GRAPHBRIDGE: GRAPH-MEDIATED TRANSFER UNDER STRUCTURAL MODALITY MISSINGNESS

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning systems are often trained in domains with rich information but deployed in domains where an entire modality is unavailable and raw feature spaces do not align. External knowledge graphs (KGs) can provide a semantic bridge across such domains, but they can be noisy, incompletely aligned to data, and may contain relations that are useful in one population but harmful in another. We propose GraphBridge, a framework for KG-mediated transfer under structural modality missingness. GraphBridge anchors raw variables from each domain to graph node families with abstention, learns a source-target-specific transferable graph by filtering relations according to scores from literature support, source and target utility, and cross-domain stability, and constructs a low-dimensional spectral bridge for source-to-target signal transport. It then learns the transformation in the source domain from this bridge to the source-only modality and applies this transformation in the target domain. We provide guarantees for anchor and graph recovery, spectral bridge perturbation, transported-feature error, target excess risk, and a validation-protected model selection. On synthetic benchmarks, GraphBridge consistently improves over target-only prediction across various settings. On a real benchmark of PLATO pharmacogenomic transfer with BioKG, GraphBridge achieves strong improvement, reaching Pearson $R = 0.824$ on GDSC cell-line transfer and nearly matching the oracle on a PDX transfer pair with $R = 0.569$ compared to oracle $R = 0.574$. These results show that GraphBridge can support transfer via noisy external knowledge graphs across structurally mismatched domains.

1 INTRODUCTION

Machine learning models are often developed in domains with rich measurements but deployed in domains without them (Le et al., 2025; Neverova et al., 2015; Wu et al., 2024). In biology, for example, cell-line studies may include molecular assays that are unavailable in patient-derived models; in healthcare, imaging or omics may be present in one cohort but absent in another (Le et al., 2025; Wu et al., 2024); in recommendation, item metadata may differ completely across platforms (Bao et al., 2023; Wang et al., 2019). Unlike ordinary missing-at-random data, the target domain lacks an entire modality, and its raw feature space may not share coordinates with the source. This setting combines challenges from missing-modality learning and heterogeneous transfer, where the information available at training and deployment can differ structurally (Bao et al., 2023; Le et al., 2025; Wu et al., 2024).

External knowledge graphs (KGs) offer a possible bridge for this gap. KGs integrate heterogeneous entities through typed relations and have been increasingly used to organize and exploit structured knowledge in biomedical and other prediction problems (Chandak et al., 2023; Romano et al., 2024; Su et al., 2023; Timón-Reina et al., 2023; Walsh et al., 2020). Such relational structure can provide a shared semantic space even when two domains do not share raw variables. However, real-world KGs are assembled from heterogeneous sources and may involve uncertain entity mappings, incomplete or noisy relations, and context-dependent associations (Chandak et al., 2023; Su et al., 2023; Walsh et al., 2020). Consequently, treating the graph as a fixed and uniformly reliable feature source can introduce inaccurate transferred signals and lead to negative transfer.

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.

We introduce **GraphBridge**, a learning framework that treats the KG as a noisy transfer substrate to be estimated rather than a fixed feature generator. GraphBridge first anchors the raw variables of each domain to graph node families with an explicit abstention option for weak or ambiguous evidence, and filters candidate relations by a score that measures a literature prior, source- and target-side utility, and a cross-domain stability term, keeping only relations that are literature-supported, useful in both domains, and stable across them. The surviving relations form a source-target-specific transferable graph. The leading eigenvectors of the transferable graph then form a low-dimensional spectral bridge shared by both domains, on which GraphBridge learns from the source alone a transport map to a latent factor of the source-only modality and applies it to target embeddings to recover the inferable component of that modality for the target predictor. Finally, an optional validation safeguard chooses between this transfer predictor and a target-only baseline on a held-out target set to avoid negative transfer. Algorithm 1 summarizes the full procedure.

Contributions. We summarize our main contributions as follows. First, we formulate a transfer setting in which a source domain observes a useful auxiliary modality that is unavailable in the target domain, the source and target raw feature spaces may be disjoint, and the only cross-domain bridge is a noisy external knowledge graph with uncertain data-to-graph alignment. Second, we propose GraphBridge, a pipeline that anchors variables to a graph with allowed abstention, learns a source-target-specific transferable graph, constructs a low-dimensional spectral bridge representation, and transports source-only modality into the target domain, with an optional validation safeguard. Third, we provide learning guarantees for the main modules in our pipeline. Fourth, we evaluate GraphBridge on diagnostic synthetic experiments and a pharmacogenomic transfer benchmark with 20 transfer pairs for GDSC and PDX cohorts, respectively, bridged by BioKG (Walsh et al., 2020), showing significant improvements over target-only baselines.

Relation to Existing Work. GraphBridge connects three lines of previous work, including missing-modality learning, domain adaptation, and KG-enhanced prediction. Missing-modality methods study learning when one or more modalities are unavailable at inference time (Le et al., 2025; Neverova et al., 2015; Wu et al., 2024), but typically operate within a common data domain or assume shared feature semantics between training and deployment. Domain adaptation and heterogeneous transfer learning address distributional or representational mismatch across domains (Bao et al., 2023; Ben-David et al., 2010), but commonly rely on shared representations, known correspondences, or some predefined cross-domain structure. KG-enhanced prediction instead exploits relational knowledge as an auxiliary inductive bias or representation (Ruiz et al., 2023; Wang et al., 2019), typically assuming that the graph and the mapping from observed variables or entities into that graph are given. GraphBridge targets their intersection, where the useful source modality is absent from the target, the raw feature spaces may have zero overlap, and the external KG is noisy enough that both alignment and transferability must be learned. A full discussion of the related work is in Section B.

2 Problem Setup

Learning Setting. We consider supervised transfer from a source domain to a target domain. The source sample $\mathcal{S} = \{(X_i^S, Z_i^S, Y_i^S) : i = 1, \ldots, n_S\} \sim P_S$ contains observed source features X^S, a source-only modality Z^S, and, when available, source labels Y^S. The target sample $\mathcal{T} = \{(X_j^T, Y_j^T) : j = 1, \ldots, n_T\} \sim P_T$ contains labels for the downstream target task but does not observe the source-only modality. The raw feature spaces $X^{(S)} \subseteq \mathbb{R}^{p_S}$ and $X^{(T)} \subseteq \mathbb{R}^{p_T}$ are not required to have overlapped features. There is an external knowledge graph $G = (V, E, R, \kappa)$, where V is the node set, E the edge set, R the relation types, and $\kappa : V \to \mathcal{C}$ maps nodes to entity categories. We use it as a noisy relational prior rather than as a trusted feature map.

GraphBridge operates through node families, which serve as an intermediate semantic space between raw variables and KG entities. Raw variables in either domain may align to one of M candidate node families or to a null anchor when no reliable alignment exists. This distinction is critical in our method since the observed KG is an external object with noise, where it may contain false edges, omit true relations, or include relations that are relevant in one domain but not another. We therefore distinguish the observed KG from the source-target-specific transferable graph used by the model. Informally, compared to a naive observed KG, a graph is transferable when it is supported by external evidence, useful for the source-only signal and target task, and stable across domains. This separation

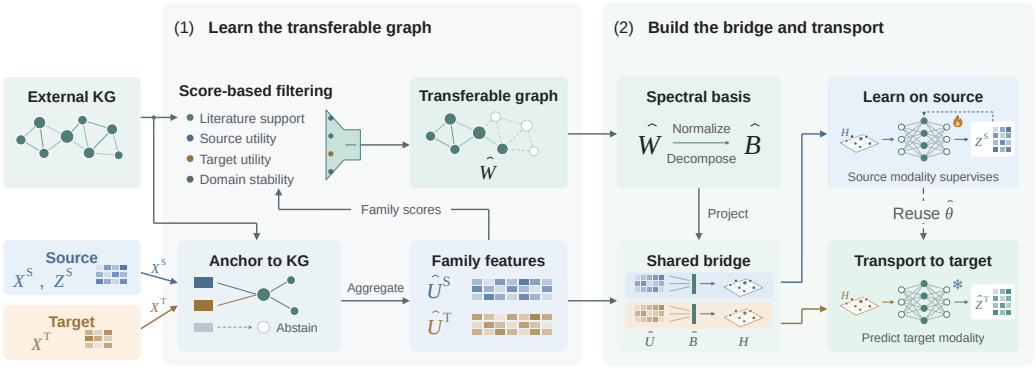

Figure 1: An overview of GraphBridge. Stage 1 builds a weighted literature prior, anchors raw variables, and filters out relations based on graph scores to build a transferable graph. Stage 2 forms the spectral basis of the transferable graph, learns the transformation in the source domain, and applies it in the target domain.

makes the learning pipeline able to handle uncertain variable-to-graph alignment, domain-specific graph relevance, and negative transfer.

Challenges. The above setting creates three challenges for transfer learning. First, the target domain has deployment-time modality absence. Specifically, Z^S is observed in the source but its target counterpart is unavailable, so the goal is not ordinary feature imputation under missing-at-random assumptions. Instead, the model should transfer only the component of the source-only signal that is predictable from information available in the target. Second, the domains may have heterogeneous raw features. Even when the two cohorts describe conceptually related objects, local measurement choices, coding systems, and preprocessing can yield different feature dimensions and non-overlapping coordinates, making raw-coordinate sharing or feature concatenation impossible. Third, the KG bridge is noisy. Variable names may map ambiguously to graph entities, some variables may have no meaningful graph counterpart, and edges may be incorrect, incomplete, or domain-specific.

3 Method: GraphBridge

GraphBridge is a pipeline for KG-mediated transfer under structural modality absence. It has two stages and one optional safeguard. Specifically, stage 1 turns the external KG into a source-target-specific transferable graph in four steps. It builds a weighted literature prior, anchors raw variables to node families with abstention, estimates source and target applicability views of the graph, and combines them into a source-target-specific transferable graph filtered out by graph-structural priors, domain-specific predictive relevance, and cross-domain stability. Stage 2 uses the transferable graph to construct a low-dimensional bridge representation shared by source and target observations, learns the component of the source-only modality that is predictable from this bridge, and transports it into the target predictor. Finally, an optional validation stage chooses between the transfer predictor and a target-only baseline. Algorithm 1 presents a concise algorithm, with the full step-by-step version provided in Section E. Fig. 1 shows an overview of our algorithm.

Stage 1: Learn a pair-specific transferable graph. Stage 1 turns the observed KG into a source–target-specific transferable graph, distinguishing the noisy literature graph from the graph the model actually uses. Specifically, GraphBridge first treats the KG as a weighted *literature prior* G^L, in which each relation carries a confidence rather than being taken as truth. It then *anchors* the raw variables of each domain $d \in \{S, T\}$ to the M node families, where alignment scores $R_{jm}^{(d)}$ rate how well variable j matches node family m, and each variable is assigned to its best node family, with an explicit abstention anchor so that a variable with weak or ambiguous evidence is left out of the graph instead of being forced in. The anchors aggregate raw variables into node-family features $\widehat{U}_i^{(d)}$,

Algorithm 1 GraphBridge: KG-mediated transfer with validation safeguard

Input: Source data $\mathcal{S}$, target data $\mathcal{T}$, external knowledge graph G, bridge dimension k

Stage 1: Learn the transferable graph

Build a weighted literature prior G^L from the KG

Anchor raw variables to node families with abstention $\rightarrow$ family features $\widehat{U}^{(d)}$

Estimate source and target applicability views of the graph

Filter the candidate relations by literature, source/target support, and cross-domain stability $\rightarrow$ transferable graph $\widehat{W}$

Stage 2: Bridge and transport

Form a spectral bridge basis $\widehat{B}$ of $\widehat{W}$ and embed both domains $\widehat{H}^{(d)} = \widehat{U}^{(d)}\widehat{B}$

Learn transport $\widehat{\theta}$ on the source

Apply $\widehat{\theta}$ to target embeddings $\rightarrow$ transported signal $\widehat{F}^{(T)}$

Optional: Validation safeguard

Fit transfer predictor $\widehat{f}_1$ on $(\widehat{H}^{(T)}, \widehat{F}^{(T)})$ and target-only baseline $\widehat{f}_0$

return $\widehat{f}_{\text{sel}}$, the better of $\widehat{f}_0$ and $\widehat{f}_1$ on a held-out target validation set

representing both domains in a shared family space. Next, because a literature relation need not hold in every cohort, GraphBridge estimates source and target applicability views that score how well each relation is supported within each domain. Finally, it filters the candidate relations by a score that combines the literature prior, the two applicability views, and a cross-domain discrepancy, retaining a relation only when it is supported by literature, useful in both domains, and stable across the two. The surviving relations define the transferable graph $\widehat{W}$ over the node families. This is the main difference between GraphBridge and standard KG feature generation. The external KG is treated as a candidate relational substrate, and the model learns which graph regions are useful and stable enough to support transfer for this particular domain pair. Section E gives the anchor scores, applicability views, and family gates in full, together with the hard-threshold reduction used in the analysis.

Stage 2: Bridge learning and source-only signal transport. The transferable graph $\widehat{W}$ defines a low-dimensional representation shared by both domains, from which GraphBridge forms a spectral bridge basis $\widehat{B}$ of $\widehat{W}$ and embeds each subject as $\widehat{H}_i^{(d)} = \widehat{U}_i^{(d)}\widehat{B}$. Because this bridge lives in node-family space rather than raw-feature space, it applies even when the source and target raw variables are disjoint. GraphBridge then learns, on the source data, a transport map $\widehat{\theta}$ from bridge embeddings to a factor $F^{(S)} = \phi(Z^{(S)})$ of the source-only modality, and applies it to the target embeddings to obtain a transported signal $\widehat{F}^{(T)}$. This signal is the component of the source-only modality that is predictable from the bridge, so it can be supplied to the target predictor at deployment time even though the target never observes the modality itself. The spectral construction and the transport objective are given in Section E.

Validation-protected model selection. Finally, GraphBridge uses a held-out target validation set to choose between the KG-mediated transfer predictor and a target-only predictor. This validation safeguard selects the most effective target predictor, allowing GraphBridge to exploit KG-mediated transfer when it is beneficial while avoiding negative transfer. We defer the precise selection rule and analysis to Section E.7.

4 Theoretical Analysis and Synthetic Validation

4.1 Theoretical Analysis

We analyze the hard-anchor and hard-gate versions of GraphBridge described in Section 3. The section presents two main results. Theorem 4.1 shows that the learned transfer substrate, including anchor assignments, relation filters, and spectral bridge basis, is consistently recovered when the KG contains a separable source-target transfer structure, with failure probability controlled by the anchor and family margins. Theorem 4.2 then bounds the end-to-end excess risk of the transfer predictor, decomposing it into three interpretable terms, including target supervision cost, source-side transport

estimation cost, and KG noise amplified by the spectral eigengap. Together, the two results justify the multi-stage GraphBridge pipeline by showing that each module contributes a controlled and additive error term. Full theorem statements and proofs are deferred to Sections G and H, respectively.

Notation and oracle objects. Let $R_T(f) = \mathbb{E}[\ell(Y^{(T)}, f(\cdot))]$ be target risk. The transfer predictor is $\widehat{f}_1$, the target-only baseline is $\widehat{f}_0$. We use disjoint target splits for graph filtering, target predictor training, validation, and final evaluation, with their sizes denoted by $n_{T,\text{gate}}$, $n_{T,\text{tr}}$, $n_{T,\text{val}}$, and $n_{T,\text{test}}$, respectively. All target-supervised variables used for graph filtering are estimated on the filtering split.

Let $A_d^\star$ and $\widehat{A}_d$ be the population and estimated anchor matrices in domain $d \in \{S, T\}$. Let $\mathcal{M}^\star \subseteq [M]$ be the population set of transferable graph's node families and $\widehat{\mathcal{M}}$ its learned estimate. The corresponding population transfer graph is $W^\star$, with top-k eigenspace $B^\star \in \mathbb{R}^{M\times k}$. GraphBridge estimates this graph by $\widehat{W}$ and its top-k eigenspace by $\widehat{B}$. The population and estimated bridge representations are

$$H_i^{(d),\star} = U_i^{(d)} B^\star, \qquad \widehat{H}_i^{(d)} = \widehat{U}_i^{(d)} \widehat{B}.$$

Because spectral embeddings are identifiable only up to rotation, bridge recovery is stated after an orthogonal alignment $Q \in \mathbb{R}^{k\times k}$.

For consistency with Section 3, we write $F_i^{(S)} = \varphi(Z_i^{(S)})$ for features extracted from the source-only modality. The target-side oracle transported feature is $F_i^{\star,(T)} = (H_i^{(T),\star})^\top \theta^\star$, and $f^\star$ denotes the oracle target predictor that uses the population bridge $H^{(T),\star}$ and the true transported feature $F^{\star,(T)}$.

Informal assumptions. The analysis uses standard recoverability conditions for the modules of GraphBridge. Anchor scores must separate correct node families from incorrect or null anchors by a margin Δ_A. Family filters must separate transferable and non-transferable graph families by a margin Δ_F, and their empirical support scores must concentrate. The population transfer graph $W^\star$ has eigengap $\text{gap} > 0$, and the learned graph perturbation satisfies $\|\widehat{W} - W^\star\|_{\text{op}} \le \eta_W$ on the family-selection event. Finally, the bridge design is well-conditioned, source and target noises are mean-zero sub-Gaussian, model parameters are bounded, and the validation split is independent of model fitting. Formal statements are given in Section F.

Theorem 4.1 (Learning a stable transfer substrate)**.** *Suppose Assumptions F.1 to F.3 hold. Let* $n_{\text{gate}} = \min\{n_S, n_{T,\text{gate}}\}$ *and* $c_F = \min\{c_s, c_t, c_d\}$. *Then the hard transfer filtering satisfies*

$$\mathbb{P}\Big(\widehat{\mathcal{M}} \neq \mathcal{M}^\star\Big) \le 6M \exp\big(-c_F n_{\text{gate}} \Delta_F^2\big)\,.$$

Moreover, define the anchor failure term

$$\varepsilon_A = 2 \sum_{d\in\{S,T\}} p_d(M+1) \exp\left(-\frac{\min\{\Delta_A^2, \Delta_0^2\}}{2\sigma_A^2}\right),$$

where Δ_0 *is the null-anchor margin and* σ_A *is the anchor-score noise scale. With probability at least*

$$1 - \varepsilon_A - 6M \exp\big(-c_F n_{\text{gate}} \Delta_F^2\big) - o(1),$$

the anchor and family-selection events hold. On this event, there exists an orthogonal matrix $Q \in \mathbb{R}^{k\times k}$ *such that*

$$\|\widehat{B} - B^\star Q\|_F \le \frac{2\sqrt{2k}}{\text{gap}} \eta_W,$$

and, for $d \in \{S, T\}$,

$$\frac{1}{n_d} \|\widehat{H}_d - H_d^\star Q\|_F^2 \le C_H \frac{k\eta_W^2}{\text{gap}^2},$$

where C_H *depends only on bounded second moments of the family-level features.*

Theorem 4.1 is the reliability statement for Stage 1 and the spectral part of Stage 2. The margins Δ_A and Δ_F quantify how ambiguous the anchor and filter decisions are; larger margins make the learned transfer substrate easier to recover. The ratio η_W/gap is the effective graph-noise level. Thus

GraphBridge can tolerate a noisy KG when the learned transfer graph is a small spectral perturbation of the population transfer graph. The exact selection statement is simplified. Section G also discusses the high-precision, moderate-recall regime, where the simplified gating can still yield a stable bridge by absorbing omitted families into the perturbation level η_W.

Theorem 4.2 (Transport and transfer-prediction error)**.** *Assume the conditions of Theorem 4.1 and Assumptions F.4 to F.6. Suppose the source-only features and target label follow the bridge model*

$$F_i^{(S)} = (H_i^{(S),\star})^\top \theta^\star + \epsilon_i^{(S)}, \qquad F_i^{\star,(T)} = (H_i^{(T),\star})^\top \theta^\star,$$

and

$$Y_i^{(T)} = (H_i^{(T),\star})^\top \beta^\star + \gamma^\star F_i^{\star,(T)} + \epsilon_i^{(T)},$$

where $\mathbb{E}[\varepsilon_i^{(d)} | H_i^{(d),\star}] = 0$*. Then, on the anchor and family-selection events, the transported feature satisfies*

$$\frac{1}{n_{T,\mathrm{tr}}} \sum_{i \in T_{\mathrm{tr}}} \left(\widehat{F}_i^{(T)} - F_i^{\star,(T)} \right)^2 \le C \left(\frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2} \right)$$

with probability at least $1 - o(1)$*. Consequently, the transfer predictor* $\widehat{f}_1$*, fitted on the target training split using* $(\widehat{H}_i^{(T)}, \widehat{F}_i^{(T)})$*, satisfies*

$$R_T(\widehat{f}_1) - R_T(f^\star) \le C \left(\frac{k+1}{n_{T,\mathrm{tr}}} + \frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2} \right)$$

with probability at least $1 - o(1)$*, conditional on the learned transfer substrate being correct.*

The three terms in Theorem 4.2 correspond to the three statistical bottlenecks of GraphBridge. The term $(k+1)/n_{T,\mathrm{tr}}$ is the supervised target estimation error for the final predictor. The term k/n_S is the source-side cost of learning the transport map from source bridge coordinates to the source-only feature. The term $k\eta_W^2/\mathrm{gap}^2$ is the cost of using a noisy KG, i.e., graph perturbation changes the spectral bridge, and the error is amplified when the population eigengap is small. When the KG is clean and the source sample is large relative to the bridge dimension, the dominant cost is target supervision, matching the sample-size behavior observed in Section 4.2.

4.2 Synthetic Validation

4.2.1 Settings

Data-generating process. A ground-truth graph $W^\star$ over M candidate node families has $M_\star$ transferable families and a spectral bridge basis $B^\star \in \mathbb{R}^{M \times k}$. Raw variables in each domain are assigned to families, yielding family-aggregated features and bridge representations. A source-only latent factor is generated as $Z_i = (H_i^{(S)})^\top \theta^\star + \varepsilon_i^{(S)}$. Target outcomes are $Y_i^{(T)} = (H_i^{(T)})^\top \beta^\star + \gamma^\star Z_i^{\star,(T)} + \varepsilon_i^{(T)}$. Full generation details and default parameters are in Section J.

Test scenarios. Five scenarios target specific failure modes, each designed to isolate a different challenge from Section 2. *Clean bridge* is the default setting with no corruption or distributional shift, serving as the reference condition. *Zero raw overlap* is designed where the source and target share no raw variable indices, so transfer is possible only through the graph-mediated bridge. *Applicability shift* tests the filtering stage's ability to filter domain-specific edges, by making a subset of transferable families reduce target-side support. *Noisy KG* has false-positive and false-negative edge rates be elevated to 10%, increasing η_W and complicating the spectral bridge recovery. *Adversarial transfer* forces the sign of the source-only contribution to the target outcome to be flipped, creating a regime where naïve transfer would hurt, testing the validation safeguard.

Baselines. We compare three estimators. *Raw*, a target-only model on $X^{(T)}$; *GraphBridge*, the full 2-staged algorithm; and *Oracle*, which receives the true target latent factor. Results are averaged over 10 seeds with standard deviations.

Table 1: AUROC (± std) of three methods in five scenarios. **Bold** marks the best non-oracle arm in each scenario. GraphBridge consistently outperforms the raw baseline with substantial improvement.

Scenario	Raw	GraphBridge	Oracle
Clean bridge	0.711 ± 0.087	$\mathbf{0.789 \pm 0.071}$	0.789 ± 0.076
Zero overlap	0.725 ± 0.078	$\mathbf{0.777 \pm 0.076}$	0.781 ± 0.076
Applicability shift	0.688 ± 0.088	$\mathbf{0.727 \pm 0.098}$	0.740 ± 0.101
Noisy KG	0.711 ± 0.085	$\mathbf{0.774 \pm 0.079}$	0.782 ± 0.080
Adversarial	0.679 ± 0.063	$\mathbf{0.737 \pm 0.027}$	0.745 ± 0.031

4.2.2 Results

GraphBridge drives the synthetic gain. Table 1 summarizes the results of three methods in five scenarios. The results show that GraphBridge provides a consistent advantage over the raw baseline across all five scenarios. GraphBridge achieves the best non-oracle performance and remains very close to the oracle upper bound. Even in the applicability-shift and adversarial scenarios, where Raw has an AUROC less than 0.7, GraphBridge substantially exceed the Raw baseline of 0.039 and 0.058, respectively. Overall, these results suggest that GraphBridge reliably closes much of the gap between the raw baseline and oracle performance while maintaining strong stability across challenging scenarios.

GraphBridge is robust to distributions. The results in Fig. 2 demonstrate the distributional robustness of GraphBridge. We test five feature distributions under the clean-bridge scenario. GraphBridge exceeds Raw in all five settings, and paired comparisons between each non-Gaussian distribution and the Gaussian baseline reveal no significant degradation. The pattern is consistent across heavy-tailed, skewed, and discrete distributions. These results indicate that GraphBridge depends on recoverable covariance and graph-structural information rather than on the Gaussian distribution itself, and suggest that the method is applicable to the mixed-type features commonly encountered in scientific domains without distributional preprocessing beyond standardization.

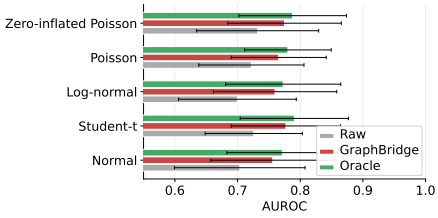

Figure 2: Distributional robustness of GraphBridge.Each row corresponds to a feature distribution. GraphBridge exceeds Raw under all five distributions.

5 Real Data Experiments

We evaluate GraphBridge on the PLATO benchmark (Ruiz et al., 2023), which most closely matches the structural modality missingness setting.

Datasets. The PLATO benchmark comprises five GDSC cancer cell-line cohorts: BC, ME, NSCLC (NS), SCLC (SC), and CH, and five PDX mouse model cohorts: BRCA (BR), CM, CRC, MNSCLC (MN), and PDAC (PD). BioKG (Walsh et al., 2020) is used as the external KG. Refer to Section L for more details.

Setup. Subgraphs are extracted from BioKG using Mehlhorn-style Steiner tree approximation. The task is drug response prediction evaluated with the Pearson correlation coefficient as in (Ruiz et al., 2023). We evaluate 20 ordered GDSC-to-GDSC and 20 PDX-to-PDX source-target pairs, respectively, using 5-fold outer CV with 3-fold inner GridSearchCV, averaged over 3 random seeds. Due to the space limit, we show the pairs of $\{\text{BC}, \text{CH}\} \mapsto \{\text{ME}, \text{NS}, \text{SC}\}$ for GDSC-to-GDSC cohorts and the pairs of $\{\text{BR}, \text{MN}\} \mapsto \{\text{CM}, \text{PD}, \text{CRC}\}$ for PDX-to-PDX cohorts. Refer to Section L for more results.

Models. We show two representative regression models: Ridge, and LightGBM (LGBM). Besides, we also compare two graph-based baselines, PLATO (Ruiz et al., 2023) and KGAT (Wang et al.,

Table 2: Pearson R for selected transfer pairs (source $\rightarrow$ target) from PLATO cell GDSC cohorts. **Bold** marks the best non-oracle arm per model and pair.

Model	Arm	BC→ME	BC→NS	BC→SC	CH→ME	CH→NS	CH→SC	$\bar{\Delta}_O$
Ridge	Raw	.160 ±.010	.213 ±.014	.204 ±.019	.160 ±.010	.214 ±.012	.219 ±.020	.694
	GraphBridge	**.398 ±.008**	**.498 ±.013**	**.399 ±.017**	**.483 ±.011**	**.451 ±.010**	**.398 ±.016**	.451
	Oracle	.900 ±.003	.876 ±.004	.893 ±.003	.899 ±.003	.874 ±.004	.893 ±.003	—
LGBM	Raw	.164 ±.007	.227 ±.004	.224 ±.023	.164 ±.007	.227 ±.004	.224 ±.023	.565
	GraphBridge	**.734 ±.011**	**.731 ±.011**	**.694 ±.012**	**.736 ±.014**	**.729 ±.036**	**.713 ±.015**	.048
	Oracle	.762 ±.011	.777 ±.012	.766 ±.015	.764 ±.011	.785 ±.031	.768 ±.016	—
PLATO	Raw	.152 ±.004	.217 ±.007	.205 ±.015	.151 ±.004	.211 ±.006	.211 ±.012	.695
	GraphBridge	**.829 ±.006**	**.719 ±.002**	**.843 ±.001**	**.822 ±.003**	**.785 ±.007**	**.783 ±.005**	.089
	Oracle	.895 ±.003	.874 ±.002	.893 ±.001	.895 ±.001	.871 ±.002	.889 ±.002	—
KGAT	Raw	**.897 ±.000**	.875 ±.002	**.894 ±.002**	.893 ±.001	.873 ±.002	**.890 ±.002**	.004
	GraphBridge	.894 ±.001	**.878 ±.003**	**.894 ±.003**	**.894 ±.001**	**.879 ±.003**	**.890 ±.002**	.003
	Oracle	.894 ±.001	.885 ±.004	.895 ±.001	.894 ±.001	.886 ±.003	.891 ±.002	—

Table 3: Pearson R for selected transfer pairs (source $\rightarrow$ target) from PLATO mouse PDX cohorts. **Bold** marks the best non-oracle arm per model and pair.

Model	Arm	BR→CM	BR→PD	BR→CRC	MN→CM	MN→PD	MN→CRC	$\bar{\Delta}_O$
Ridge	Raw	.361 ±.076	.355 ±.087	.515 ±.075	**.364 ±.076**	.356 ±.083	.517 ±.074	.088
	GraphBridge	**.444 ±.082**	**.391 ±.085**	**.548 ±.071**	.362 ±.077	**.421 ±.076**	**.569 ±.079**	.044
	Oracle	.506 ±.084	.426 ±.086	.568 ±.070	.506 ±.085	.418 ±.084	.574 ±.070	—
LGBM	Raw	.359 ±.074	**.364 ±.080**	**.510 ±.072**	.359 ±.074	**.364 ±.081**	.512 ±.071	.067
	GraphBridge	**.402 ±.075**	.326 ±.094	.503 ±.062	**.406 ±.087**	.346 ±.075	**.559 ±.066**	.054
	Oracle	.468 ±.065	.387 ±.083	.545 ±.069	.469 ±.066	.401 ±.076	.598 ±.060	—
PLATO	Raw	.255 ±.018	.248 ±.063	.460 ±.035	.248 ±.036	.246 ±.062	**.505 ±.035**	.118
	GraphBridge	**.336 ±.021**	**.389 ±.050**	**.520 ±.020**	**.249 ±.036**	**.330 ±.061**	.499 ±.142	.058
	Oracle	.418 ±.036	.386 ±.039	.549 ±.040	.425 ±.057	.395 ±.035	.495 ±.135	—
KGAT	Raw	.265 ±.022	.303 ±.045	.512 ±.034	**.261 ±.055**	.305 ±.035	**.517 ±.050**	.005
	GraphBridge	**.266 ±.024**	**.307 ±.046**	**.514 ±.034**	**.261 ±.055**	**.321 ±.046**	.511 ±.039	.002
	Oracle	.268 ±.024	.310 ±.048	.516 ±.036	.264 ±.059	.324 ±.048	.511 ±.039	—

2019). Results of Lasso, ElasticNet, and SVR are shown in Section L. Each model is evaluated with three methods: *Raw* (target observed features only), *GraphBridge*, and *Oracle* (true target missing modality).

Results. Table 2 reports Pearson R for selected GDSC cell-line transfer pairs. Raw cell-line features yield baseline correlations of $R = 0.227$ at maximum, whereas GraphBridge raises performance to $R = 0.824$ at maximum. Even the worst performance of $R = 0.398$ exceeds the best one of the baseline. LightGBM with GraphBridge achieves $R \geq 0.69$ on every pair shown, closing around 90% of the Raw-to-Oracle gap. The large gains reflect the fact that raw GDSC features alone are weakly predictive across cancer types, so the KG-mediated bridge provides the primary signal for useful transfer. For graph-based methods, results show that the PLATO algorithm itself suffers from poor performance without any modality transfer. With only the observed modality, it cannot extract enough information for prediction. On the other hand, KGAT utilizes the KG structure and attention mechanism, leading to a better result with Raw features. Nevertheless, with GraphBridge, there is still a minor improvement with the transferred modality. To further show the effectiveness of GraphBridge

Table 4: Normalized MSE (NMSE) for selected transfer pairs (source→target) from PLATO cell GDSC cohorts.

Metric	BC→ME	BC→NS	BC→SC	CH→ME	CH→NS	CH→SC
NMSE	.180 ±.004	.323 ±.003	.172 ±.003	.233 ±.002	.289 ±.005	.391 ±.001

Table 5: Effect of δ on model performance.

Method	$\delta = 0.00$	$\delta = 0.10$	$\delta = 0.30$	$\delta = 0.50$	$\delta = 0.70$
Raw	.164 ±.007	.164 ±.007	.164 ±.007	.164 ±.007	.164 ±.007
GraphBridge	.718 ±.020	.713 ±.023	.681 ±.024	.687 ±.012	.644 ±.035

to reconstruct the missing modality, we show the normalized MSE (NMSE) of the reconstruction in Table 4. Results show that most NMSEs are below 0.3 and the highest NMSE is less than 0.4, implying the missing modality is well transferred from the source to the target set.

To evaluate sensitivity to anchor quality, we conduct an experiment in which a fraction δ of the anchors is corrupted through random reassignment, with $\delta \in \{0, 0.1, 0.3, 0.5, 0.7\}$. Experiments are conducted on the BC→ME pair, with LGBM as the predictor. Results in Table 5 show that as δ increases, more perturbation is introduced in the anchoring and the performance is worse. However, even under $\delta = 0.7$, GraphBridge still outperforms Raw significantly, emphasizing the importance of different modalities even with bad structure information.

Table 3 reports Pearson R for selected PDX transfer pairs. First, GraphBridge consistently improves over raw features on most Ridge PDX pairs. The improvement is most significant for the BR→CM pair from 0.361 to 0.444 and the MN→PD pair from 0.356 to 0.421. LGBM shows mixed behavior, where GraphBridge helps on some pairs but underperforms Raw on others, suggesting that nonlinear target-only models can already exploit part of the available signal. Results of graph-based methods are consistent with Table 2, where GraphBridge outperforms with PLATO while the improvement is marginal with KGAT.

Discussion. The PLATO results reveal two transfer regimes corresponding to the information content of target-only features. In the GDSC cell-line setting, raw target-only features carry little cross-cancer predictive signal, and GraphBridge provides large gains. In the PDX mouse-model setting, raw features already achieve moderate correlations, and improvements are smaller but consistent for most models. The larger residual Oracle gap indicates additional transportable signal beyond what the current linear bridge extracts. This contrast aligns with the excess-risk bound (Theorem 4.2), that the bridge term is most beneficial when the target-only risk is large relative to the estimation error from graph noise and limited samples. From a practical standpoint, even modest PDX gains are relevant in pharmacogenomics where small correlation improvements affect drug ranking decisions, while the GDSC results demonstrate that GraphBridge can enable effective prediction in cross-cancer settings where raw features alone are nearly uninformative.

6 Conclusion

We introduced GraphBridge, a framework that treats a noisy external knowledge graph as a learnable transfer substrate for settings where the target domain lacks an entire source modality and raw feature spaces may be disjoint. End-to-end guarantees ensure that the method with a validation selector never performs worse than the target-only baseline up to a standard held-out validation penalty. Meanwhile, synthetic and real-data experiments confirm consistent gains, particularly when target-only features are least informative. The results suggest that external knowledge graphs can bridge domains with no shared raw features, provided that alignment, graph selection, and model choice are all learned with appropriate safeguards.

AI DISCLOSURE

In this work, we used generative AI tools to polish, find related work, and check mathematical proofs. We have not used generative AI tools for other tasks with required disclosure, and the rest of the required disclosure tasks are not applicable to this work. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. Spice: Semantic propositional image caption evaluation. In *European conference on computer vision*, pages 382–398. Springer, 2016.

Runxue Bao, Yiming Sun, Yuhe Gao, Jindong Wang, Qiang Yang, Zhi-Hong Mao, and Ye Ye. A recent survey of heterogeneous transfer learning. *arXiv preprint arXiv:2310.08459*, 2023.

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):151–175, 2010.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific data*, 10(1):67, 2023.

Yingshan Chang, Mridu Narang, Hisami Suzuki, Guihong Cao, Jianfeng Gao, and Yonatan Bisk. Webqa: Multihop and multimodal qa. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16495–16504, 2022.

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

Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. Clap learning audio concepts from natural language supervision. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. The webnlg challenge: Generating text from rdf data. In *Proceedings of the 10th international conference on natural language generation*, pages 124–133, 2017.

Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. Imagebind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15180–15190, 2023.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 7514–7528, 2021.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. volume 30, 2017.

Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning*, pages 4904–4916. PMLR, 2021.

Justin Johnson, Agrim Gupta, and Li Fei-Fei. Image generation from scene graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1219–1228, 2018.

K Kilgour et al. Frechet audio distance: A reference-free metric for evaluating music enhancem. InterSpeech, 2019.

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. Text generation from knowledge graphs with graph transformers. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2284–2293, 2019.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision*, 123(1):32–73, 2017.

Lien P Le, Thu Nguyen, Michael A Riegler, Pål Halvorsen, and Binh T Nguyen. Multimodal missing data in healthcare: A comprehensive review and future directions. *Computer Science Review*, 56: 100720, 2025.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023.

Di Lu, Spencer Whitehead, Lifu Huang, Heng Ji, and Shih-Fu Chang. Entity-aware image caption generation. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 4013–4023, 2018.

Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to explain: Multimodal reasoning via thought chains for science question answering. volume 35, pages 2507–2521, 2022.

Natalia Neverova, Christian Wolf, Graham Taylor, and Florian Nebout. Moddrop: adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (8):1692–1706, 2015.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021.

Joseph D Romano, Van Truong, Rachit Kumar, Mythreye Venkatesan, Britney E Graham, Yun Hao, Nick Matsumoto, Xi Li, Zhiping Wang, Marylyn D Ritchie, et al. The alzheimer's knowledge base: a knowledge graph for alzheimer disease research. *Journal of Medical Internet Research*, 26: e46777, 2024.

Camilo Ruiz, Hongyu Ren, Kexin Huang, and Jure Leskovec. High dimensional, tabular deep learning with an auxiliary knowledge graph. volume 36, pages 26348–26371, 2023.

Chang Su, Yu Hou, Manqi Zhou, Suraj Rajendran, Jacqueline RMA Maasch, Zehra Abedi, Haotan Zhang, Zilong Bai, Anthony Cuturrufo, Winston Guo, et al. Biomedical discovery through the integrative biomedical knowledge hub (ibkh). *Iscience*, 26(4), 2023.

Santiago Timón-Reina, Mariano Rincón, Rafael Martínez-Tomás, Bjørn-Eivind Kirsebom, and Tormod Fladby. A knowledge graph framework for dementia research data. *Applied Sciences*, 13 (18):10497, 2023.

Ramakrishna Vedantam, C Lawrence Zitnick, and Devi Parikh. Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4566–4575, 2015.

Brian Walsh, Sameh K Mohamed, and Vít Nováček. Biokg: A knowledge graph for relational learning on biological data. pages 3173–3180, 2020.

Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019.

Renjie Wu, Hu Wang, Hsiang-Ting Chen, and Gustavo Carneiro. Deep multimodal learning with missing modality: A survey. *arXiv preprint arXiv:2409.07825*, 2024.

Hu Xu, Gargi Ghosh, Po-Yao Huang, Dmytro Okhonko, Armen Aghajanyan, Florian Metze, Luke Zettlemoyer, and Christoph Feichtenhofer. Videoclip: Contrastive pre-training for zero-shot video-text understanding. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 6787–6800, 2021.

Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Chen. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. pages 9556–9567, 2023. URL https://api.semanticscholar.org/CorpusID:265466525.

Wentian Zhao and Xinxiao Wu. Boosting entity-aware image captioning with multi-modal knowledge graph. *IEEE Transactions on Multimedia*, 26:2659–2670, 2023.

APPENDIX CONTENTS

A Knowledge Graphs in Biomedicine **14**

B Related Work **14**

C Scope and Limitations **15**

D Notation and Population Setup **16**

D.1 Objects and Symbols 16

D.2 Source and Target Data 16

D.3 The External Graph and Its Population Views 17

D.4 Anchor-Induced Family Features 17

D.5 Transfer Families and the Population Transfer Graph 17

D.6 Bridge Basis and Latent Representations 17

D.7 Transported Source-Only Factor and Target Model 17

E Step-by-Step Algorithm Definition **18**

E.1 Step A: Weighted Literature Prior 18

E.2 Step B: Soft Anchoring with Abstention 18

E.3 Step C: Source and Target Applicability Views 19

E.4 Step D: Pair-Specific Transfer Graph 19

E.5 Step E: Spectral Bridge Encoding 20

E.6 Step F: Source-Only Modality Transport 20

E.7 Step G: Target Prediction and Validation Safeguard 20

F Assumptions **21**

G Finite-Sample Results **22**

G.1 Anchor Consistency 22

G.2 Transfer-Family Selection Consistency 23

G.3 Bridge Basis Recovery 23

G.4 Transport Error 23

G.5 Target Excess Risk 23

G.6 Validation Safeguard Against Negative Transfer 24

H Proofs of Main Results **24**

H.1 Proof of Theorem G.1 24

H.2 Proof of Theorem G.2 25

H.3 Proof of Theorem G.3 25

H.4 Auxiliary Lemma for Least-Squares Perturbation Bounds 26

H.5 Proof of Theorem G.4 26
H.6 Proof of Theorem G.5 26
H.7 Proof of Theorem G.6 27
H.8 Proof of Corollary G.1 27

I Remarks **28**

J Synthetic Data-Generating Process **28**
J.1 Population Graph 29
J.2 Bridge Basis 29
J.3 Feature Generation 29
J.4 Source-Only Modality 29
J.5 Outcome Model 29
J.6 External Graph Corruption 29
J.7 Default Parameters 29
J.8 Scenario Perturbations 29

K Synthetic Experiments **29**

L Details of Real Data Experiments **30**

A Knowledge Graphs in Biomedicine

A rapidly growing literature builds domain-specific knowledge graphs to organize evidence, support hypothesis generation, identify biomarkers, and prioritize therapeutics. In biomedicine, disease-specific resources such as AlzKB (Romano et al., 2024) and DemKG (Timón-Reina et al., 2023) organize dementia-related knowledge, while broader graphs such as PrimeKG (Chandak et al., 2023) and iBKH (Su et al., 2023) integrate multiple biomedical sources into large-scale knowledge networks. BioKG (Walsh et al., 2020) provides a multi-relational graph spanning genes, diseases, drugs, pathways, and molecular functions. These resources demonstrate the value of graph-structured priors and provide essential infrastructure for disease modeling.

Most such work treats the graph primarily as a curated knowledge base or as an input to downstream analytics. Less attention has been given to estimating a source-target-specific transferable graph for cross-cohort or cross-domain learning under structured modality absence—the focus of the present paper.

B Related Work

Multimodal representation learning and cross-modal alignment. Modern multimodal learning has largely advanced through shared latent spaces learned from paired data. CLIP and ALIGN showed that large-scale contrastive image–text pretraining yields strong zero-shot transfer and retrieval performance (Radford et al., 2021; Jia et al., 2021). BLIP-2 further demonstrated an efficient bridge from visual encoders to language models for generative vision–language transfer (Li et al., 2023). Beyond images, VideoCLIP and CLAP extended contrastive pretraining to video–text and audio–text alignment, respectively (Xu et al., 2021; Elizalde et al., 2023), while ImageBind generalized joint alignment to a larger family of modalities and even exhibited emergent cross-modal generation behaviors (Girdhar et al., 2023). However, these methods usually keep knowledge implicit in datasets or model parameters, rather than representing it explicitly as a relational graph.

Graph-guided modality transfer and generation. A closer body of work uses graphs as explicit control signals for mapping one modality to another. In vision, scene-graph-to-image generation conditions image synthesis on object–relation structure (Johnson et al., 2018). In language generation, RDF/KG-to-text benchmarks and models such as WebNLG and GraphWriter translate graph-structured inputs into fluent natural language (Gardent et al., 2017; Koncel-Kedziorski et al., 2019). Knowledge-augmented captioning further injects symbolic structure into image-to-text generation: entity-aware captioning uses KG-based collective inference to fill specific entities (Lu et al., 2018), and later work explicitly constructs a multimodal KG to connect visual objects, named entities, and textual relations for more informative captions (Zhao and Wu, 2023). These works show that graph structure can regularize or drive cross-modal generation, but they typically treat the graph as an input controller or a local grounding aid, rather than as a transport medium for transferring a missing modality across domains.

Missing modalities, benchmarks, and evaluation. Orthogonal work studies robustness to incomplete observations. Early methods such as ModDrop improved multimodal robustness under missing or noisy inputs (Neverova et al., 2015), and recent surveys systematize a broad family of missing-modality learning methods (Wu et al., 2024). Yet most of this literature focuses on fusion, imputation, or robustness within a single multimodal task, rather than using a KG to bridge a structurally absent modality across domains. Evaluation in the broader area spans structured benchmarks such as Visual Genome for scene graphs (Krishna et al., 2017), ScienceQA and WebQA for multimodal reasoning (Lu et al., 2022; Chang et al., 2022), and MMMU for expert-level multimodal understanding and knowledge-rich reasoning (Yue et al., 2023). For generation-oriented tasks, common metrics include CIDEr, SPICE, CLIPScore, FID, and FAD (Vedantam et al., 2015; Anderson et al., 2016; Hessel et al., 2021; Heusel et al., 2017; Kilgour et al., 2019). Taken together, prior work has established strong components for graph representation, cross-modal alignment, graph-guided generation, and knowledge-enhanced reasoning, but there remains comparatively less work on using a KG itself as the explicit medium for modality transport under structural modality absence.

C Scope and Limitations

GraphBridge helps only when the source-only signal has a component that is inferable through a stable source-target semantic bridge. If variables cannot be reliably anchored to graph families, if the external graph is badly misleading, or if source and target transferable subspaces are nearly disjoint (Remark 6), transport cannot be expected to help. The validation safeguard protects against choosing a harmful transfer predictor, but it does not create a transferable signal where none exists.

The theoretical analysis is intentionally simplified, as it uses hard anchors, hard gates, a spectral bridge, linear transport, and standard concentration and spectral-gap assumptions. These assumptions yield a tractable theoretical analysis that isolates how anchoring, gating, graph perturbation, and validation-based model selection contribute to reliable transfer. At the same time, they leave several important extensions beyond the current analysis, including nonlinear representation learners, temporal graphs, multi-source transfer settings, and inference after data-dependent model selection.

Any deployment in biomedical settings should be treated as decision support only: graph bias, population mismatch, and negative transfer can produce misleading predictions, and clinical use would require external validation and appropriate governance. Natural extensions include soft anchors and gates, generalized outcome models, cross-fitting for post-selection stability, multi-source transfer, and uncertainty-aware bridge estimation.

Scope of the theory. We analyze a simplified but end-to-end version of the pipeline, using hard anchors, hard gates, a spectral bridge, and linear transport. This abstraction preserves the core transfer mechanism while allowing errors to be propagated through all stages. The theory establishes stability of the learned bridge, reliable transport of source-only signals, and validation-protected selection against negative transfer. The broader predictive benefits of transported factors, especially with limited target data, regularization, and nonlinear predictors, are evaluated empirically in **??** and Section 5.

D NOTATION AND POPULATION SETUP

This section fixes the notation used throughout the appendix. We separate three layers of objects. The first layer is the observed data and the external knowledge graph. The second layer is the intermediate quantities produced by the implemented pipeline of Section E. The third layer is the population and oracle quantities used in the theoretical analysis, marked with a star $\star$. The implemented pipeline uses soft anchors, data-driven applicability views, and soft gates, while the analysis in Section G studies a hard-threshold reduction of the same pipeline that keeps support recovery, spectral perturbation, and regression arguments transparent.

D.1 OBJECTS AND SYMBOLS

Symbol	Meaning
$d \in \{S, T\}$	Domain indicator for source S or target T.
n_d	Sample size in domain d.
p_d	Number of observed raw variables in domain d. No equality or overlap is assumed between p_S and p_T.
M	Number of candidate node families, also called concept clusters, in the external graph.
$X_i^{(d)} \in \mathbb{R}^{p_d}$	Raw observed features for subject i in domain d.
Z_i	Source-only modality for source subject i, with $F_i^{(S)} = \varphi(Z_i)$ a fixed-dimensional factor extracted from it.
$Y_i^{(T)}$	Target response for the downstream task.
$G = (V, E, R, \kappa)$	Observed typed knowledge graph with nodes V, edges E, relation types R, and entity-type map κ.
$G^L \in \mathbb{R}^{M \times M}$	Literature view of the graph after family aggregation, with prior weights q^L.
$q_e^L \in [0, 1]$	Literature prior confidence on edge e, from Step A.
$\pi_{j\cdot}^{(d)}$	Soft anchor distribution of raw variable j over the M node families and the null state, from Step B.
$a_j^{(d),\star} \in \{0, 1, \ldots, M\}$	True anchor of raw variable j in domain d, where 0 is the null state with no compatible family.
$A_d^\star$	True one-hot anchor matrix in $\{0, 1\}^{p_d \times (M+1)}$.
$U_i^{(d)} \in \mathbb{R}^M$	Family-aggregated graph features, the node proxies, after applying the anchors.
$c_{e,S}$, $c_{e,T}$	Data-driven source and target support for edge e, from Step C.
$G_S^\star(P_S)$, $G_T^\star(P_T)$	Source and target applicability views of the graph, from Step C.
$d_{e,ST}$	Source–target discrepancy of edge e, from Step D.
ω_e^{tr}	Pair-specific transfer weight on edge e, from Step D.
$\mathcal{M}^\star \subseteq [M]$	Set of transferable node families.
$g_m^\star \in \{0, 1\}$	Population transfer gate for node family m, equal to 1 exactly when $m \in \mathcal{M}^\star$.
$G_{S \leftrightarrow T}^\star(P_S, P_T)$	Population transferable graph, the subgraph that is safe for transfer.
$W^\star \in \mathbb{R}^{M \times M}$	Population transfer graph matrix on the transferable families.
$B^\star \in \mathbb{R}^{M \times k}$	Top-k orthonormal eigenvectors of $W^\star$, the bridge basis.
$H_i^{(d),\star} = U_i^{(d)} B^\star \in \mathbb{R}^k$	Latent bridge representation of subject i in domain d.
$\theta^\star \in \mathbb{R}^k$	Transport parameter from the bridge to the source-only factor.

D.2 SOURCE AND TARGET DATA

We observe

$$\mathcal{S} = \{(X_i^{(S)}, Z_i, Y_i^{(S)}) : 1 \le i \le n_S\}, \qquad \mathcal{T} = \{(X_i^{(T)}, Y_i^{(T)}) : 1 \le i \le n_T\}.$$

The source records the source-only modality Z, while the target does not. The raw features do not have to overlap across domains, and the bridge is induced through the graph families and the anchor maps. In Section 4 we write $F_i^{(S)} = \varphi(Z_i)$ for the fixed-dimensional factor extracted from the source-only modality, while here we use Z for brevity, and the displayed rates are unchanged for any fixed-dimensional factor up to constants.

D.3 The External Graph and Its Population Views

The observed graph G is a noisy *literature view* that may contain false edges, omit true relations, or include relations relevant in one population but not another. To make the inferential target precise we distinguish four graph objects, all defined at the family level after aggregation. Let $G^{\mathrm{bio},\star}$ be a latent broad-population relation graph. Let $G_S^\star(P_S)$ and $G_T^\star(P_T)$ be its *applicability views*, the subgraphs whose relations remain supported under the source and target distributions, respectively. Finally, let

$$G_{S\leftrightarrow T}^\star(P_S, P_T)$$

be the *transferable graph*, the subgraph that is both structurally plausible and jointly useful for transfer between the specific source and target populations. The literature view G^L, with prior weights $q^L \in [0,1]$, is the only graph object the algorithm observes, and Steps A–D estimate the remaining objects from data, culminating in an estimate of $G_{S\leftrightarrow T}^\star$.

D.4 Anchor-Induced Family Features

The implemented pipeline produces a soft anchor distribution $\pi_{j\cdot}^{(d)}$ over the M node families and a null abstention state in Step B. For the analysis we use the corresponding hard anchor. For domain d, the true anchor matrix $A_d^\star \in \{0,1\}^{p_d\times(M+1)}$ has exactly one nonzero entry per row. If raw variable j maps to family $m \geq 1$, then $a_j^{(d),\star} = m$, and if it is incompatible with the graph, then $a_j^{(d),\star} = 0$.

Let $\bar{A}_d^\star \in \{0,1\}^{p_d\times M}$ denote the anchor matrix with the null column removed. The family-aggregated features are

$$U_i^{(d)} = X_i^{(d)} \bar{A}_d^\star \in \mathbb{R}^M. \tag{1}$$

While this definition is simple, in that each family feature is just the standardized sum of all raw variables assigned to that family, weighted aggregation can be handled by replacing $\bar{A}_d^\star$ with a real-valued loading matrix such as the soft anchors $\pi^{(d)}$.

D.5 Transfer Families and the Population Transfer Graph

Let $\mathcal{M}^\star \subseteq [M]$ be the set of node families that are both structurally meaningful and jointly useful for transfer between the source and target populations, and that therefore index the transferable graph $G_{S\leftrightarrow T}^\star(P_S, P_T)$. Define the population gate vector $g^\star \in \{0,1\}^M$ by $g_m^\star = \mathbb{I}\{m \in \mathcal{M}^\star\}$. Let $\widetilde{G}^\star \in \mathbb{R}^{M\times M}$ be an ideal family-level graph on all candidate families. The population transfer graph is

$$W^\star = \mathrm{diag}(g^\star)\, \widetilde{G}^\star\, \mathrm{diag}(g^\star), \tag{2}$$

the family graph restricted to the transferable families $\mathcal{M}^\star$.

D.6 Bridge Basis and Latent Representations

Let $B^\star \in \mathbb{R}^{M\times k}$ collect the top k eigenvectors of $W^\star$, with $(B^\star)^\top B^\star = I_k$. The latent bridge representation for subject i in domain d is

$$H_i^{(d),\star} = U_i^{(d)} B^\star \in \mathbb{R}^k. \tag{3}$$

Zero raw overlap is not a problem as long as both domains project onto this shared bridge subspace.

D.7 Transported Source-Only Factor and Target Model

We model the source-only factor by

$$Z_i = (H_i^{(S),\star})^\top \theta^\star + \varepsilon_i^{(S)}, \qquad \mathbb{E}[\varepsilon_i^{(S)} \mid H_i^{(S),\star}] = 0, \tag{4}$$

where $\theta^\star \in \mathbb{R}^k$ is the transport parameter.

On the target side, define the latent transported signal

$$Z_i^{\star,(T)} := (H_i^{(T),\star})^\top \theta^\star. \tag{5}$$

The downstream target label is assumed to satisfy

$$Y_i^{(T)} = (H_i^{(T),\star})^\top \beta^\star + \gamma^\star Z_i^{\star,(T)} + \varepsilon_i^{(T)}, \qquad \mathbb{E}[\varepsilon_i^{(T)} \mid H_i^{(T),\star}] = 0, \tag{6}$$

for unknown $\beta^\star \in \mathbb{R}^k$ and $\gamma^\star \in \mathbb{R}$. This is the analysis model. The target does not literally observe Z, yet it carries a latent source-only component that is inferable only through the bridge.

E Step-by-Step Algorithm Definition

We now describe the algorithm in the form in which it is implemented and analyzed. The implemented pipeline runs seven inspectable steps, labeled A through G, each emitting an interpretable intermediate object so that the procedure is modular and debuggable, as summarized in Algorithm 2. Steps A–D realize Stage 1 of the main text and turn the observed literature graph into a pair-specific transferable graph through a weighted literature prior in Step A, soft semantic anchoring with abstention in Step B, data-driven source and target applicability views in Step C, and pair-specific transfer-graph construction in Step D. Steps E and F realize Stage 2 through a spectral bridge basis in Step E and source-only modality transport in Step F. Step G fits the validation-protected target predictor. We describe each step as implemented, with soft anchors, data-driven views, soft gates, a normalized spectral bridge, and regularized transport and prediction. Where the theory simplifies a step, we flag the *analyzed reduction* studied in Section G, namely hard anchors and gates, a plain spectral bridge, and linear least-squares transport and prediction. We use a medical KG as a running example, and we collect the default settings of all hyperparameters in Section L.

E.1 Step A: Weighted Literature Prior

The only graph object the algorithm observes is the literature view. For each candidate edge $e = (u, v)$ between node families u and v, the literature assigns a prior confidence

$$q_e^L = \mathrm{clip}_{[0,1]}\big(c(e) \cdot \mathbb{I}\{e \text{ present}\}\big), \tag{7}$$

where $c(e)$ is the per-edge literature confidence and $\mathbb{I}\{e \text{ present}\}$ indicates that the edge is reported. Assembling these weights into a symmetric matrix yields the literature-view graph $G^L \in \mathbb{R}^{M\times M}$ with $G^L_{uv} = q_e^L$. All later steps reweight and prune G^L rather than treating it as truth. The M *node families*, also called concept clusters, are obtained by clustering KG entity embeddings, so G^L is the node-family-aggregated literature graph, and its edges may optionally be grouped into *relation families* for the gating in Step D.

E.2 Step B: Soft Anchoring with Abstention

For each raw variable j in domain d, let $R^{(d)}_{j,m}$ be a score for anchoring variable j to node family $m \in [M]$, and let $R^{(d)}_{j,0}$ be a score for the null abstention state. The score combines a *prior* term, proportional to the candidate logit $\ell_{j,m}$ that encodes how confidently variable j maps to family m, with a *neighborhood-compatibility* term that rewards mapping j to a family whose graph neighbors $N(m)$ in G^L are empirically associated with the other variables of the cohort,

$$R^{(d)}_{j,m} = \underbrace{\lambda_{\mathrm{pr}}\, \ell_{j,m}}_{\text{prior}} + \underbrace{\lambda_{\mathrm{nbr}} \sum_k C^{(d)}_{jk}\, \mathrm{mem}_{k,N(m)}/|N(m)|}_{\text{neighborhood compatibility}},$$

where $C^{(d)} = |\mathrm{corr}(X^{(d)})|$, the term $\mathrm{mem}_{k,N(m)}$ measures how strongly variable k loads onto the neighbors of m, and $\lambda_{\mathrm{pr}}, \lambda_{\mathrm{nbr}}$ are weights. A temperature-scaled softmax over the M node families and the null state yields the soft anchor distribution

$$\pi^{(d)}_{j\cdot} = \mathrm{softmax}\big(R^{(d)}_{j\cdot}/\varsigma\big) \in \Delta^M, \tag{8}$$

with temperature ς and an explicit null bias. The null state implements abstention, so variables with weak or ambiguous KG evidence are excluded from graph-based transfer rather than becoming noisy semantic features.

Analyzed reduction, hard anchor. For the theory we replace (8) by its hard counterpart. Given threshold τ_A,

$$\hat{a}_j^{(d)} = \begin{cases} \arg\max_{1\le m\le M} R_{j,m}^{(d)}, & \text{if } \max_{1\le m\le M} R_{j,m}^{(d)} \ge \tau_A, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Let $\hat{A}_d$ and $\hat{\tilde{A}}_d$ be the corresponding one-hot matrices, where $\hat{\tilde{A}}_d$ drops the null column, and define the family-aggregated features, also called node proxies,

$$\hat{U}_i^{(d)} = X_i^{(d)} \hat{\tilde{A}}_d. \tag{10}$$

The implemented pipeline uses the same construction with $\hat{\tilde{A}}_d$ replaced by the soft anchors $\pi^{(d)}$ after dropping the null column and renormalizing the rows.

E.3 Step C: Source and Target Applicability Views

Step C estimates which edges of G^L are supported *within each domain* and produces data-driven estimates of the applicability views $G_S^\star(P_S)$ and $G_T^\star(P_T)$. We run it independently for each domain $d \in \{S, T\}$. The family features $\hat{U}^{(d)}$ of (10) serve as node proxies, and for each candidate edge $e = (u, v)$ the data-driven support is the standardized absolute correlation of its endpoint proxies,

$$c_{e,d} = z\Big(\big|\mathrm{corr}_d\big(\hat{U}_{\cdot u}^{(d)}, \hat{U}_{\cdot v}^{(d)}\big)\big|\Big), \tag{11}$$

where $z(\cdot)$ denotes z-scoring across edges. The calibrated per-domain view weight combines the literature prior and the data-driven support through an increasing map,

$$\psi_d(c_{e,d}) = \mathrm{sigmoid}\big(\beta_q \,\mathrm{logit}(q_e^L + \epsilon) + \beta_c \, c_{e,d}\big). \tag{12}$$

The reweighted graphs $\{\psi_S(c_{e,S})\}_e$ and $\{\psi_T(c_{e,T})\}_e$ are the estimated source and target applicability views.

Analyzed reduction, family-level scores. For the analysis we replace the data-driven edge support and view weights by scalar scores attached to each node family m, which play the role of $c_{e,S}, c_{e,T}$ and the cross-domain shift aggregated to node families. We define the population source utility, target utility, and domain discrepancy by

$$s_m^\star = \Big|\mathrm{Cov}(U_{1m}^{(S)}, Z_1)\Big|, \qquad t_m^\star = \Big|\mathrm{Cov}(U_{1m}^{(T)}, Y_1^{(T)})\Big|, \qquad d_m^\star = \Big|\mathbb{E}[U_{1m}^{(S)}] - \mathbb{E}[U_{1m}^{(T)}]\Big|. \tag{13}$$

We let $\ell_m^\star \in [0, 1]$ denote the population literature-prior score for node family m, the node-family-level summary of q^L, and we let $\hat{s}_m, \hat{t}_m, \hat{d}_m$ be the corresponding estimators.

E.4 Step D: Pair-Specific Transfer Graph

Step D merges the source and target views into one pair-specific transferable graph, an estimate of $G_{S\leftrightarrow T}^\star(P_S, P_T)$. The per-edge source–target discrepancy is

$$d_{e,ST} = |c_{e,S} - c_{e,T}|, \tag{14}$$

and, by default, GraphBridge groups the candidate edges into *relation families*, each a set of edges sharing a relation type or weight-quantile bucket. This grouping is optional, and letting each edge be its own family recovers a purely edge-level gate. For a relation family, let $\bar{c}_S, \bar{c}_T, \bar{q}^L$, and $\bar{d}_{ST}$ be the within-family averages of the source support, target support, literature prior, and discrepancy, which reduce to the edge's own values without grouping. The family receives a soft gate

$$g = \mathrm{sigmoid}\big(\alpha_c(\bar{c}_S + \bar{c}_T) + \alpha_q \, \bar{q}^L - \alpha_d \, \bar{d}_{ST} - b\big), \tag{15}$$

shared by all of its edges, where $\alpha_c, \alpha_q, \alpha_d$ and b are gate coefficients. The gate suppresses groups of relations whose support pattern differs across domains. Writing $g(e)$ for the gate applied to edge e, the resulting transfer weight is the product

$$\omega_e^{\mathrm{tr}} = g(e) \; \psi_S(c_{e,S}) \; \psi_T(c_{e,T}) \; \rho(d_{e,ST}), \qquad \rho(d) = e^{-\nu d}, \tag{16}$$

with decay $\nu > 0$. An edge is kept only when it is structurally supported through large ψ_S, ψ_T, literature- and family-consistent through a large gate $g(e)$, and stable across domains through a small $d_{e,ST}$. Edges with $\omega_e^{\mathrm{tr}} \geq \tau_\omega$ are selected and assembled into the symmetric weighted transferable graph $\hat{W}$. The product form (16) is the operational realization of the four cleaning principles of literature support, source compatibility, target compatibility, and cross-domain stability, and it matches the canonical transfer weight $\omega_e^{\mathrm{tr}} = q_e^L \, \psi_S(c_{e,S}) \, \psi_T(c_{e,T}) \, \rho(d_{e,ST})$ of the main text, with the family gate $g(e)$ generalizing the bare literature factor q_e^L by pooling the prior with within-family support and discrepancy.

Analyzed reduction, hard family gate. For the theory we replace the soft gate (15) by a hard gate over *node families*, driven by the node-family scores (13), so that selection acts directly on the node-family graph $\hat{W}$. Given estimators $\hat{s}_m, \hat{t}_m, \hat{d}_m$,

$$\hat{g}_m = \mathbb{I}\{\ell_m^\star \geq \tau_\ell, \ \hat{s}_m \geq \tau_s, \ \hat{t}_m \geq \tau_t, \ \hat{d}_m \leq \tau_d\}, \tag{17}$$

and $\hat{\mathcal{M}} = \{m : \hat{g}_m = 1\}$. With $G^L \in \mathbb{R}^{M \times M}$ the literature graph after family aggregation, the estimated transfer graph is

$$\hat{W} = \mathrm{diag}(\hat{g}) \, G^L \, \mathrm{diag}(\hat{g}). \tag{18}$$

E.5 Step E: Spectral Bridge Encoding

Given the transferable graph $\hat{W}$, Step E builds a low-dimensional bridge basis. We symmetrize $\hat{W}$ and zero its diagonal, form the degree vector $D = \mathrm{rowsum}(\hat{W})$ with zero entries reset to one, and degree-normalize with a small ridge $c_0 > 0$,

$$\hat{\mathcal{S}} = D^{-1/2} \, (\hat{W} + c_0 I) \, D^{-1/2}. \tag{19}$$

Let $\hat{B} \in \mathbb{R}^{M \times k}$ collect the top k eigenvectors of $\hat{\mathcal{S}}$, scaled by the square roots of their eigenvalues, where k is the bridge dimension. The estimated bridge representations are

$$\hat{H}_i^{(d)} = \hat{U}_i^{(d)} \hat{B}. \tag{20}$$

Analyzed reduction. The analysis sets $c_0 = 0$ and drops the degree normalization and eigenvalue scaling, taking $\hat{B}$ to be the top-k orthonormal eigenvectors of $\hat{W}$ itself, which keeps the spectral Davis–Kahan perturbation argument transparent.

E.6 Step F: Source-Only Modality Transport

On the source training sample, GraphBridge fits a transport map from the bridge embedding $\hat{H}_i^{(S)}$ to the source-only factor Z_i, or to a fixed-dimensional factor $F_i^{(S)} = \varphi(Z_i)$ of it. The default is ridge regression with a standardizer, and lasso, a small MLP, or a gradient-boosted model may be substituted. Writing $\hat{\theta}$ for the fitted map, the factor is transported to the target by

$$\hat{Z}_i^{(T)} = (\hat{H}_i^{(T)})^\top \hat{\theta}. \tag{21}$$

Analyzed reduction. The analysis takes φ linear and drops the ridge penalty, so $\hat{\theta}$ is the ordinary least-squares solution

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^k}{\arg\min} \, \frac{1}{n_S} \sum_{i=1}^{n_S} \big(Z_i - (\hat{H}_i^{(S)})^\top \theta\big)^2, \tag{22}$$

and the transported factor (21) is linear in the bridge embedding.

E.7 Step G: Target Prediction and Validation Safeguard

We define the transfer-augmented target design

$$\hat{X}_{i,\mathrm{tr}}^{(T)} = \big((\hat{H}_i^{(T)})^\top, \ \hat{Z}_i^{(T)}\big)^\top \in \mathbb{R}^{k+1}.$$

Algorithm 2 Step-by-step GraphBridge estimator

Input: Source sample $\mathcal{S}$, target sample $\mathcal{T}$, literature graph G^L with priors q^L, thresholds $\tau_A, \tau_\ell, \tau_s, \tau_t, \tau_d, \tau_\omega$, decay ν, bridge dimension k.

// Split target data before any target-supervised computation

Split $\mathcal{T}$ into disjoint folds $\mathcal{T}_{\text{gate}}$ for graph selection, $\mathcal{T}_{\text{tr}}$ for predictor training, $\mathcal{T}_{\text{val}}$ for the safeguard, and $\mathcal{T}_{\text{test}}$ for reporting.

// Step A — weighted literature prior

Build G^L from edge confidences q_e^L by (7).

// Step B — soft anchoring with abstention

Compute anchor scores $R_{j,m}^{(S)}, R_{j,m}^{(T)}$ and form soft anchors $\pi^{(S)}, \pi^{(T)}$ by (8), with hard reduction (9).

Aggregate family features $\hat{U}_i^{(S)}, \hat{U}_i^{(T)}$ by (10).

// Step C — source and target applicability views

Estimate edge support $c_{e,S}, c_{e,T}$ by (11) and view weights ψ_S, ψ_T by (12), and form family scores $\hat{s}_m$ on $\mathcal{S}$, $\hat{t}_m$ on $\mathcal{T}_{\text{gate}}$, and $\hat{d}_m$ on $\mathcal{S} \cup \mathcal{T}_{\text{gate}}$.

// Step D — pair-specific transfer graph

Form transfer weights ω_e^{tr} by (16), select edges with $\omega_e^{\text{tr}} \geq \tau_\omega$, and assemble $\hat{W}$, with hard reduction given by (17) and (18).

// Step E — spectral bridge encoding

Compute the top-k spectral basis $\hat{B}$ from (19) and form $\hat{H}_i^{(S)}, \hat{H}_i^{(T)}$ by (20).

// Step F — source-only modality transport

Fit $\hat{\theta}$ by source least squares (22) and transport $\hat{Z}_i^{(T)}$ by (21).

// Step G — prediction and safeguard

Fit transfer predictor $\hat{f}_1$ and target-only predictor $\hat{f}_0$ on $\mathcal{T}_{\text{tr}}$.

Select $\hat{f}_{\text{sel}}$ on $\mathcal{T}_{\text{val}}$ by (24) and report performance on $\mathcal{T}_{\text{test}}$.

return $\pi^{(S)}, \pi^{(T)}, \hat{\mathcal{M}}, \hat{W}, \hat{B}, \hat{\theta}, \hat{f}_1, \hat{f}_{\text{sel}}$.

We fit the transfer predictor by least squares on a target training split,

$$\hat{\psi}_{\text{tr}} = \arg\min_{\psi \in \mathbb{R}^{k+1}} \frac{1}{n_{T,\text{tr}}} \sum_{i \in \mathcal{I}_{T,\text{tr}}} \big(Y_i^{(T)} - (\hat{X}_{i,\text{tr}}^{(T)})^\top \psi\big)^2. \tag{23}$$

We also fit a target-only baseline $\hat{f}_0$, and we let $\hat{f}_1$ be the transfer predictor. Their validation predictions are

$$\hat{Y}_{i,0}^{(T)} = \hat{f}_0(X_i^{(T)}), \qquad \hat{Y}_{i,1}^{(T)} = \hat{f}_1(\hat{X}_{i,\text{tr}}^{(T)}),$$

and a validation split selects

$$\hat{m}_{\text{sel}} \in \arg\min_{m \in \{0,1\}} \frac{1}{n_{T,\text{val}}} \sum_{i \in \mathcal{I}_{T,\text{val}}} \ell\big(Y_i^{(T)}, \hat{Y}_{i,m}^{(T)}\big), \qquad \hat{f}_{\text{sel}} = \hat{f}_{\hat{m}_{\text{sel}}}. \tag{24}$$

In the multi-model evaluation, several arms are compared. These are the target-only *raw* arm, the *bridge* arm on $\hat{H}^{(T)}$, the *transfer* arm on $[\hat{H}^{(T)}, \hat{Z}^{(T)}]$, an *oracle* arm with the true source-only factor, and diagnostic baselines such as a PCA-compressed bridge and a naive-KG variant that skips graph cleaning. Each arm is fit with several regressors, including ridge, SVR, gradient boosting, lasso, and elastic net.

Analyzed reduction. The analysis uses linear least-squares predictors and restricts the safeguard (24) to a binary choice between the transfer predictor $\hat{f}_1$ and a target-only baseline $\hat{f}_0$, which is what the negative-transfer guarantee requires.

F Assumptions

We state assumptions in a way that is strong enough for clean proofs but flexible enough for later relaxation.

Assumption F.1 (Anchor score separation and sub-Gaussian noise). *For each domain d and raw variable j, there exist deterministic means $\mu_{j,m}^{(d)}$ such that*

$$R_{j,m}^{(d)} = \mu_{j,m}^{(d)} + \xi_{j,m}^{(d)}, \qquad 1 \le m \le M,$$

where $\xi_{j,m}^{(d)}$ is mean-zero σ_A^2-sub-Gaussian. In addition, there exist margins $\Delta_A, \Delta_0 > 0$ such that:

(i) if $a_j^{(d),\star} = m^\star \ge 1$, then $\mu_{j,m^\star}^{(d)} \ge \tau_A + 2\Delta_A$ and $\mu_{j,m^\star}^{(d)} - \max_{m \ne m^\star} \mu_{j,m}^{(d)} \ge 2\Delta_A$;

(ii) if $a_j^{(d),\star} = 0$, then $\max_{1 \le m \le M} \mu_{j,m}^{(d)} \le \tau_A - 2\Delta_0$.

Assumption F.2 (Family score concentration and transfer margins). *For each family m, the estimators $\hat{s}_m, \hat{t}_m, \hat{d}_m$ satisfy, for every $x > 0$,*

$$\mathbb{P}\big(|\hat{s}_m - s_m^\star| > x\big) \le 2e^{-c_s n_S x^2}, \quad \mathbb{P}\big(|\hat{t}_m - t_m^\star| > x\big) \le 2e^{-c_t n_T x^2}, \quad \mathbb{P}\Big(\Big|\hat{d}_m - d_m^\star\Big| > x\Big) \le 2e^{-c_d n x^2},$$

where $n = \min(n_S, n_T)$. There exists $\Delta_F > 0$ such that for each m:

(i) if $m \in \mathcal{M}^\star$, then $\ell_m^\star \ge \tau_\ell$, $s_m^\star \ge \tau_s + 2\Delta_F$, $t_m^\star \ge \tau_t + 2\Delta_F$, and $d_m^\star \le \tau_d - 2\Delta_F$;

(ii) if $m \notin \mathcal{M}^\star$, then at least one of the following holds: $\ell_m^\star < \tau_\ell$, $s_m^\star \le \tau_s - 2\Delta_F$, $t_m^\star \le \tau_t - 2\Delta_F$, or $d_m^\star \ge \tau_d + 2\Delta_F$.

Assumption F.3 (Graph perturbation and eigengap). *The population transfer graph $W^\star$ has eigengap*

$$\lambda_k(W^\star) - \lambda_{k+1}(W^\star) \ge \mathrm{gap} > 0.$$

Moreover, on the family-selection event $\{\hat{\mathcal{M}} = \mathcal{M}^\star\}$, the graph perturbation satisfies

$$\big\|\hat{W} - W^\star\big\|_{\mathrm{op}} \le \eta_W.$$

Assumption F.4 (Design regularity). *There exist constants $0 < c_- \le c_+ < \infty$ such that with probability at least $1 - o(1)$,*

$$c_- I_k \preceq \frac{1}{n_S}(H_S^\star)^\top H_S^\star \preceq c_+ I_k, \qquad c_- I_k \preceq \frac{1}{n_T}(H_T^\star)^\top H_T^\star \preceq c_+ I_k.$$

Rows of $H_S^\star$ and $H_T^\star$ are sub-Gaussian with uniformly bounded second moments.

Assumption F.5 (Noise). *The noises $\varepsilon_i^{(S)}$ and $\varepsilon_i^{(T)}$ are conditionally mean-zero and σ_Z^2-sub-Gaussian and σ_Y^2-sub-Gaussian, respectively.*

Assumption F.6 (Bounded bridge parameters). $\|\theta^\star\|_2 \le C_\theta$, $\|\beta^\star\|_2 \le C_\beta$, *and* $|\gamma^\star| \le C_\gamma$.

G Finite-Sample Results

G.1 Anchor Consistency

Theorem G.1 (Anchor recovery with abstention). *Under Assumption F.1, for any domain d and any raw variable j,*

$$\mathbb{P}\big(\hat{a}_j^{(d)} \ne a_j^{(d),\star}\big) \le 2(M+1)\exp\Big(-\frac{\min\{\Delta_A^2, \Delta_0^2\}}{2\sigma_A^2}\Big).$$

Hence, by a union bound,

$$\mathbb{P}\big(\hat{A}_d \ne A_d^\star\big) \le 2p_d(M+1)\exp\Big(-\frac{\min\{\Delta_A^2, \Delta_0^2\}}{2\sigma_A^2}\Big).$$

G.2 Transfer-Family Selection Consistency

Remark 1 (Conservative gating and the precision–recall asymmetry). Assumption F.2 is stated symmetrically: the family margin Δ_F separates $m \in \mathcal{M}^\star$ from $m \notin \mathcal{M}^\star$ in both directions. In practice, the discrepancy score $\hat{d}_m$ may not concentrate as fast as the utility scores $\hat{s}_m, \hat{t}_m$, leading to a gate that achieves high edge precision ($\hat{\mathcal{M}} \subseteq \mathcal{M}^\star$) but moderate recall ($\mathcal{M}^\star \not\subseteq \hat{\mathcal{M}}$). This conservative regime is consistent with the mechanism-level checks in **??**, where gating improves transfer-edge precision under applicability shift and KG corruption. Importantly, Theorem G.3 remains valid under conservative gating: if $\hat{\mathcal{M}} \subseteq \mathcal{M}^\star$, the estimated transfer graph $\hat{W} = \mathrm{diag}(\hat{g})\, G\, \mathrm{diag}(\hat{g})$ is a subgraph of the true transfer graph, and the perturbation bound $\left\|\hat{W} - W^\star\right\|_{\mathrm{op}} \le \eta_W$ holds with a larger η_W that reflects the omitted families.

Theorem G.2 (Family gate selection consistency). *Under Assumption F.2, if $\Delta_F > 0$ is the family margin and the gate is defined by (17), then*

$$\mathbb{P}(\hat{\mathcal{M}} \neq \mathcal{M}^\star) \le 6M \exp(-c_F n \Delta_F^2),$$

where $c_F = \min\{c_s, c_t, c_d\}$ and $n = \min(n_S, n_T)$. Consequently, if $\log M = o(n\Delta_F^2)$, then $\mathbb{P}(\hat{\mathcal{M}} = \mathcal{M}^\star) \to 1$.

G.3 Bridge Basis Recovery

Theorem G.3 (Spectral bridge-space recovery). *Assume Assumption F.3. On the event $\{\hat{\mathcal{M}} = \mathcal{M}^\star\}$, there exists an orthogonal matrix $Q \in \mathbb{R}^{k\times k}$ such that*

$$\left\|\hat{B} - B^\star Q\right\|_F \le \frac{2\sqrt{2k}}{\mathrm{gap}}\, \eta_W. \tag{25}$$

If, in addition, $\hat{A}_S = A_S^\star$ and $\hat{A}_T = A_T^\star$, then the estimated bridge representations satisfy

$$\frac{1}{n_d}\left\|\hat{H}_d - H_d^\star Q\right\|_F^2 \le \frac{8\,\lambda_{\max}\big((U_d^\star)^\top U_d^\star/n_d\big)\,k}{\mathrm{gap}^2}\,\eta_W^2, \qquad d \in \{S, T\}. \tag{26}$$

G.4 Transport Error

Theorem G.4 (Transport-parameter estimation and target latent-factor error). *Assume Assumption F.3, Assumption F.4, Assumption F.5, and Assumption F.6, and suppose the anchor and family-selection events hold. Let Q be as in Theorem G.3. Then with probability at least $1 - o(1)$,*

$$\left\|\hat{\theta} - Q^\top \theta^\star\right\|_2 \le C\left(\sqrt{\frac{k}{n_S}} + \frac{\sqrt{k}\,\eta_W}{\mathrm{gap}}\right), \tag{27}$$

for a constant C depending only on $c_-, c_+, \sigma_Z, C_\theta$. Moreover,

$$\frac{1}{n_T}\sum_{i=1}^{n_T}\big(\hat{Z}_i^{(T)} - Z_i^{\star,(T)}\big)^2 \le C'\left(\frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2}\right) \tag{28}$$

with probability at least $1 - o(1)$.

G.5 Target Excess Risk

Theorem G.5 (Transfer predictor excess risk). *Assume Assumption F.3, Assumption F.4, Assumption F.5, and Assumption F.6, together with the structural model (6). Suppose the anchor and family-selection events hold, and let $\hat{f}_1$ be the transfer predictor fitted on $(\hat{H}_i^{(T)}, \hat{Z}_i^{(T)})$. Then for a fresh target draw $(X_{\mathrm{new}}^{(T)}, Y_{\mathrm{new}}^{(T)})$,*

$$\mathcal{R}_T(\hat{f}_1) - \mathcal{R}_T(f^\star) \le C\left(\frac{k+1}{n_T} + \frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2}\right) \tag{29}$$

with probability at least $1 - o(1)$, where $f^\star$ is the oracle predictor using the true bridge representation and the true transported latent factor.

G.6 Validation Safeguard Against Negative Transfer

Theorem G.6 (Model selection guarantee for the validation selector)**.** *Suppose the loss $\ell(y, \hat{y}) \in [0, 1]$ almost surely. Let $\hat{f}_0$ be any target-only baseline predictor, $\hat{f}_1$ the transfer predictor, and $\hat{f}_{\text{sel}}$ the selector defined in (24) on a validation set of size $n_{T,\text{val}}$. Then for every $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\mathcal{R}_T(\hat{f}_{\text{sel}}) \le \min\{\mathcal{R}_T(\hat{f}_0), \mathcal{R}_T(\hat{f}_1)\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\text{val}}}}. \tag{30}$$

Remark 2 (Linear-theory collinearity). In the simplified linear theory, $Z_i^{\star,(T)} = (H_i^{(T),\star})^\top \theta^\star$, so the transported latent feature lies in the same linear span as the bridge representation. Consequently, the risk proof does not require the augmented target design $[\hat{H}_T, \hat{Z}_T]$ to be full rank. The displayed k/n_S term is retained from the transport-error bound and is harmless but generally loose for purely linear target least squares with $\hat{H}_T$ already included. Benefits from the transported feature beyond this span require nonlinear prediction, restricted target models, or source-only signal structure not represented by the linear bridge model.

Remark 3 (Source sample size and the collinearity regime). The term k/n_S in (29) captures estimation error of the transport parameter $\hat{\theta}$ on the source sample. This term is active when $Z^{(S)}$ has variance orthogonal to the bridge representation $H^{(S)}$ — that is, when the source-only latent feature carries information not already encoded in the graph. In a synthetic data-generating process where both $Z^{(S)}$ and $H^{(S)}$ are linear functions of the same latent feature, the two are collinear and $\theta^\star \approx 0$, making $\hat{\theta}$ trivially accurate regardless of n_S. The k/n_S term is therefore empirically vacuous in such degenerate settings and manifests only when the source modality carries signal beyond what the bridge subspace alone captures.

Corollary G.1 (End-to-end bound with negative-transfer safeguard)**.** *Under the assumptions of Theorem G.1, Theorem G.2, Theorem G.3, Theorem G.4, Theorem G.5, and Theorem G.6, the validation-selected predictor obeys*

$$\mathcal{R}_T(\hat{f}_{\text{sel}}) \le \min\left\{\mathcal{R}_T(\hat{f}_0),\ \mathcal{R}_T(f^\star) + C\left(\frac{k+1}{n_T} + \frac{k}{n_S} + \frac{k\eta_W^2}{\text{gap}^2}\right)\right\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\text{val}}}}$$

with probability at least

$$1 - \delta - 2\sum_{d\in\{S,T\}} p_d(M+1)e^{-\min\{\Delta_A^2, \Delta_0^2\}/(2\sigma_A^2)} - 6Me^{-c_F n \Delta_F^2} - o(1).$$

H Proofs of Main Results

H.1 Proof of Theorem G.1

Fix a domain d and variable j.

Case 1: Non-Null Anchor. Suppose $a_j^{(d),\star} = m^\star \ge 1$. By construction, the event $\{\hat{a}_j^{(d)} \ne m^\star\}$ can only happen in one of two ways:

(a) the true score fails the threshold, namely $R_{j,m^\star}^{(d)} < \tau_A$, or

(b) some competing score beats the true score, namely $R_{j,m}^{(d)} \ge R_{j,m^\star}^{(d)}$ for some $m \ne m^\star$.

For event (a), by Assumption F.1, $\mu_{j,m^\star}^{(d)} \ge \tau_A + 2\Delta_A$. Hence

$$\{R_{j,m^\star}^{(d)} < \tau_A\} \subseteq \{\left|R_{j,m^\star}^{(d)} - \mu_{j,m^\star}^{(d)}\right| \ge \Delta_A\}.$$

Because $R_{j,m^\star}^{(d)} - \mu_{j,m^\star}^{(d)}$ is σ_A^2-sub-Gaussian,

$$\mathbb{P}\big(R_{j,m^\star}^{(d)} < \tau_A\big) \le 2\exp\Big(-\frac{\Delta_A^2}{2\sigma_A^2}\Big). \tag{31}$$

For event (b), if $R^{(d)}_{j,m} \geq R^{(d)}_{j,m^\star}$, then at least one of the deviations $R^{(d)}_{j,m} - \mu^{(d)}_{j,m}$ or $R^{(d)}_{j,m^\star} - \mu^{(d)}_{j,m^\star}$ has magnitude at least Δ_A. Therefore,

$$\mathbb{P}\big(R^{(d)}_{j,m} \geq R^{(d)}_{j,m^\star}\big) \leq 4\exp\Big(-\frac{\Delta_A^2}{2\sigma_A^2}\Big).$$

Summing over $m \neq m^\star$ and combining with (31),

$$\mathbb{P}\big(\hat{a}^{(d)}_j \neq m^\star\big) \leq 2(M+1)\exp\Big(-\frac{\Delta_A^2}{2\sigma_A^2}\Big).$$

Case 2: Null Anchor. Suppose $a^{(d),\star}_j = 0$. Then $\hat{a}^{(d)}_j \neq 0$ means there exists some $m \in [M]$ such that $R^{(d)}_{j,m} \geq \tau_A$. Since $\mu^{(d)}_{j,m} \leq \tau_A - 2\Delta_0$,

$$\mathbb{P}\big(\hat{a}^{(d)}_j \neq 0\big) \leq 2M\exp\Big(-\frac{\Delta_0^2}{2\sigma_A^2}\Big) \leq 2(M+1)\exp\Big(-\frac{\Delta_0^2}{2\sigma_A^2}\Big).$$

Combining the non-null and null cases proves the row-wise claim. The matrix claim follows from a union bound over $j = 1, \ldots, p_d$. □

H.2 Proof of Theorem G.2

Define the good score-concentration event

$$\mathcal{E}_F = \bigcap_{m=1}^{M} \Big\{|\hat{s}_m - s^\star_m| \leq \Delta_F,\ |\hat{t}_m - t^\star_m| \leq \Delta_F,\ \Big|\hat{d}_m - d^\star_m\Big| \leq \Delta_F\Big\}.$$

By Assumption F.2 and a union bound,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_F^c) &\leq \sum_{m=1}^{M} \mathbb{P}\big(|\hat{s}_m - s^\star_m| > \Delta_F\big) + \sum_{m=1}^{M} \mathbb{P}\big(|\hat{t}_m - t^\star_m| > \Delta_F\big) + \sum_{m=1}^{M} \mathbb{P}\Big(\Big|\hat{d}_m - d^\star_m\Big| > \Delta_F\Big) \\
&\leq 2Me^{-c_s n_S \Delta_F^2} + 2Me^{-c_t n_T \Delta_F^2} + 2Me^{-c_d n \Delta_F^2} \\
&\leq 6Me^{-c_F n \Delta_F^2},
\end{aligned}$$

where $c_F = \min\{c_s, c_t, c_d\}$ and $n = \min(n_S, n_T)$.

Now work on the event $\mathcal{E}_F$. If $m \in \mathcal{M}^\star$, then by Assumption F.2,

$$\hat{s}_m \geq s^\star_m - \Delta_F \geq \tau_s + \Delta_F > \tau_s, \qquad \hat{t}_m \geq t^\star_m - \Delta_F \geq \tau_t + \Delta_F > \tau_t,$$

while $\hat{d}_m \leq d^\star_m + \Delta_F \leq \tau_d - \Delta_F < \tau_d$. Also $\ell^\star_m \geq \tau_\ell$. Therefore $\hat{g}_m = 1$.

If $m \notin \mathcal{M}^\star$, then at least one of the transfer conditions fails by a margin $2\Delta_F$. If, for instance, $s^\star_m \leq \tau_s - 2\Delta_F$, then on $\mathcal{E}_F$,

$$\hat{s}_m \leq s^\star_m + \Delta_F \leq \tau_s - \Delta_F < \tau_s,$$

so $\hat{g}_m = 0$. The same logic applies if the failing coordinate is $t^\star_m$, $d^\star_m$, or the deterministic literature criterion $\ell^\star_m < \tau_\ell$. Hence $\hat{g}_m = g^\star_m$ for every m, and therefore $\hat{\mathcal{M}} = \mathcal{M}^\star$ on $\mathcal{E}_F$.

Thus $\mathbb{P}(\hat{\mathcal{M}} \neq \mathcal{M}^\star) \leq 6Me^{-c_F n \Delta_F^2}$. □

H.3 Proof of Theorem G.3

On the event $\{\hat{\mathcal{M}} = \mathcal{M}^\star\}$, we have $\mathrm{diag}(\hat{g}) = \mathrm{diag}(g^\star)$, so $\hat{W}$ is a perturbation of $W^\star$ with operator norm at most η_W. By the Davis–Kahan $\sin\Theta$ theorem (Davis and Kahan, 1970; Yu et al., 2015),

$$\big\|\sin\Theta(\hat{\mathcal{U}}, \mathcal{U})\big\|_{\mathrm{op}} \leq \frac{\big\|\hat{W} - W^\star\big\|_{\mathrm{op}}}{\mathrm{gap}} \leq \frac{\eta_W}{\mathrm{gap}}. \tag{32}$$

The standard Procrustes consequence yields an orthogonal matrix $Q \in \mathbb{R}^{k\times k}$ such that

$$\left\|\hat{B} - B^\star Q\right\|_F \le 2\sqrt{2k}\left\|\sin\Theta(\hat{\mathcal{U}},\mathcal{U})\right\|_{\mathrm{op}} \le \frac{2\sqrt{2k}}{\mathrm{gap}}\eta_W,$$

proving (25).

Now assume $\hat{A}_d = A_d^\star$. Then $\hat{U}_d = U_d^\star$, so $\hat{H}_d - H_d^\star Q = U_d^\star(\hat{B} - B^\star Q)$. Taking Frobenius norms, squaring, and dividing by n_d gives (26). □

H.4 Auxiliary Lemma for Least-Squares Perturbation Bounds

Lemma H.1 (Perturbed least squares). *Let $\widetilde{H} \in \mathbb{R}^{n\times k}$ and $\hat{H} \in \mathbb{R}^{n\times k}$ satisfy $\lambda_{\min}(\widetilde{H}^\top\widetilde{H}/n) \ge c_- > 0$ and $\|\hat{H} - \widetilde{H}\|_{\mathrm{op}}/\sqrt{n} \le r$. If $y = \widetilde{H}\vartheta^\star + \varepsilon$ and $\hat{\vartheta} = (\hat{H}^\top\hat{H})^{-1}\hat{H}^\top y$, then*

$$\left\|\hat{\vartheta} - \vartheta^\star\right\|_2 \le \frac{2}{c_-}\left\|\frac{1}{n}\hat{H}^\top\varepsilon\right\|_2 + \frac{2}{c_-}\left\|\frac{1}{n}\hat{H}^\top(\widetilde{H} - \hat{H})\right\|_{\mathrm{op}}\|\vartheta^\star\|_2.$$

Moreover, if rows of $\hat{H}$ are sub-Gaussian and ε is independent sub-Gaussian noise, then $\left\|\hat{\vartheta} - \vartheta^\star\right\|_2 = O_p(\sqrt{k/n} + r)$.

H.5 Proof of Theorem G.4

On the anchor and family-selection events, Theorem G.3 provides Q such that $\|\hat{H}_S - H_S^\star Q\|_{\mathrm{op}}/\sqrt{n_S} \le C_B\sqrt{k}\eta_W/\mathrm{gap} =: r_S$. Applying Lemma H.1 with $(\hat{H}, \widetilde{H}, \vartheta^\star) = (\hat{H}_S, H_S^\star Q, Q^\top\theta^\star)$ yields (27).

For the transported factor error, decompose $\hat{Z}_T - Z_T^\star = \hat{H}_T(\hat{\theta} - Q^\top\theta^\star) + (\hat{H}_T - H_T^\star Q)Q^\top\theta^\star$ and apply $(a+b)^2 \le 2a^2 + 2b^2$ to obtain (28). □

H.6 Proof of Theorem G.5

In the simplified linear target model,

$$Y_i^{(T)} = (H_i^{(T),\star})^\top\beta^\star + \gamma^\star(H_i^{(T),\star})^\top\theta^\star + \varepsilon_i^{(T)} = (H_i^{(T),\star})^\top\alpha^\star + \varepsilon_i^{(T)},$$

where $\alpha^\star = \beta^\star + \gamma^\star\theta^\star$. Also $\hat{Z}_T = \hat{H}_T\hat{\theta}$, so the columns of the augmented design $[\hat{H}_T, \hat{Z}_T]$ have the same span as the columns of $\hat{H}_T$. Thus the least-squares fitted values of the transfer predictor are the same as the fitted values obtained by regressing Y_T on $\hat{H}_T$ alone, even if the augmented coefficients are not unique.

By Theorem G.3,

$$\|\hat{H}_T - H_T^\star Q\|_{\mathrm{op}}/\sqrt{n_T} \le \|\hat{H}_T - H_T^\star Q\|_F/\sqrt{n_T} = O_p(\sqrt{k}\eta_W/\mathrm{gap}).$$

Applying Lemma H.1 with $(\hat{H}, \widetilde{H}, \vartheta^\star) = (\hat{H}_T, H_T^\star Q, Q^\top\alpha^\star)$ and using Assumption F.4, Assumption F.5, and Assumption F.6 gives

$$\|\hat{\alpha} - Q^\top\alpha^\star\|_2 = O_p\left(\sqrt{\frac{k}{n_T}} + \frac{\sqrt{k}\eta_W}{\mathrm{gap}}\right).$$

Computing the prediction error for a fresh target draw with bounded second moment then yields

$$\mathcal{R}_T(\hat{f}_1) - \mathcal{R}_T(f^\star) \le C\left(\frac{k}{n_T} + \frac{k\eta_W^2}{\mathrm{gap}^2}\right).$$

This is stronger than (29), since the displayed bound in (29) includes the nonnegative transport-estimation term k/n_S and uses $k+1$ in place of k. □

H.7 Proof of Theorem G.6

Fix the trained candidate predictors and let $\hat{L}_m = n_{T,\mathrm{val}}^{-1} \sum_{i \in \mathcal{I}_{T,\mathrm{val}}} \ell(Y_i^{(T)}, \hat{Y}_{i,m}^{(T)})$ and $L_m = \mathcal{R}_T(\hat{f}_m)$. By Hoeffding's inequality (loss bounded in $[0,1]$) and a union bound over $m \in \{0,1\}$, with probability at least $1-\delta$, both $\left|\hat{L}_m - L_m\right| \le \epsilon$ hold with $\epsilon = \sqrt{\log(4/\delta)/(2n_{T,\mathrm{val}})}$. On this event,

$$\mathcal{R}_T(\hat{f}_{\mathrm{sel}}) \le \hat{L}_{\mathrm{sel}} + \epsilon \le \min\{\hat{L}_0, \hat{L}_1\} + \epsilon \le \min\{L_0, L_1\} + 2\epsilon.$$

Substituting the value of ϵ proves (30). □

H.8 Proof of Corollary G.1

Proof. Define the anchor-recovery event

$$\mathcal{E}_A = \{\widehat{A}_S = A_S^\star,\ \widehat{A}_T = A_T^\star\},$$

and the family-selection event

$$\mathcal{E}_F = \{\widehat{\mathcal{M}} = \mathcal{M}^\star\}.$$

By Theorem G.1 and a union bound over the source and target domains,

$$\Pr(\mathcal{E}_A^c) \le 2 \sum_{d \in \{S,T\}} p_d(M+1) \exp\left\{-\frac{\min\{\Delta_A^2, \Delta_0^2\}}{2\sigma_A^2}\right\}.$$

Let

$$\delta_A := 2 \sum_{d \in \{S,T\}} p_d(M+1) \exp\left\{-\frac{\min\{\Delta_A^2, \Delta_0^2\}}{2\sigma_A^2}\right\}.$$

By Theorem G.2,

$$\Pr(\mathcal{E}_F^c) \le 6M \exp\{-c_F n \Delta_F^2\},$$

where $c_F = \min\{c_s, c_t, c_d\}$ and

$$n = \min\{n_S, n_T\}.$$

Thus, by a union bound,

$$\Pr(\mathcal{E}_A \cap \mathcal{E}_F) \ge 1 - \delta_A - 6M \exp\{-c_F n \Delta_F^2\}.$$

Now work on the event $\mathcal{E}_A \cap \mathcal{E}_F$. Under the assumptions of Theorems G.3–G.5, Theorem G.5 implies that, with probability at least $1 - o(1)$ conditional on $\mathcal{E}_A \cap \mathcal{E}_F$,

$$R_T(\widehat{f}_1) - R_T(f^\star) \le C\left(\frac{k+1}{n_T} + \frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2}\right).$$

Equivalently, defining

$$r_{\mathrm{tr}} := C\left(\frac{k+1}{n_T} + \frac{k}{n_S} + \frac{k\eta_W^2}{\mathrm{gap}^2}\right),$$

there exists an event $\mathcal{E}_R$ such that

$$\mathcal{E}_R = \left\{R_T(\widehat{f}_1) \le R_T(f^\star) + r_{\mathrm{tr}}\right\},$$

and

$$\Pr(\mathcal{E}_R^c \cap \mathcal{E}_A \cap \mathcal{E}_F) \le o(1).$$

Therefore,

$$\Pr(\mathcal{E}_A \cap \mathcal{E}_F \cap \mathcal{E}_R) \ge 1 - \delta_A - 6M \exp\{-c_F n \Delta_F^2\} - o(1).$$

Next define the validation event

$$\mathcal{E}_{\mathrm{val}} = \left\{R_T(\widehat{f}_{\mathrm{sel}}) \le \min\{R_T(\widehat{f}_0), R_T(\widehat{f}_1)\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\mathrm{val}}}}\right\}.$$

By Theorem G.6, since the validation loss is bounded in $[0,1]$ and the validation split is independent of the data used to fit $\widehat{f}_0$ and $\widehat{f}_1$,

$$\Pr(\mathcal{E}_{\mathrm{val}}) \geq 1-\delta.$$

Applying another union bound gives

$$\Pr(\mathcal{E}_A \cap \mathcal{E}_F \cap \mathcal{E}_R \cap \mathcal{E}_{\mathrm{val}}) \geq 1-\delta-\delta_A-6M\exp\{-c_F n\Delta_F^2\}-o(1).$$

It remains to prove the risk inequality on this intersection of events. On $\mathcal{E}_{\mathrm{val}}$,

$$R_T(\widehat{f}_{\mathrm{sel}}) \leq \min\{R_T(\widehat{f}_0), R_T(\widehat{f}_1)\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\mathrm{val}}}}.$$

On $\mathcal{E}_R$,

$$R_T(\widehat{f}_1) \leq R_T(f^\star) + r_{\mathrm{tr}}.$$

Hence,

$$\min\{R_T(\widehat{f}_0), R_T(\widehat{f}_1)\} \leq \min\{R_T(\widehat{f}_0), R_T(f^\star)+r_{\mathrm{tr}}\}.$$

Substituting the definition of r_{tr}, we obtain

$$R_T(\widehat{f}_{\mathrm{sel}}) \leq \min\left\{R_T(\widehat{f}_0), R_T(f^\star) + C\left(\frac{k+1}{n_T}+\frac{k}{n_S}+\frac{k\eta_W^2}{\mathrm{gap}^2}\right)\right\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\mathrm{val}}}}.$$

Combining this deterministic implication with the probability bound above yields

$$R_T(\widehat{f}_{\mathrm{sel}}) \leq \min\left\{R_T(\widehat{f}_0), R_T(f^\star) + C\left(\frac{k+1}{n_T}+\frac{k}{n_S}+\frac{k\eta_W^2}{\mathrm{gap}^2}\right)\right\} + 2\sqrt{\frac{\log(4/\delta)}{2n_{T,\mathrm{val}}}}$$

with probability at least

$$1-\delta-2\sum_{d\in\{S,T\}} p_d(M+1)\exp\left\{-\frac{\min\{\Delta_A^2,\Delta_0^2\}}{2\sigma_A^2}\right\} - 6M\exp\{-c_F n\Delta_F^2\} - o(1).$$

□

I Remarks

Remark 4 (Zero raw overlap). The theory does not require any overlap between the coordinates of $X^{(S)}$ and $X^{(T)}$. The only overlap requirement is that the domains map into the same transferable family space and hence into the same bridge subspace.

Remark 5 (What is still missing). This analysis does not yet prove an impossibility theorem for disjoint bridge subspaces, does not cover generalized linear models or survival outcomes, and does not yet account for cross-fitting or post-selection inference. Those are natural next steps.

Remark 6 (Population degeneracy under disjoint bridge subspaces). Let $B_S \in \mathbb{R}^{M\times k_S}$ and $B_T \in \mathbb{R}^{M\times k_T}$ denote the population bridge bases induced by the source and target transferable graphs, and define the overlap measure $\rho = \|B_T^\top B_S\|_{\mathrm{op}}$. When $\rho = 0$ (equivalently, $\mathrm{col}(B_S) \perp \mathrm{col}(B_T)$), the source-side latent factor has no component in the target bridge space, so the transport step cannot improve upon the bridge-only predictor even with perfect transport and arbitrarily large n_S. In particular, with $\theta_T^\star := (B_T^\top B_S)\theta_S^\star$,

$$\rho = 0 \implies \theta_T^\star = 0 \implies \mathbb{E}\Big[\widehat{Z}^T \mid H^T\Big] = 0 \implies R_T(\widehat{f}_{\mathrm{tr}}) = R_T(\widehat{f}_{\mathrm{br}}) \quad \text{in population,}$$

so the excess risk of the transfer predictor over the bridge-only baseline is exactly zero. Practically, this makes subspace overlap a natural diagnostic: if $\widehat{Z}^T$ is near zero across the target cohort, it is consistent with $\mathrm{col}(B_S) \approx \mathrm{col}(B_T)^\perp$ and the transport step is expected to add little.

J Synthetic Data-Generating Process

This appendix provides full details of the synthetic data-generating process summarized in **??**.

J.1 Population Graph

The population graph $G^{\mathrm{bio},\star}$ is defined over M candidate relation families. A subset $\mathcal{M}^\star \subseteq [M]$ of families is designated as transferable, with $|\mathcal{M}^\star| = M_\star$. An $M \times M$ symmetric family-level adjacency matrix $\widetilde{G}^\star$ is constructed by sampling edges within and between transferable families with specified density; edges involving non-transferable families are included at lower density to create noise. The population transfer graph is $W^\star = \mathrm{diag}(g^\star)\,\widetilde{G}^\star\,\mathrm{diag}(g^\star)$ with gate vector $g^\star_m = \mathbb{I}\{m \in \mathcal{M}^\star\}$.

J.2 Bridge Basis

The top-k eigenvectors of $W^\star$ form the population bridge basis $B^\star \in \mathbb{R}^{M\times k}$.

J.3 Feature Generation

For each domain $d \in \{S, T\}$, p_d raw variables are generated. Each variable j is assigned to a family $a_j^{(d)} \in \{0, 1, \ldots, M\}$ (with 0 denoting null/unassigned). Family-aggregated features are $U_i^{(d)} \in \mathbb{R}^M$, obtained by summing all raw variables assigned to each family. The bridge representation for subject i is $H_i^{(d)} = U_i^{(d)} B^\star \in \mathbb{R}^k$, with additive sub-Gaussian noise on raw features controlled by parameter σ_X.

J.4 Source-Only Modality

The source-only latent factor is generated as $Z_i = (H_i^{(S)})^\top \theta^\star + \varepsilon_i^{(S)}$, where $\theta^\star \in \mathbb{R}^k$ and $\varepsilon^{(S)} \sim \mathcal{N}(0, \sigma_Z^2)$.

J.5 Outcome Model

Target outcomes are generated from a logistic or linear model: $Y_i^{(T)} = (H_i^{(T)})^\top \beta^\star + \gamma^\star Z_i^{\star,(T)} + \varepsilon_i^{(T)}$, where $Z_i^{\star,(T)} = (H_i^{(T)})^\top \theta^\star$ is the latent transported signal.

J.6 External Graph Corruption

The observed external graph is constructed from $\widetilde{G}^\star$ by introducing false-positive edges (rate α_{FP}) and removing true edges (rate α_{FN}).

J.7 Default Parameters

Unless otherwise noted, we use $n_S = 400$, $n_T = 200$, $p_S = p_T = 60$, $M = 10$, $k = 3$, $M_\star = 6$, $\sigma_X = 0.5$, $\sigma_Z = 1.0$, $\alpha_{\mathrm{FP}} = \alpha_{\mathrm{FN}} = 0.05$. Results are averaged over 10 random seeds unless otherwise stated.

J.8 Scenario Perturbations

Five main scenarios probe specific failure modes:

(i) *Clean bridge*: no corruption or shift.
(ii) *Zero raw overlap*: source and target share no raw variable indices.
(iii) *Applicability shift*: some transferable families have reduced target support.
(iv) *Noisy KG*: elevated $\alpha_{\mathrm{FP}}, \alpha_{\mathrm{FN}}$ up to 10%.
(v) *Adversarial transfer*: sign of imaging contribution to target outcome is flipped.

K Synthetic Experiments

Validation safeguard prevents negative transfer. Fig. 3 studies the behavior of the validation safeguard under the adversarial scenario. Despite this design, GraphBridge (0.737 ± 0.027) exceeds

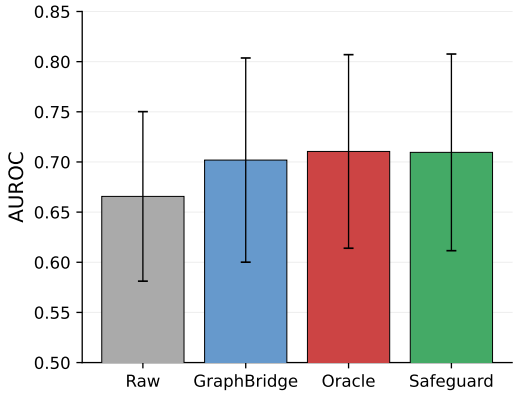

Figure 3: Effect of Validation safeguard.

Table 6: Summary of PLATO benchmark datasets.

Dataset	Type	Samples (n)	Genes (d_g)	Drugs (d_r)	Total ($d_g + d_r$)
BC	Cell line	17,074	18,061	427	18,488
CH	Cell line	1,847	18,006	406	18,412
ME	Cell line	17,015	17,902	422	18,324
NSCLC	Cell line	28,004	18,108	426	18,534
SCLC	Cell line	17,889	18,237	427	18,664
BRCA	Mouse PDX	768	13,235	18	13,253
CM	Mouse PDX	484	12,983	12	12,995
CRC	Mouse PDX	803	12,493	21	12,514
MNSCLC	Mouse PDX	523	13,390	19	13,409
PDAC	Mouse PDX	576	12,732	15	12,747

Raw (0.679 ± 0.063) by $+0.058$ and trails the Oracle (0.745 ± 0.031) by only 0.008. Two mechanisms contribute to this robustness. First, the target-side ridge predictor can learn negative coefficients for the transported dimensions whose contribution has been sign-flipped, effectively absorbing the adversarial manipulation at the prediction step. Second, the validation safeguard (Theorem G.6) provides a distribution-free fallback. It compares the transfer and target-only predictors on held-out data and retains whichever achieves lower validation loss, guaranteeing that the selected model is no worse than Raw up to an $O(n_{\text{val}}^{-1/2})$ penalty.

L Details of Real Data Experiments

Compute resources. All the experiments are conducted on a single AMD EPYC 7543 32-Core Processor. Total running time for each cohort of PLATO is approximately 36 hours. Synthetic experiments run on CPU in under 10 minutes per scenario.

Datasets. The PLATO benchmark comprises five cohorts of genomics of drug sensitivity in cancer (GDSC) and five patient-derived xenograft (PDX) mouse model cohorts. The GDSC cohorts include breast carcinoma (BC), chondrosarcoma (CH), melanoma (ME), non-small cell lung cancer (NSCLC; abbr. NS), and small cell lung cancer (SCLC; abbr. SC). The PDX cohorts include breast carcinoma (BRCA; abbr. BR), cutaneous melanoma (CM), colorectal cancer (CRC), non-small cell lung carcinoma (MNSCLC; abbr. MN), and pancreatic ductal adenocarcinoma (PDAC; abbr. PD). Table 6 shows the statistics of the datasets in the PLATO benchmark. The knowledge graph is BioKG (Walsh et al., 2020), a multi-relational biomedical graph containing approximately 101,000 entities, over 3 million edges, 99 relation types, and 7 node types.

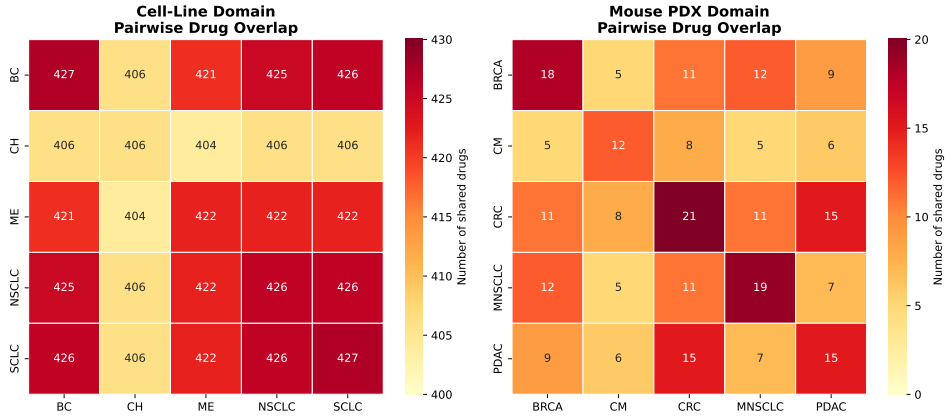

Figure 4: The numbers of overlapped drugs between datasets in each domain.

Setup. As shown in Fig. 4, we first examine the overlapped drugs between dataset pairs inside each domain. Since the number of genes is much larger than that of drugs, to better compare the impact of the dimensions of the missing and observed modalities, we set the total number of sampled features as 500, and determine the number of genes as $500 - d_{\text{overlap}}$, where d_{overlap} is the dimension of overlapped drugs. In that case, the cell-line cohorts will have more drugs as the missing modality ($Z^{(d)}$), while the mouse PDX cohorts will have more genes as the observed modality ($X^{(d)}$).

With the selected features, we build an undirected graph $G = (\mathcal{V}, \mathcal{E})$ from the BioKG. The full BioKG contains $|\mathcal{V}| = 100{,}937$ nodes and $|\mathcal{E}| = 3{,}066{,}156$ edges. The node set includes all entity indices $\{0, \ldots, |\mathcal{V}| - 1\}$; edges are drawn from the `head`–`tail` columns of the KG edge table, treated as undirected (i.e., both (h, t) and (t, h) are present in G). All edge weights are uniform (unweighted graph).

Default configuration. Unless stated otherwise, the steps of Section E use the following defaults. The anchor score in Step B uses prior weight $\lambda_{\text{pr}} = 1.0$ and neighborhood weight $\lambda_{\text{nbr}} = 1.15$, and the soft anchor uses temperature $\varsigma = 0.85$. The view weight in Step C uses prior weight $\beta_q = 0.7$ and support weight $\beta_c = 1.7$. The soft gate in Step D uses support weight $\alpha_c = 2.1$, prior weight $\alpha_q = 0.5$, discrepancy weight $\alpha_d = 3.0$, and bias $b = 0.8$, the discrepancy decay is $\nu = 3.0$, and an edge is kept when $\omega_e^{\text{tr}} \geq \tau_\omega$ with $\tau_\omega = 0.12$. The spectral ridge in Step E is $c_0 = 10^{-4}$. The default transport model in Step F is ridge regression with regularization strength 1.0. We cluster the BioKG entity embeddings into $M = 200$ node families, and we set the bridge dimension k to twice the dimension of the source-only modality.

Table 7: PLATO cell-line results (uni-directional, pairs 1–10): Pearson R (mean ± std) for drug-mask transfer. Bold marks the best non-oracle arm per model and pair; Oracle is separated by a rule.

	BC→CH				BC→ME				BC→NS				BC→SC				CH→BC			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.081 ±.041	.089 ±.044	.000 ±.000	.000 ±.000	.153 ±.017	.164 ±.007	**.000 ±.000**	.077 ±.054	.214 ±.005	.227 ±.004	**.000 ±.000**	.160 ±.004	.214 ±.022	.224 ±.023	**.000 ±.000**	.058 ±.081	.210 ±.010	.220 ±.005	**.000 ±.000**	.095 ±.086
Bridge	.619 ±.022	.554 ±.033	.484 ±.244	.471 ±.225	.791 ±.015	.784 ±.007	.000 ±.000	.271 ±.348	.796 ±.014	.788 ±.005	.000 ±.000	**.325 ±.313**	.738 ±.039	.761 ±.007	.000 ±.000	.283 ±.311	.699 ±.043	.703 ±.009	.000 ±.000	.213 ±.320
C. Br.	.577 ±.026	**.554 ±.035**	**.534 ±.185**	**.592 ±.118**	**.796 ±.011**	**.790 ±.005**	.000 ±.000	.274 ±.350	**.798 ±.011**	**.791 ±.006**	.000 ±.000	.299 ±.322	**.747 ±.027**	**.791 ±.006**	.000 ±.000	**.293 ±.317**	.708 ±.024	**.713 ±.007**	.000 ±.000	**.218 ±.327**
Transfer	**.631 ±.020**	.278 ±.037	.235 ±.128	.293 ±.164	.706 ±.212	.734 ±.011	.000 ±.000	**.286 ±.355**	.571 ±.334	.731 ±.011	.000 ±.000	.323 ±.314	.656 ±.256	.694 ±.012	.000 ±.000	.290 ±.317	**.709 ±.024**	.657 ±.012	.000 ±.000	.213 ±.320
Oracle	.752 ±.021	.089 ±.044	.814 ±.016	.814 ±.016	.891 ±.009	.762 ±.011	.000 ±.000	.557 ±.364	.870 ±.009	.777 ±.012	.000 ±.000	.535 ±.263	.877 ±.024	.766 ±.015	.000 ±.000	.392 ±.368	.855 ±.022	.768 ±.013	.000 ±.000	.459 ±.419

	CH→ME				CH→NS				CH→SC				ME→BC				ME→CH			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.154 ±.018	.164 ±.007	**.000 ±.000**	.072 ±.050	.220 ±.008	.227 ±.004	**.000 ±.000**	.133 ±.006	.208 ±.023	.224 ±.023	**.000 ±.000**	.064 ±.088	.206 ±.027	.220 ±.005	**.000 ±.000**	.097 ±.089	.079 ±.027	.077 ±.026	**.000 ±.000**	.026 ±.038
Bridge	.819 ±.038	.817 ±.006	.000 ±.000	.251 ±.377	.782 ±.039	.795 ±.007	.000 ±.000	.282 ±.325	.778 ±.039	.778 ±.009	.000 ±.000	.239 ±.359	**.759 ±.028**	.749 ±.011	.000 ±.000	.263 ±.335	**.711 ±.017**	.625 ±.021	.000 ±.000	**.211 ±.318**
C. Br.	**.824 ±.031**	**.827 ±.004**	.000 ±.000	**.253 ±.379**	**.788 ±.024**	**.797 ±.007**	.000 ±.000	**.300 ±.325**	**.784 ±.033**	**.783 ±.007**	.000 ±.000	**.245 ±.368**	.747 ±.029	**.751 ±.011**	.000 ±.000	.228 ±.307	.610 ±.066	**.676 ±.018**	.000 ±.000	.152 ±.229
Transfer	.824 ±.031	.736 ±.014	.000 ±.000	.252 ±.378	.787 ±.026	.729 ±.036	.000 ±.000	.291 ±.323	.783 ±.035	.713 ±.015	.000 ±.000	.241 ±.362	.702 ±.131	.706 ±.015	.000 ±.000	**.267 ±.335**	.219 ±.061	.279 ±.033	.000 ±.000	.148 ±.273
Oracle	.878 ±.035	.764 ±.011	.000 ±.000	.603 ±.402	.862 ±.026	.785 ±.031	.000 ±.000	.575 ±.266	.868 ±.029	.768 ±.016	.000 ±.000	.298 ±.411	.853 ±.029	.769 ±.014	.000 ±.000	.526 ±.351	.780 ±.015	.077 ±.026	.000 ±.000	.318 ±.398

Table 8: PLATO cell-line results (uni-directional, pairs 11–20): Pearson R (mean ± std) for drug-mask transfer. Bold marks the best non-oracle arm per model and pair; Oracle is separated by a rule.

	ME→NS				ME→SC				NS→BC				NS→CH				NS→ME			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.215 ±.014	.227 ±.004	**.000 ±.000**	.120 ±.010	.214 ±.028	.224 ±.023	**.000 ±.000**	.056 ±.078	.209 ±.015	.220 ±.005	**.000 ±.000**	.104 ±.096	.074 ±.026	.077 ±.026	**.000 ±.000**	.034 ±.044	.154 ±.013	.164 ±.007	**.000 ±.000**	.078 ±.052
Bridge	.736 ±.051	.756 ±.006	.000 ±.000	.221 ±.332	.826 ±.050	.821 ±.008	.000 ±.000	.348 ±.330	.713 ±.055	.742 ±.003	.000 ±.000	.222 ±.334	.571 ±.134	.526 ±.051	.000 ±.000	.191 ±.306	.742 ±.038	.752 ±.011	.000 ±.000	.246 ±.319
C. Br.	**.750 ±.031**	**.777 ±.006**	.000 ±.000	**.226 ±.339**	**.836 ±.028**	**.824 ±.006**	.000 ±.000	.323 ±.338	**.723 ±.037**	**.745 ±.005**	.000 ±.000	**.226 ±.339**	**.657 ±.048**	**.618 ±.023**	.000 ±.000	**.233 ±.350**	**.750 ±.017**	**.759 ±.011**	.000 ±.000	.256 ±.332
Transfer	.483 ±.319	.700 ±.011	.000 ±.000	.225 ±.337	.765 ±.230	.744 ±.022	.000 ±.000	**.371 ±.326**	.303 ±.302	.707 ±.010	.000 ±.000	.225 ±.337	-.008 ±.020	.271 ±.039	.000 ±.000	.084 ±.248	.627 ±.250	.714 ±.012	.000 ±.000	**.263 ±.334**
Oracle	.854 ±.032	.776 ±.012	.000 ±.000	.768 ±.007	.879 ±.031	.764 ±.017	.000 ±.000	.439 ±.334	.844 ±.034	.769 ±.014	.000 ±.000	.463 ±.423	.779 ±.018	.077 ±.026	.000 ±.000	.343 ±.427	.887 ±.021	.764 ±.011	.000 ±.000	.560 ±.363

	NS→SC				SC→BC				SC→CH				SC→ME				SC→NS			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.217 ±.027	.224 ±.023	**.000 ±.000**	.059 ±.081	.213 ±.005	.220 ±.005	**.000 ±.000**	.088 ±.080	.073 ±.033	.077 ±.026	**.000 ±.000**	.034 ±.044	.155 ±.010	.164 ±.007	**.000 ±.000**	.092 ±.061	.193 ±.019	.227 ±.004	**.000 ±.000**	.119 ±.004
Bridge	.782 ±.043	.769 ±.009	.000 ±.000	**.356 ±.307**	.801 ±.048	.806 ±.010	.000 ±.000	.273 ±.360	.235 ±.157	.490 ±.039	.000 ±.000	.034 ±.065	.813 ±.039	.816 ±.005	.000 ±.000	.343 ±.324	.737 ±.052	.758 ±.004	.000 ±.000	.223 ±.335
C. Br.	.788 ±.023	**.769 ±.008**	.000 ±.000	.349 ±.311	**.812 ±.032**	**.814 ±.010**	.000 ±.000	.277 ±.355	**.529 ±.039**	**.492 ±.040**	.000 ±.000	.197 ±.285	**.824 ±.025**	**.822 ±.006**	.000 ±.000	.343 ±.327	**.746 ±.035**	**.765 ±.004**	.000 ±.000	.223 ±.334
Transfer	**.788 ±.026**	.724 ±.016	.000 ±.000	.348 ±.309	.529 ±.335	.751 ±.015	.000 ±.000	**.279 ±.357**	.147 ±.097	.209 ±.023	.000 ±.000	**.215 ±.324**	.351 ±.360	.738 ±.013	.000 ±.000	**.368 ±.329**	.495 ±.303	.696 ±.010	.000 ±.000	**.225 ±.337**
Oracle	.885 ±.024	.764 ±.016	.000 ±.000	.439 ±.334	.853 ±.026	.771 ±.013	.000 ±.000	.524 ±.349	.779 ±.017	.077 ±.026	.000 ±.000	.343 ±.427	.879 ±.024	.764 ±.011	.000 ±.000	.511 ±.310	.852 ±.032	.778 ±.012	.000 ±.000	.767 ±.008

Table 9: PLATO mouse PDX results (uni-directional, pairs 1–10): Pearson R (mean ± std) for drug-mask transfer. Bold marks the best non-oracle arm per model and pair; Oracle is separated by a rule.

	BR→CM				BR→CRC				BR→MN				BR→PD				CM→BR			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	**.346 ±.080**	.359 ±.076	.000 ±.000	**.052 ±.128**	.491 ±.083	.510 ±.074	.485 ±.087	.480 ±.087	.630 ±.079	**.637 ±.078**	**.580 ±.199**	**.605 ±.081**	.352 ±.101	**.364 ±.081**	**.249 ±.148**	**.272 ±.098**	.390 ±.045	.428 ±.034	.429 ±.047	.430 ±.047
Bridge	.342 ±.086	**.439 ±.089**	.025 ±.098	.000 ±.000	.536 ±.065	.514 ±.054	.450 ±.143	.506 ±.082	**.661 ±.080**	.615 ±.106	.472 ±.304	.602 ±.105	.349 ±.107	.350 ±.090	.092 ±.144	.269 ±.127	.397 ±.046	**.456 ±.055**	.430 ±.047	.439 ±.045
C. Br.	.342 ±.086	.428 ±.084	**.025 ±.098**	.046 ±.122	.531 ±.067	**.521 ±.066**	.509 ±.084	.524 ±.086	.636 ±.076	.601 ±.109	.410 ±.307	.595 ±.088	.353 ±.105	.341 ±.096	.100 ±.155	.257 ±.122	.396 ±.046	.452 ±.056	.430 ±.047	.439 ±.045
Transfer	.345 ±.082	.400 ±.077	.025 ±.098	.046 ±.122	**.540 ±.060**	.503 ±.060	**.510 ±.080**	**.533 ±.080**	.646 ±.085	.626 ±.090	.475 ±.305	.602 ±.105	**.357 ±.107**	.329 ±.093	.100 ±.155	.254 ±.118	**.400 ±.046**	.453 ±.060	**.431 ±.047**	**.439 ±.046**
Oracle	.443 ±.076	.468 ±.066	.508 ±.083	.504 ±.082	.538 ±.081	.545 ±.070	.518 ±.081	.537 ±.080	.658 ±.079	.668 ±.081	.594 ±.204	.610 ±.092	.408 ±.082	.387 ±.084	.378 ±.090	.406 ±.100	.437 ±.043	.472 ±.042	.471 ±.049	.502 ±.048

	CM→CRC				CM→MN				CM→PD				CRC→BR				CRC→CM			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.491 ±.081	**.511 ±.072**	**.473 ±.080**	.504 ±.070	.631 ±.079	.637 ±.078	**.317 ±.370**	**.546 ±.090**	.349 ±.113	.363 ±.082	**.179 ±.189**	.271 ±.089	.396 ±.037	.429 ±.034	.430 ±.059	.430 ±.052	.337 ±.080	.358 ±.076	.000 ±.000	.000 ±.000
Bridge	.496 ±.080	.489 ±.061	.442 ±.144	.497 ±.071	.636 ±.072	.644 ±.082	.085 ±.224	.328 ±.285	.356 ±.107	.366 ±.084	.091 ±.166	.279 ±.098	.472 ±.053	.503 ±.051	.486 ±.063	.508 ±.051	.373 ±.068	**.532 ±.099**	**.031 ±.119**	**.031 ±.120**
C. Br.	.496 ±.081	.492 ±.053	.444 ±.144	**.505 ±.069**	.630 ±.079	.642 ±.083	.085 ±.224	.328 ±.285	.356 ±.106	**.368 ±.083**	.123 ±.191	**.279 ±.099**	.483 ±.059	**.515 ±.054**	**.511 ±.050**	**.517 ±.051**	.327 ±.073	.460 ±.100	.000 ±.000	.015 ±.059
Transfer	**.500 ±.081**	.499 ±.053	.442 ±.144	.500 ±.068	**.637 ±.072**	**.646 ±.075**	.085 ±.224	.328 ±.285	**.363 ±.112**	.361 ±.077	.122 ±.191	.271 ±.089	**.487 ±.044**	.495 ±.055	.488 ±.063	.510 ±.051	**.380 ±.062**	.460 ±.076	.000 ±.000	.031 ±.119
Oracle	.511 ±.081	.523 ±.065	.477 ±.079	.529 ±.075	.648 ±.079	.669 ±.079	.324 ±.378	.546 ±.090	.407 ±.093	.393 ±.076	.395 ±.112	.364 ±.118	.501 ±.032	.540 ±.047	.518 ±.061	.564 ±.051	.586 ±.093	.590 ±.068	.499 ±.159	.536 ±.069

Table 10: PLATO mouse PDX results (uni-directional, pairs 11–20): Pearson R (mean ± std) for drug-mask transfer. Bold marks the best non-oracle arm per model and pair; Oracle is separated by a rule.

	CRC→MN				CRC→PD				MN→BR				MN→CM				MN→CRC			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.630 ±.079	.638 ±.078	**.615 ±.088**	.616 ±.080	.343 ±.113	.363 ±.082	.317 ±.094	.299 ±.099	.391 ±.046	.428 ±.034	.447 ±.049	.448 ±.049	.320 ±.074	.359 ±.075	**.228 ±.170**	**.253 ±.109**	.490 ±.082	.512 ±.072	.489 ±.089	.495 ±.087
Bridge	.649 ±.073	.663 ±.066	.471 ±.253	.564 ±.175	.392 ±.092	**.456 ±.061**	.340 ±.164	.397 ±.115	.398 ±.038	.484 ±.061	**.449 ±.049**	.454 ±.047	.335 ±.079	.391 ±.086	.154 ±.180	.228 ±.140	.523 ±.076	**.597 ±.059**	.518 ±.084	.532 ±.081
C. Br.	.633 ±.080	.648 ±.083	.574 ±.173	**.622 ±.080**	**.403 ±.094**	.446 ±.081	**.378 ±.127**	**.412 ±.095**	**.399 ±.039**	.467 ±.054	.445 ±.052	.449 ±.048	.312 ±.069	**.434 ±.082**	.057 ±.122	.142 ±.154	.505 ±.084	.532 ±.072	.496 ±.088	.510 ±.084
Transfer	**.653 ±.080**	**.675 ±.063**	.525 ±.228	.607 ±.075	.393 ±.091	.441 ±.075	.320 ±.185	.398 ±.115	.390 ±.044	**.489 ±.059**	.448 ±.049	**.455 ±.046**	**.337 ±.087**	.406 ±.084	.155 ±.180	.228 ±.140	**.537 ±.074**	.560 ±.065	**.519 ±.084**	**.534 ±.080**
Oracle	.675 ±.076	.680 ±.081	.646 ±.088	.653 ±.076	.481 ±.073	.526 ±.075	.464 ±.093	.505 ±.069	.483 ±.048	.569 ±.045	.529 ±.053	.575 ±.049	.456 ±.083	.469 ±.067	.487 ±.081	.498 ±.087	.557 ±.076	.598 ±.061	.531 ±.085	.573 ±.078

	MN→PD				PD→BR				PD→CM				PD→CRC				PD→MN			
	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN	SV	LG	La	EN
Raw	.346 ±.106	.364 ±.082	.207 ±.176	.268 ±.085	.390 ±.046	**.428 ±.034**	.447 ±.052	.440 ±.038	.318 ±.069	.359 ±.076	**.000 ±.000**	**.000 ±.000**	.490 ±.082	**.510 ±.074**	**.480 ±.070**	**.498 ±.061**	.630 ±.079	**.637 ±.078**	**.639 ±.057**	**.616 ±.071**
Bridge	**.385 ±.101**	**.367 ±.072**	**.311 ±.125**	**.340 ±.101**	.421 ±.049	.413 ±.068	**.476 ±.052**	.476 ±.042	.347 ±.070	.392 ±.101	.000 ±.000	.000 ±.000	.523 ±.078	.509 ±.066	.462 ±.143	.413 ±.201	.634 ±.078	.629 ±.090	.291 ±.325	.537 ±.231
C. Br.	.347 ±.100	.360 ±.069	.058 ±.126	.268 ±.085	.402 ±.045	.402 ±.068	.448 ±.051	.451 ±.043	.347 ±.070	.397 ±.100	.000 ±.000	.000 ±.000	.503 ±.083	.497 ±.069	.452 ±.142	.464 ±.145	.633 ±.078	.635 ±.086	.470 ±.297	.493 ±.263
Transfer	.375 ±.100	.338 ±.079	.143 ±.168	.268 ±.085	**.421 ±.049**	.405 ±.072	.475 ±.052	**.477 ±.043**	**.351 ±.066**	**.399 ±.091**	.000 ±.000	.000 ±.000	**.531 ±.077**	.494 ±.068	.471 ±.145	.477 ±.148	**.643 ±.078**	.624 ±.082	.039 ±.153	.397 ±.295
Oracle	.412 ±.086	.401 ±.077	.360 ±.085	.340 ±.110	.481 ±.030	.523 ±.037	.526 ±.061	.552 ±.038	.443 ±.080	.486 ±.064	.000 ±.000	.080 ±.195	.536 ±.081	.550 ±.072	.533 ±.080	.580 ±.071	.664 ±.084	.671 ±.077	.626 ±.076	.644 ±.071