,T\}$, let $\boldsymbol{Z}_i \in \mathbb{R}^{N_i\times w}$ and let $\boldsymbol{Q}_i \in \mathbb{R}^{w\times w}$ be orthogonal. Fix a shared first-layer weight $\boldsymbol{\Theta} \in \mathbb{R}^{w\times h}$, where $h \ge 1$. A replacement weight $\boldsymbol{\Theta}' \in \mathbb{R}^{w\times h}$ preserves all first-layer outputs after the changes $\boldsymbol{Z}_i \mapsto \boldsymbol{Z}_i\boldsymbol{Q}_i$ if and only if*

$$\boldsymbol{Z}_i(\boldsymbol{Q}_i\boldsymbol{\Theta}' - \boldsymbol{\Theta}) = 0 \qquad \textit{for every } i.$$

*If every $\boldsymbol{Z}_i$ has full column rank, this condition reduces to $\boldsymbol{Q}_i\boldsymbol{\Theta}' = \boldsymbol{\Theta}$ for every $i$, and these equations need not admit a shared solution. A common coordinate change $\boldsymbol{Q}_i = \boldsymbol{Q}$ always admits $\boldsymbol{\Theta}' = \boldsymbol{Q}^\top\boldsymbol{\Theta}$. With separate unrestricted weights $\boldsymbol{\Theta}_i$, each graph instead admits $\boldsymbol{\Theta}_i' = \boldsymbol{Q}_i^\top\boldsymbol{\Theta}_i$.*

*Proof of Proposition 9.* Equality of the old and new first-layer outputs is exactly $\boldsymbol{Z}_i\boldsymbol{Q}_i\boldsymbol{\Theta}' = \boldsymbol{Z}_i\boldsymbol{\Theta}$, which gives the stated annihilation condition. Full column rank makes multiplication by $\boldsymbol{Z}_i$ injective, so it is equivalent to $\boldsymbol{Q}_i\boldsymbol{\Theta}' = \boldsymbol{\Theta}$. For two full-column-rank inputs, take $\boldsymbol{Q}_1 = \boldsymbol{I}_w$, $\boldsymbol{Q}_2 = -\boldsymbol{I}_w$, and $\boldsymbol{\Theta} \ne 0$. The equations then require both $\boldsymbol{\Theta}' = \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}' = -\boldsymbol{\Theta}$, which is impossible. For a common orthogonal change, substituting $\boldsymbol{\Theta}' = \boldsymbol{Q}^\top\boldsymbol{\Theta}$ preserves each product. The same identity holds separately for $\boldsymbol{\Theta}_i' = \boldsymbol{Q}_i^\top\boldsymbol{\Theta}_i$. □

Rank-deficient inputs require only the stated annihilation condition. These are hypothesis-class identities, not optimization equivalences, and do not apply to a fresh classifier located after a frozen nonlinear encoder.

### A.9 Predictive limits of geometric alignment

Target labeling can remain ambiguous even when feature coordinates are fixed and known. Consider a target with observable scalar feature $x$ drawn from a symmetric distribution with no mass at zero and identical uninformative topology in two worlds. Fix all source observations. In world one the target label is $\mathbf{1}\{x > 0\}$, whereas in world two it is its complement. The unlabeled input to a preprocessing rule has the same distribution in both worlds. For any resulting predictor that sees no target labels or additional coupling, its binary errors in the two worlds sum to one, pointwise, and therefore at least one expected error is at least $1/2$. The source data cannot resolve which target labeling world holds. The obstruction concerns target labeling, even with identical feature geometry. Target support labels can distinguish the worlds, but enter the subsequent GFM adaptation rather than our preprocessing step.

Classical binary domain-adaptation bounds likewise separate source risk, post-transform marginal discrepancy, and the best joint labeling error in a specified hypothesis class (Ben-David et al., 2010). Projector affinity alone does not control the latter two terms. Such bounds assume a common binary example and label space, including topology or neighborhood context if used by the predictor. They are not performance guarantees for this benchmark's heterogeneous multiclass, support-adapted macro-F1. Our results instead characterize the interface's geometric objectives and conditional movement.

## B Data, transforms, and benchmark pipelines

### B.1 Benchmark unit, data, and target access

**Comparison unit.** A benchmark pipeline comprises a model implementation, its source-training procedure, and its target-adaptation procedure. The seven pipeline names in the main table identify these complete benchmark implementations, rather than unchanged reproductions of the published systems. Their objectives, target heads, and computational substitutions differ. We therefore compare preprocessing within each pipeline, pairing target episodes across interfaces.

**Graph data.** Features remain in their archived native column order. Graphs use undirected coalesced edges, integer node labels, and graph-specific class spaces. Dataset tensor hashes accompany the evidence package. The data inventory and splits are defined by the benchmark registry rather than being pipeline-specific. Cora and Physics differ from their source relatives, but this is not a leave-one-domain-family-out benchmark: related source and target families can coexist.

Table 4 identifies the source and target graphs. Actor uses the Geom-GCN release, and Chameleon and Squirrel use its classification preprocessing (Pei et al., 2020). BlogCatalog uses the attributed release distributed on the PANE authors' dataset page (Yang et al., 2020). The benchmark uses the splits defined below. Inventory edge counts are directed COO entries after undirected coalescing, so an undirected pair generally contributes two entries.

**Episodes and target access.** For seeds 0, 1, and 2, class-stratified splits assign 50% of nodes to a training pool, 25% to validation, and 25% to test. Each target support set selects the requested number per class from its training pool. Within a seed, test nodes form the same query set across transforms and shot counts, and support membership is paired across transforms at each shot count. The seeds vary the node split, support sampling, and model randomness, as well as randomized transforms where applicable. Labels outside the support set are not exposed to adaptation. Unlabeled features and graph edges are available for the full target graph, including query nodes. Each SoLIFT interface computes a target-local transform from these unlabeled features while leaving its fitted source state unchanged. Neither interface consumes support indices or labels, and transformed target features are reused across shot counts. Source labels may be used by a pipeline's source-training procedure, so *label-free* describes preprocessing only.

Table 4: Source and target graphs. Citations identify benchmark releases, while statistics describe the archived, preprocessed graphs used here.

| Graph | Role | Nodes | Edges | Features | Classes |
|---|---|---|---|---|---|
| citeseer (Yang et al., 2016) | Source | 3,327 | 9,104 | 3703 | 6 |
| pubmed (Yang et al., 2016) | Source | 19,717 | 88,648 | 500 | 3 |
| arxiv (Hu et al., 2020) | Source | 169,343 | 2,315,598 | 128 | 40 |
| wikics (Mernyei & Cangea, 2020) | Source | 11,701 | 431,206 | 300 | 10 |
| amazon–ratings (Platonov et al., 2023) | Source | 24,492 | 186,100 | 300 | 5 |
| amazon–photo (Shchur et al., 2018) | Source | 7,650 | 238,162 | 745 | 8 |
| coauthor–cs (Shchur et al., 2018) | Source | 18,333 | 163,788 | 6805 | 15 |
| actor (Pei et al., 2020) | Source | 7,600 | 53,318 | 932 | 5 |
| chameleon (Rozemberczki et al., 2021) | Source | 2,277 | 62,742 | 2325 | 5 |
| cora (Yang et al., 2016) | Target | 2,708 | 10,556 | 1433 | 7 |
| amazon–computers (Shchur et al., 2018) | Target | 13,752 | 491,722 | 767 | 10 |
| coauthor–physics (Shchur et al., 2018) | Target | 34,493 | 495,924 | 8415 | 5 |
| blogcatalog (Yang et al., 2020) | Target | 5,196 | 343,486 | 8189 | 6 |
| squirrel (Rozemberczki et al., 2021) | Target | 5,201 | 396,706 | 2089 | 5 |

### B.2 Matched preprocessing comparison

**Matched policies.** For every pipeline, preprocessing method, and model seed, a fresh source model is trained under the same pipeline-specific hyperparameters, training caps, checkpoint-selection rules, and early-stopping policy. Actual stopping epochs and selected parameters can differ across methods. Target support and query episodes are paired across methods.

**Source fitting and training.** SoLIFT first centers each source and computes its native local frame, then zero-extends the frames and fits the full-source consensus. SPLIT retains $g$ common directions and computes source residual scores, whereas FUSION combines each source's local frame with evidence from the other sources and orients the result to the full-source reference. The source model is then trained on the transformed source features, and its checkpoint is saved.

**Target transformation and adaptation.** For a target, the stored source state is zero-extended as needed. SPLIT computes common and residual-PCA scores, whereas FUSION jointly selects its subspace from target-local PCA and all-source evidence and orients it to the full-source reference. Both interfaces emit width-$w$ scores. Finally, the pipeline-specific target-adaptation procedure uses the requested support labels, and evaluation uses the fixed queries. The target is never added to source fitting. Source exclusion applies only to FUSION's joint selection operator. Its orientation reference remains fitted to all sources.

### B.3 Transform definitions and implementation conventions

All main comparison arms output width 64.

**RP.** A Gaussian graph-local map (Bingham & Mannila, 2001) is seeded by transform seed and input width and scaled by $1/\sqrt{d_i}$. It acts on uncentered raw $\boldsymbol{X}_i$. Maps are reused for equal widths under this seed convention, without asserting semantic correspondence.

**PCA.** Center each graph and use `torch.pca_lowrank` with sketch width $\min(\min(N_i, d_i), w + 8)$ and the default two power iterations. If reduction is unnecessary, use exact SVD. Form centered scores in the returned right-singular frame.

**RSVD.** Center each graph and use `torch.svd_lowrank` with eight oversampling columns and two power iterations, again falling back to exact SVD when appropriate. Some PCA and RSVD cells coincide. Both approximate the same centered low-rank objective.

**SVD.** Apply exact SVD to the centered matrix, with scores $\boldsymbol{L}_i^x(:, 1{:}r_i)\,\mathrm{diag}(\sigma_{i1}, \ldots, \sigma_{ir_i})$, using the local-SVD notation and returned rank $r_i$ of Equation (3). Deterministic cache entries use transform seed zero while model seeds still vary. For all three local spectral baselines, orient score columns by the sign of their centered third moment, leave a zero sign positive, and append zeros when $r_i < w$.

**SPLIT.** Use the full-source, no-target-truncation implementation with deterministically seeded local PCA and loading-based sign orientation. Local seeds depend on transform seed and shape. Source frames are computed at nominal width 64. Thin consensus SVD supplies the common frame, and each graph's residual PCA supplies the remaining width. G48/L16 is the single main-table configuration. Five allocations are reported descriptively.

**FUSION.** Derive local PCA seeds deterministically from the transform seed and graph identifier, with scoped random-number state. Basis signs use the largest absolute loading. Full and source-excluded consensus factors are constructed from the nominal projectors. Targets wider than all sources extend the source factors with zeros. FUSION uses thin SVD, Procrustes uses the full-source reference, and output padding supplies the requested width when necessary.

**Numerical comparability.** SPLIT's G0/L64 arm is a local-PCA endpoint under its own solver and seed policy, so it is not expected to reproduce the separately implemented PCA baseline bitwise. SPLIT and FUSION likewise use different randomized local-basis conventions. The archived comparisons include these implementation choices and do not isolate one symbolic objective while holding every numerical detail identical.

### B.4 GFM INTERFACES, TRAINING, AND ADAPTATION

For each model, we distinguish the published feature interface, relevant release-code choices, and the width-64 boundary used in our benchmark. The subsequent training and adaptation settings specify the evaluated pipeline. Retained launch configurations and available provenance are included in the accompanying evidence files.

**AnyGraph (Xia & Huang, 2026).** The published model constructs fixed-dimensional inputs from SVD of node features and normalized adjacency before graph propagation and expert residual encoders. Its released feature transform uses uncentered SVD with square-root singular-value scaling. Our benchmark supplies external width-64 features through the original-feature path, bypassing both native SVD branches. It retains class-node augmentation, feature-coordinate flipping, graph propagation, and expert residual encoders. Class nodes have zero input features, while support-to-class edges convey support labels. Query labels are evaluator-owned. Coordinate flipping acts before the learned encoder, so an unrestricted first linear map alone does not establish invariance to input rotations.

The source/expert model is freshly trained for each transform and model seed, with up to 150 epochs, eight experts, eight fully connected layers, batch size 4,096, source patience 12, a 30-epoch minimum, and source evaluation every five epochs. The checkpoint maximizes mean source-validation accuracy. Target prediction uses the source-trained model without optimizer updates.

**MF-GIA (Zhuo & Luo, 2026).** The method unifies feature dimensions before its GNN and gradient-fingerprint computation, then uses domain-conditioned modulation of encoded features. The paper specifies graph-wise SVD, whereas the pinned released loader applies graph-wise PCA when the feature width differs from the requested width. Our benchmark replaces this initial unification step with the selected width-64 interface for source and target graphs, before the backbone and fingerprint modules.

The canonical runner builds the released backbone, domain embedder, and episodic pretraining model. Source class identifiers are offset in the shared label bank, while fingerprint probes retain graph-local class identifiers and mask labels outside source training nodes. The checkpoint minimizes the source-validation episodic loss. On a target, support-label gradients produce its domain fingerprint without target optimizer updates. Learned feature modulation acts on encoded node features, which are normalized and classified by normalized support-class prototypes. This target wrapper is the evaluated path, and the full released refinement pipeline is not claimed as reproduced. The canonical settings use up to 300 pretraining epochs, 300 fingerprint epochs, 20 training episodes, 5-shot/10-way source episodes, and patience 30.

**MDGFM (Wang et al., 2025).** Graph-wise PCA precedes domain and shared tokens, topology refinement, and GCN encoding. Its released implementation also fits PCA on the target graph before adaptation. Our benchmark replaces this PCA boundary with the external width-64 interface. Source features enter the token and topology modules, while target features enter the token-modulated GCN used by the encoder-head adaptation procedure.

The archived source run uses 60 epochs, hidden width 256, and a 1,024-node per-source cap for its training batches, while transform fitting still uses all source feature rows. The checkpoint minimizes source training loss. Target adaptation is `encoder-head`, with head learning rate $10^{-2}$ and encoder learning rate $10^{-3}$, rather than a universally frozen encoder. The source learning rate is $2 \times 10^{-3}$. The target budget is 100 epochs with patience 25, selecting the state with the lowest support loss. The benchmark implementation preserves the released source-slot architecture, but this does not imply an arbitrary-source-capacity theorem.

**GraphGlue (Sun et al., 2026).** The model learns local and global manifold representations to align graphs. Its published setup unifies source dimensions using random projection or SVD. The released source loader defaults to standardization followed by Gaussian projection, whereas the target adapter accepts raw features through a learned input map before the frozen encoder. Our benchmark supplies width-64 features to both paths before rooted-subgraph sampling and their input maps. The native comparison also applies standardization and Gaussian projection to both sources and targets, a distinction from the released target path.

All transforms use the matched fast protocol: hidden and attention widths 512, two layers, 32 generators, dropout 0.1, two-hop neighborhoods with 10 neighbors per hop, source batch size 512, and five source epochs at learning rate $3 \times 10^{-4}$ with one warmup epoch. The wrapper uses the released local/global manifold objectives and refinement logic, retaining the final source state. Target adaptation allows up to 1,000 epochs at $10^{-3}$ with patience 20 and batch size 256, selecting by the support classification loss plus geometric regularization. Its fresh unrestricted input map can absorb a target-only orthogonal coordinate change without changing the class of representable functions. Results are budget-specific and do not claim to reproduce the full original training schedule.

**BRIDGE (Yuan et al., 2025).** The method separates dimensional unification from learned feature masking and target prompting. The paper describes an SVD-plus-MLP dimension aligner, while its released node implementation applies per-graph PCA before source masks and the shared GCN, and before target prompts. Our benchmark replaces this PCA boundary with each width-64 interface and retains the subsequent masks, encoder, and prompt routing. Its graph-spectral regularization is separate from this feature reduction.

The shared GCN uses hidden width 256, three layers, and dropout 0.1. Source-specific masks precede the GCN, and target adaptation uses prompts/routing. The source cap is 10,000 epochs at learning rate $10^{-5}$ with patience 50. The checkpoint minimizes the source objective, and early stopping can use far fewer epochs. The target cap is 400 epochs at approximately $9.624 \times 10^{-4}$ with patience 50. The frozen encoder is used with prompts selected by the support-supervised adaptation objective. Other archived settings include 50 negatives and batches of 4,096 link anchors. Elementwise masks and prompts are coordinate-sensitive and do not provide arbitrary orthogonal-map absorption.

**GraphLoRA (Yang et al., 2025).** The method adapts a pretrained GNN through learned feature matching and low-rank updates. Its released code offers optional graph-local SVD before source pretraining and before the learned target projector. Our native comparison uses this uncentered SVD reduction at width 64. The other interfaces replace that reduction, supplying source features to the GAT and target features to the projector before the frozen GAT and LoRA branches. The learned projector and structure-aware distribution-matching objective remain downstream.

Our GraphLoRA-Generalist pipeline pretrains a released-style two-layer GAT by sampled GRACE on all nine sources for 300 round-robin rounds (2,700 graph updates), retaining the final state. Each target receives a fresh square projector, a rank-32 LoRA adapter, and a classifier, trained with the base model frozen and selected by support classification loss. Dense PPR and reconstruction are replaced with deterministic Andersen PPR anchor blocks ($\alpha = 0.05$, $\varepsilon = 10^{-4}$), sampled graph contrast, and balanced edge/non-edge structural loss. The GraphLoRA benchmark implementation retains the released loss form, but these computational substitutions are material. Target hyperparameters were calibrated on held-out source graphs, not target query labels, and then reused across preprocessing methods. The fresh target projector can absorb a target-only orthogonal coordinate change without

changing the class of representable functions. The shared source encoder generally cannot absorb graph-specific source rotations.

**GraphFM (Lachi et al., 2025).** The model processes graph tokens through a shared latent transformer. Its published feature tokenizer maps raw attributes through dataset-specific MLPs, without requiring PCA or SVD. In our fixed-width arms, the selected interface precedes the source-specific MLPs and the fresh target MLP. Their outputs then combine with separately computed SignNet positional encodings before the shared backbone. This preprocessing stage is added by the benchmark, while the native arm feeds raw attributes to the MLPs.

The GraphFM-Generalist-18M wrapper uses 256 latent tokens of width 256, width-64 feature tokens, 32-dimensional SignNet positional encodings, four-head cross attention, ten shared self-attention blocks, and four decoder blocks. Source-specific feature MLPs precede the shared backbone. Balanced source updates use LAMB, a cosine schedule, two warmup epochs, eight-step gradient accumulation, a maximum of 400 epochs, source validation from epoch 100, and patience 40. The checkpoint maximizes mean source-validation macro-F1. Targets receive a fresh feature MLP and classifier, trained for exactly 20 support-only steps while the backbone remains frozen. The decoder allocation is adjusted to the 18M geometry, and PyTorch scaled-dot-product attention substitutes for unavailable xFormers. Native raw-feature MLP parameter counts differ from fixed-width arms and are not a width-matched preprocessing comparison.

### B.5 Evaluation estimands and artifact reproduction

**Main-table summaries.** Write $\mathrm{F1}_{p,\tau,s,e,m}$ for macro-F1 of pipeline $p$, benchmark target $\tau$, shot count $s$, seed $e$, and method $m$. The target index $\tau \in \{1, \dots, 5\}$ enumerates the benchmark graphs, each playing the role of $g_T$ in Section 3, with $s \in \{1, 3, 5, 10, 20\}$ and $e \in \{0, 1, 2\}$. A cell in the main table is

$$\overline{\mathrm{F1}}_{p,m} = \frac{1}{3} \sum_{e} \left( \frac{1}{25} \sum_{\tau,s} 100\mathrm{F1}_{p,\tau,s,e,m} \right),$$

with sample standard deviation over the three parenthesized values. For the *avg* column, we first average the seven pipelines within each seed, then report the mean and sample standard deviation of the three seed means.

**Uncertainty across targets.** For the seven-pipeline comparisons of methods $a, b$, the target-level paired effect and its equal-target mean are

$$\Delta_{\tau}^{a,b} = \frac{1}{7 \cdot 5 \cdot 3} \sum_{p,s,e} 100(\mathrm{F1}_{p,\tau,s,e,a} - \mathrm{F1}_{p,\tau,s,e,b}),$$

$$\overline{\Delta}^{a,b} = \frac{1}{5} \sum_{\tau=1}^{5} \Delta_{\tau}^{a,b}.$$

For the three-pipeline ablations, the pipeline average instead runs over AnyGraph, GraphLoRA, and GraphFM. Bootstrap replicates average five entries sampled with replacement from $(\Delta_{\tau}^{a,b})_{\tau=1}^{5}$. Enumerating every ordered draw yields the exact empirical bootstrap distribution, from which the 2.5% and 97.5% quantiles are taken. Sign-flip replicates average $(\zeta_{\tau}\Delta_{\tau}^{a,b})$ for $\zeta_{\tau} \in \{-1, 1\}$. The two-sided value is the fraction with absolute mean at least $|\overline{\Delta}^{a,b}|$. These are target-level procedures, not resampling tests that assume independent support counts or encoders. Five targets yield coarse sign-flip resolution and limited external inference. Secondary analyses and split sweeps are descriptive.

**Retained records.** The evidence package contains 6,075 retained main/extended records, including 3,150 for the six main methods. Every pipeline-method arm has 75 unique target/shot/seed records, and the four native arms have the same per-arm count. All main SPLIT comparisons use G48/L16. No missing score is imputed. The included analysis script verifies these properties and reproduces every reported mean against archived aggregates to numerical tolerance.

**Artifact reproduction.** From the paper source directory, rebuild the evidence tables with `python scripts/build_evidence.py`. Then compile `iclr2027_conference.tex` with `latexmk -pdf`. The script uses the immutable benchmark source files when available, otherwise the bundled filtered inputs. It does not train models. Training reproduction requires the original pinned reference code, canonical graph tensors, cache manifests, and per-pipeline environments. An anonymous full-code release remains an author action. Feature caches bind source identities, graph hashes, width, rank allocation, solver settings, and seed. Precomputed target features cannot replace source-only fitting of a transform state.

# C Extended experimental results

The following analyses complement Section 4 in the same order: overall performance, source–local design choices, target effects, and learned representations. They use the retained benchmark records and the aggregation rules in Appendix B.5.

## C.1 Accuracy and native-interface comparisons

**Accuracy.** The main text reports macro-F1 for the six width-64 preprocessing arms. Accuracy tests whether the aggregate improvement also appears under a second metric, while native interfaces provide pipeline-specific context. Accuracy was the original registry's primary metric, whereas the emphasis on macro-F1 and the allocation analysis are retrospective. Reaggregating the archived episodes gives mean accuracy $47.86 \pm 1.12\%$ for SVD and $50.40 \pm 0.50\%$ for FUSION: +2.55 percentage points (pp). These SDs describe three seed means, each averaging seven pipelines, five targets, and five support sizes. The paired seed differences have SD 0.68 pp. FUSION exceeds SVD in all seven pipeline accuracy means.

Table 5: Accuracy (%) for the six interfaces, averaged equally over targets, support sizes, and seeds.

| Pipeline | RP | PCA | RSVD | SVD | SPLIT | FUSION |
|---|---|---|---|---|---|---|
| AnyGraph | 36.67 | 43.60 | 42.50 | 45.47 | 49.40 | 53.41 |
| MF-GIA | 31.19 | 46.73 | 46.73 | 47.35 | 48.83 | 48.08 |
| MDGFM | 41.25 | 51.00 | 50.45 | 50.93 | 51.36 | 54.70 |
| GraphGlue | 35.13 | 39.23 | 40.59 | 42.09 | 44.85 | 43.05 |
| BRIDGE | 46.28 | 54.60 | 54.62 | 55.24 | 54.28 | 56.66 |
| GraphLoRA | 53.53 | 56.72 | 56.70 | 56.79 | 56.29 | 57.89 |
| GraphFM | 21.50 | 36.88 | 36.88 | 37.12 | 38.56 | 39.04 |

**Native-interface comparisons.** The native interfaces in Table 6 follow Appendix B.4, including the documented benchmark substitutions. They provide context for four pipelines rather than a comparison with every original published system. Their partial-coverage mean is not ranked against the full-coverage methods.

Table 6: Native interfaces and width-64 alternatives: macro-F1 (%, mean $\pm$ seed SD).

| Pipeline | Native | SVD | SPLIT | FUSION |
|---|---|---|---|---|
| GraphGlue | $21.68_{\pm 1.58}$ | $37.04_{\pm 1.23}$ | $39.74_{\pm 1.77}$ | $37.47_{\pm 2.75}$ |
| BRIDGE | $54.03_{\pm 3.20}$ | $53.91_{\pm 3.05}$ | $52.77_{\pm 3.38}$ | $55.07_{\pm 1.83}$ |
| GraphLoRA | $56.73_{\pm 0.67}$ | $56.57_{\pm 0.70}$ | $55.68_{\pm 1.00}$ | $57.21_{\pm 0.68}$ |
| GraphFM | $33.03_{\pm 1.54}$ | $32.26_{\pm 2.87}$ | $32.37_{\pm 3.16}$ | $34.14_{\pm 3.02}$ |

## C.2 Source–local allocation

The complete SPLIT sweep tests how capacity allocation affects each pipeline at width 64 (Table 2a). G48/L16 remains the main setting, while the sweep is descriptive and does not select a separate allocation for each pipeline. G0/L64 uses SPLIT's local solver and seed convention, so it need not reproduce a separately run PCA or SVD baseline exactly. These comparisons assess capacity allocation at the tested width, not a common latent rank.

### C.3 SELECTION AND ORIENTATION ABLATIONS

**Interventions and comparison scope.** The complete three-variant comparison covers AnyGraph, GraphLoRA, and GraphFM. Variants reuse FUSION's fitted source state and, where needed, its local frames and orientation reference. Local plus orientation replaces joint selection with the local frame, whereas consensus-only removes the current graph's local term. The variant without orientation retains the fused subspace but omits its final rotation. Each intervention applies to source and target preprocessing, followed by fresh downstream training under the same pipeline policy. Each variant is paired with FUSION on the same target/shot/seed cells before averaging within target and then across five targets. These conditional contrasts do not form a full factorial decomposition. Intervals use target-cluster bootstrap over these five targets. Implementation conventions follow Appendix B.3.

Table 7: Three-pipeline ablations: macro-F1 (%) and variant minus FUSION (pp).

| Variant | F1 | Δ vs. FUSION | 95% interval |
|---|---|---|---|
| Without orientation | 47.30 | -0.15 | [-0.95, +0.73] |
| Local + orientation | 47.06 | -0.40 | [-0.67, -0.19] |
| Consensus only | 32.32 | -15.13 | [-26.22, -5.94] |

**Effects and uncertainty.** Table 7 reports aggregate scores and effects, while Table 2b gives the per-pipeline effects. The orientation interval crosses zero, and its exact two-sided target-level sign-flip value is $p = 0.9375$. The other two intervals lie below zero, but both have $p = 0.0625$, the smallest possible two-sided value with five nonzero target effects. Resampling assumptions are discussed in Appendix C.4. Per-pipeline effects further show why the aggregate should be interpreted cautiously: removing orientation helps AnyGraph and GraphFM but hurts GraphLoRA, while local plus orientation slightly exceeds FUSION on GraphFM. These comparisons support retaining local evidence and show a modest average benefit from joint selection with orientation retained on this subset. They do not establish a uniform benefit from orientation or explain the full seven-pipeline gain over SVD.

### C.4 TARGET EFFECTS AND UNCERTAINTY

**Target heterogeneity.** Figure 4 resolves the FUSION–SVD effects by pipeline: every pipeline gains on Physics and loses on Computers. Other targets show mixed responses. These descriptive means average over support sizes and seeds.

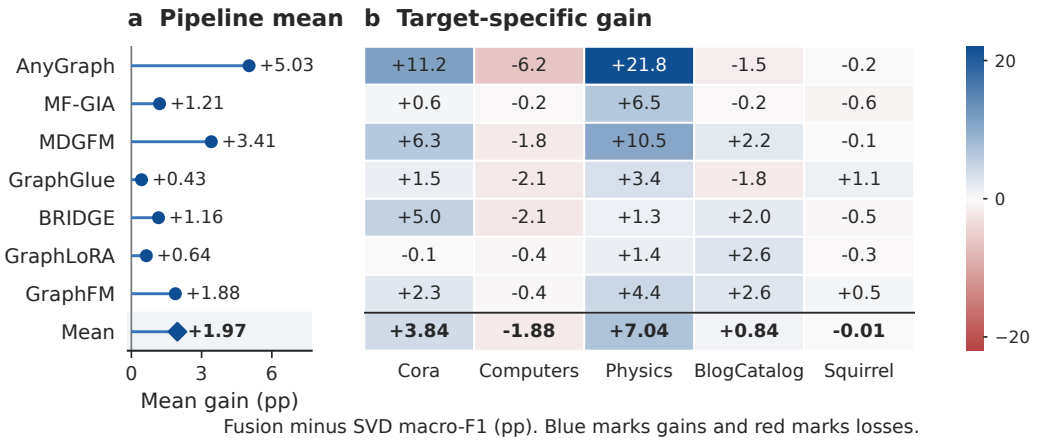



Figure 4: FUSION minus SVD in macro-F1 (pp), averaged over targets (a) and resolved by target (b). The bottom row averages pipelines.

Table 8 aggregates over the same seven pipelines, five support counts, and three seeds used in the main comparison. FUSION LOTO reports the mean over the other four targets when the named target is omitted. All five FUSION LOTO means remain positive, ranging from 0.70 to 2.93 pp. SPLIT's

remaining mean is $-0.31$ when Physics is excluded, showing how a favorable aggregate can depend on target composition.

Table 8: Target-level macro-F1 differences from SVD and leave-one-target-out (LOTO) means (pp).

| Target | SPLIT–SVD | FUSION–SVD | FUSION LOTO |
|---|---|---|---|
| Cora | +0.32 | +3.84 | +1.50 |
| Computers | -1.53 | -1.88 | +2.93 |
| Physics | +4.31 | +7.04 | +0.70 |
| BlogCatalog | -0.01 | +0.84 | +2.25 |
| Squirrel | -0.01 | -0.01 | +2.46 |

**Uncertainty across targets.** Each primary contrast contains 525 paired pipeline/target/shot/seed cells but only five target graphs. We first average differences within each target and then resample the five target-level effects. The exploratory 95% percentile interval is $[-0.59, 4.99]$ pp for FUSION–SVD and $[-0.86, 2.59]$ for SPLIT–SVD. Exact two-sided sign-flip values on the same effects are $p = 0.375$ and $p = 0.750$, respectively. The bootstrap treats targets as exchangeable sampling units, and the sign-flip test assumes sign exchangeability under its null. The selected targets are not a random sample of graph domains, so these exploratory quantities do not establish population-wide improvement. This target-level uncertainty differs from the three-seed SDs reported in Tables 1 and 3.

### C.5 BRIDGE–CORA REPRESENTATION ANALYSIS

**Case and visualization protocol.** For Figure 3, we replay support-only target prompt adaptation from the original source-selected checkpoints. The case was selected retrospectively after inspecting MDGFM, BRIDGE, and GraphFM on Cora and Physics. The displayed episode uses the same 681 query nodes across methods, five support labels per class, and model seed 0.

Final classifier-input representations are L2-normalized before t-SNE (van der Maaten & Hinton, 2008) is fitted independently for each method with a common random initialization (seed 17), perplexity 30, 2,000 iterations, early exaggeration 12, and automatic learning rate. Panel coordinates and global cluster sizes are not compared across fits. The neighborhood score uses Euclidean distances between normalized query representations, excluding self-neighbors. Same-class fractions are averaged within each class and then equally across classes. Query labels are used only for coloring and scoring, not adaptation, neighbor selection, or t-SNE fitting. Diagnostics use saved replays, whose scores can differ slightly from the archived results in Table 1.

**Neighborhood consistency.** For the BRIDGE–Cora case in Figure 3, we quantify local class consistency before t-SNE in the normalized 256-dimensional classifier-input space. The class-balanced same-class fraction among ten nearest query neighbors is 63.3% for FUSION versus 54.1% for SVD in the displayed episode. Three-seed means are 61.0%, 62.1%, 56.2%, and 65.9% for PCA, SVD, SPLIT, and FUSION, respectively. FUSION exceeds SVD in each seed on this neighborhood measure. The t-SNE panels visualize the final representations, with class overlap remaining. This selected case provides evidence of improved local consistency after the complete training and adaptation procedure. The statistic measures neighborhood consistency rather than classifier accuracy. It does not isolate a FUSION component or validate the input-geometry theory.

## D RESOURCE ACCOUNTING

**Standalone preprocessing.** The common standalone preprocessing audit covers all nine sources and five targets and averages the recorded transform seeds. SPLIT has the largest recorded GPU peak in this audit. Thin-factor computation avoids dense ambient covariance matrices, but does not imply lower end-to-end memory. FUSION was measured separately and is excluded from the common audit. Its measurements below have a different scope and do not support a uniform efficiency ranking.

Table 9: Standalone preprocessing time (s) and peak memory (MiB).

| Method | Time s | GPU MiB | CPU MiB |
|---|---|---|---|
| RP | 1.70 | 2043 | 2284 |
| PCA | 2.22 | 3024 | 2521 |
| RSVD | 2.19 | 3025 | 2522 |
| SVD | 298.84 | 6839 | 2512 |
| SPLIT | 2.88 | 18532 | 2467 |

**Pipeline runtime and memory.** Table 10 distinguishes cached and cold-equivalent runtime. Cached time records execution with reusable feature caches, while cold-equivalent time adds preprocessing under each runner's accounting convention. Rows average three seeds and cover source training and target episodes. SPLIT uses G48/L16. AnyGraph–FUSION uses one cumulative endpoint per shot run and counts the reused base materialization once per seed. Its recorded totals exclude untimed shot-specific graph rewriting and work after the final result snapshot, so they are not complete end-to-end wall times. For the first five SPLIT/SVD pipelines, GPU values include the cold preprocessing peak. Generalist and FUSION rows retain runner-reported peaks with different stage coverage, and missing or zero-as-unrecorded CPU values are dashes. The recorded times and peaks therefore retain runner-specific measurement scope.

Table 10: Recorded runtime (s) and memory (MiB). Measurement scopes differ across runners.

| Pipeline | Method | Cached s | Cold s | GPU MiB | CPU MiB |
|---|---|---|---|---|---|
| AnyGraph | SVD | 1474.0 | 1772.8 | 6839 | – |
| | SPLIT | 1469.5 | 1472.3 | 18532 | – |
| | FUSION | 1465.6 | 1466.9 | 9069 | 2329 |
| MF-GIA | SVD | 1966.2 | 2265.0 | 22132 | – |
| | SPLIT | 1946.2 | 1949.0 | 22132 | – |
| | FUSION | 1976.6 | 1978.0 | 22132 | – |
| MDGFM | SVD | 50.5 | 349.3 | 6839 | – |
| | SPLIT | 49.9 | 52.7 | 18532 | – |
| | FUSION | 50.5 | 51.7 | 650 | 3058 |
| GraphGlue | SVD | 4124.8 | 4423.6 | 6839 | – |
| | SPLIT | 4093.3 | 4096.1 | 18532 | – |
| | FUSION | 4148.1 | 4149.3 | 5073 | 3689 |
| BRIDGE | SVD | 5131.9 | 5430.7 | 10572 | – |
| | SPLIT | 5125.5 | 5128.3 | 18532 | – |
| | FUSION | 5148.5 | 5149.7 | 10572 | 2936 |
| GraphLoRA | SVD | 1148.9 | 1454.3 | 18426 | 2821 |
| | SPLIT | 1156.6 | 1166.1 | 18426 | 2834 |
| | FUSION | 666.1 | 716.7 | 18414 | 2964 |
| GraphFM | SVD | 853.6 | 1191.6 | 8712 | 3158 |
| | SPLIT | 806.8 | 848.9 | 8712 | 3143 |
| | FUSION | 808.0 | 958.8 | 8709 | 3284 |