TASK-CONDITIONED DUAL-GRAPH RESIDUAL WITH MARGINAL CONSTRAINED REFINEMENT FOR CIRCRNA-MIRNA INTERACTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate circRNA-miRNA interaction prediction requires integrating heterogeneous biological similarities with unevenly distributed interaction evidence. Different RNAs may rely on different similarity views, while recorded interaction networks vary substantially in endpoint connectivity. TAPER learns RNA-specific preferences over sequence, structural, and composition similarities from interaction-recovery utility estimated within each training fold. These preferences are used to construct task-conditioned biological graphs. A degree-based bipartite reference is then fitted to the training interaction network, and the standardized interaction residual is separated into graph-smooth and complementary components. The graph-smooth component is combined with a graph-regularized factor model to obtain base interaction probabilities, while the complementary component provides neighborhood evidence for candidate-specific refinement. A Bernoulli projection preserves the row and column marginals of the base predictor during this refinement. In five-fold transductive evaluation, TAPER achieves AUCs of 0.9721, 0.9827, and 0.9691 on CMI-9905, CMI-9589, and CMI-20208, respectively. Ablations on CMI-9905 show lower AUC and AUPR after each component removal. TAPER outperforms the tested mechanism alternatives, while literature evidence and expression profiles contextualize its top-ranked predictions.

1 INTRODUCTION

MicroRNAs (miRNAs) regulate gene expression through sequence-specific interactions with target transcripts (Jouravleva & Zamore, 2025). Specific circular RNAs (circRNAs) also bind miRNAs. Cdr1as/ciRS-7 provides a well-studied example, with numerous miR-7 binding sites and context-dependent effects on neuronal regulation (Cerda-Jara et al., 2026). Computational models prioritize additional circRNA-miRNA pairs for experimental testing.

Existing predictors combine recorded interactions with biological information derived from RNA sequence or structure (Guo et al., 2024; Cao et al., 2025). These similarities are commonly converted into RNA graphs before interaction prediction. Their predictive relevance can vary across individual RNAs, which makes a shared fusion rule restrictive when different nodes depend on different biological evidence. We therefore construct task-conditioned biological graphs. Interaction recovery within the training fold estimates RNA-specific preferences over sequence, structural, and composition views, and these preferences determine the similarity edges used by subsequent prediction.

Recorded interactions are unevenly distributed across RNAs, so endpoint connectivity can influence the score assigned to a candidate pair. Degree effects have been shown to affect link prediction under common evaluation protocols (Aiyappa et al., 2025). We account for this variation by fitting a degree-conditioned bipartite reference to the training network. Standardized deviations from this reference represent the residual interaction evidence. We decompose this evidence over the task-conditioned miRNA and circRNA graphs into a graph-smooth component and a complementary residual.

This decomposition provides evidence for two successive prediction stages. The graph-smooth component supports global prediction, whose probabilities establish an interaction budget for each RNA

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

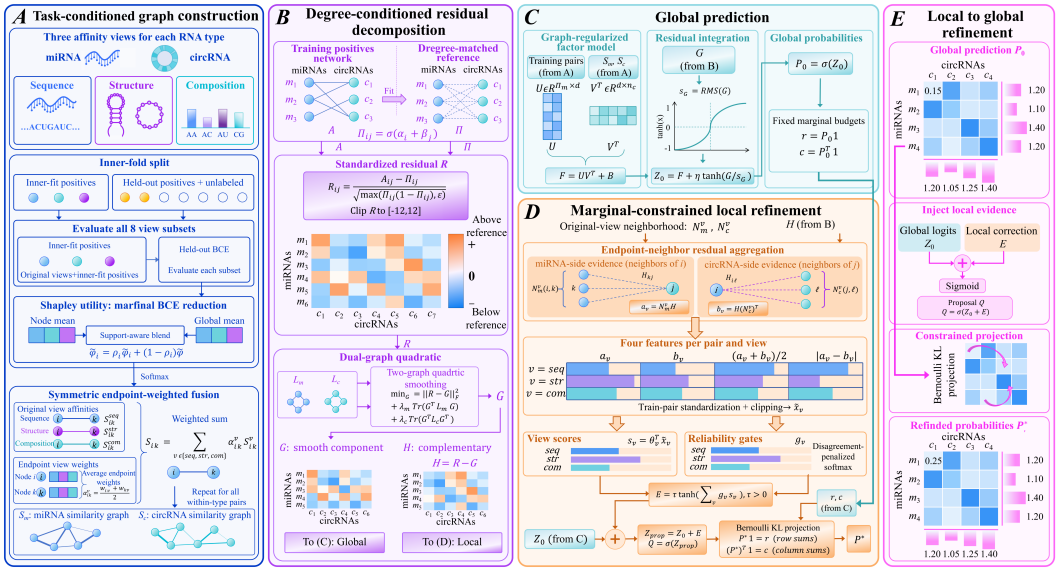

Figure 1: Overview of TAPER. **(A)** Inner-fold Shapley utility estimates RNA-specific preferences over sequence, structural, and composition similarities. Endpoint preferences determine the fused graph weights. **(B)** Training interactions are standardized against a degree-conditioned bipartite reference. Dual-graph regularization decomposes the resulting residual into graph-smooth and complementary components. **(C)** The graph-smooth component augments a graph-regularized factor model to produce base interaction probabilities, whose row and column sums define the marginal budgets. **(D)** Endpoint neighborhoods aggregate complementary residuals into pair-specific features. Disagreement-aware view gating produces local logit corrections, followed by Bernoulli marginal projection. **(E)** It demonstrates how local evidence is redistributed among candidate partners within the marginal budget of RNA levels determined by the global prediction.

through their row and column sums. Local refinement then adjusts individual pair scores using neighborhood evidence derived from the complementary residual. To retain the RNA-level budgets established by the global predictor, we constrain these adjustments to preserve its row and column marginals. Local evidence thus determines how probability mass is allocated across candidate partners. We integrate these stages into TAPER, as illustrated in Figure 1.

Our contributions are as follows:

Task-conditioned biological graph construction. We construct biological graphs from RNA-specific view utilities estimated through inner-fold interaction recovery. Shapley contributions determine node preferences, which symmetric endpoint weighting translates into graph edges.

Degree-conditioned dual-graph residual decomposition. We decompose standardized deviations from a degree-conditioned interaction reference through dual-graph smoothing with endpoint-offset penalties. The graph-smooth component supports global prediction, while the complementary residual supplies candidate-specific evidence for local refinement.

Marginal-constrained evidence reallocation. We formulate local refinement as probability reallocation under budgets established by the global predictor. Neighborhood residual evidence produces candidate-specific logit corrections, and a Bernoulli information projection enforces the global row and column marginal constraints.

2 Related Work

2.1 Biological graph construction for RNA interaction prediction

Recent circRNA-miRNA predictors increasingly combine sequence-derived information with graph-based interaction representations. BGF-CMAP integrates sequence embeddings with homogeneous

graph representations obtained from LINE and graph factorization, while DGCLCMI couples sequence feature extraction with neural graph collaborative filtering to capture interaction-dependent representations (Guo et al., 2024; Cao et al., 2025). More recent methods further enrich the relational structure used for prediction. MRHRL represents multiple higher-order relations with hypergraphs and aggregates complementary information through view attention. HCLAMCMI similarly constructs complementary hypergraphs and combines hypergraph convolution, contrastive learning, and channel attention for interaction prediction (Chen et al., 2025). ZeroStem combines graph Transformer representations with state-space modeling of meta-path subgraphs for zero-shot CMI prediction (Wei et al., 2026).

Existing methods integrate biological views at different stages of prediction. SNF combines similarity networks constructed from individual data types into a fused network (Wang et al., 2014). MRHRL applies view attention to aggregate hypergraph representations and jointly optimizes association prediction with hyperedge reconstruction (Yin et al., 2024b). TAPER estimates RNA-specific view preferences from interaction-recovery utility within the training fold and uses the preferences of both endpoints to determine similarity edge weights before residual decomposition.

2.2 Degree effects and residual modeling in link prediction

Degree-dependent behavior has been identified as an important factor in general link prediction. Aiyappa et al. (2025) show that conventional edge sampling can favor predictions correlated with node degree on general network benchmarks. This observation motivates using degrees in the observed training network to parameterize a bipartite maximum-entropy reference for interaction propensity. Saracco et al. (2015) construct ensembles whose expected degrees match the observed connectivity on both sides of a bipartite network. We use this perspective to distinguish the interaction propensity associated with recorded endpoint connectivity from residual pair-specific evidence.

Shahid et al. (2016) regularize matrix recovery over both row and column graphs, encouraging smoothness along the two dimensions. TAPER applies dual-graph regularization to standardized deviations from a degree-based interaction reference and uses the complementary residual for candidate-specific refinement.

2.3 Local evidence refinement and constrained probability modeling

Neighborhood structure provides candidate-specific features for link prediction. Zhang & Chen (2018) demonstrate that enclosing subgraphs around candidate links can support link prediction by exposing pair-specific structural patterns. Related RNA interaction models also exploit local or consistency-based information during prediction. For example, Yin et al. (2024a) reconstruct similarity networks from learned representations and apply network consistency projection to improve circRNA-miRNA association scores. These approaches illustrate the value of candidate-level structural evidence and post-prediction consistency.

Posterior regularization constrains expectations of latent-variable distributions through KL-based inference (Ganchev et al., 2010). TAPER uses a constrained probability update for local refinement. It minimizes the sum of pairwise Bernoulli KL divergences while preserving the row and column probability budgets established by the global predictor.

3 Preliminaries

3.1 Problem setting and notation

Let $\mathcal{M} = \{\mu_1, \ldots, \mu_m\}$ and $\mathcal{C} = \{c_1, \ldots, c_n\}$ denote the miRNAs and circRNAs. Matrix rows index miRNAs and columns index circRNAs. For an outer evaluation fold, let $\mathcal{P} \subseteq \mathcal{M} \times \mathcal{C}$ be the positive training pairs and $\mathcal{U}$ the permitted unlabeled training pool. The observed-positive adjacency is $A \in \{0,1\}^{m \times n}$, where $A_{ij} = 1$ exactly on $\mathcal{P}$. A zero records the absence of a positive training edge. Supervised losses use explicit masks, with validation and test pairs excluded from training supervision.

For RNA type $t \in \{m, c\}$, let $N_m = m$ and $N_c = n$. The three within-type similarity matrices are S^t_{seq}, S^t_{str}, and $S^t_{com} \in [0,1]^{N_t \times N_t}$. Prediction concerns a cross-type pair (μ_i, c_j) and returns a

probability-scale score P_{ij}. The default setting is transductive: both RNA sets and their sequence-derived similarities are available during fitting. Association-dependent quantities are constructed from the training partition. Validation selects model settings, and test labels are reserved for final evaluation.

The fused graphs have normalized Laplacians $L_m \in \mathbb{R}^{m\times m}$ and $L_c \in \mathbb{R}^{n\times n}$. For an interaction-shaped matrix X, $L_m X$ acts along miRNAs and $X L_c$ acts along circRNAs. We write $\mathbf{1}_q$ for a vector of ones and $\sigma(z) = (1 + e^{-z})^{-1}$ for the sigmoid. The row and column probability totals of P are $P\mathbf{1}_n$ and $P^\top \mathbf{1}_m$.

4 METHOD

TAPER first constructs task-conditioned RNA graphs and computes the degree-referenced residual R. Dual-graph decomposition yields the graph-smooth component G and complementary residual E. The factor logits F and a bounded transformation of G form the base logits Z^0. Neighborhood features derived from E produce a local correction D, yielding the proposal $Q = \sigma(Z^0 + D)$. Finite Bernoulli projection updates produce the final scores $P^{(T)}$ using the row and column sums of $B = \sigma(Z^0)$ as target marginals.

4.1 BIOLOGICAL SIMILARITY VIEWS

Each RNA is represented by three sequence-derived affinity views. Sequence affinity is computed from normalized weighted edit distance, structural affinity from the overlap of RNAfold-predicted base-pair structures, and composition affinity from the rank relationship between dinucleotide-frequency profiles. These views provide complementary within-type similarity information for both miRNAs and circRNAs. Their complete definitions, preprocessing rules, and boundary conventions are provided in Appendix A.2.

4.2 TASK-CONDITIONED BIOLOGICAL GRAPH CONSTRUCTION

We estimate node-specific view weights through an inner partition of $\mathcal{P}$. For each inner holdout, each view predicts held-out positives and sampled pairs from $\mathcal{U}$ by similarity-weighted propagation over the remaining positive edges. We evaluate all eight coalitions of the three views. For each evaluated pair e, the Shapley contribution ϕ_{ev} measures the average reduction in binary cross-entropy when view v is added to a coalition (Chen et al., 2023).

For an RNA node a, let N_a denote its number of evaluated pair occurrences. We shrink its mean contribution $\bar{\phi}_{av}$ toward the global mean $\bar{\phi}_v$. With shrinkage strength κ and temperature T, we obtain

$$\widetilde{\phi}_{av} = \frac{N_a \bar{\phi}_{av} + \kappa \bar{\phi}_v}{N_a + \kappa}, \qquad w_{av} = \frac{\exp(\widetilde{\phi}_{av}/T)}{\sum_{u=1}^{3} \exp(\widetilde{\phi}_{au}/T)}. \tag{1}$$

Here, κ controls shrinkage toward the global view utility, while T controls weight concentration. The normalized weight w_{av} represents node a's preference for view v. Each similarity edge combines the preferences of its two endpoints symmetrically, yielding an undirected within-type graph with RNA-specific edge weights. Appendix B details the coalition predictor and weight estimation procedure.

4.3 DEGREE-CONDITIONED DUAL-GRAPH RESIDUAL DECOMPOSITION

We model interaction propensity with a maximum-entropy Bernoulli reference whose expected row and column degrees match the training interaction graph (Saracco et al., 2015). The reference probability Π_{ij} captures the degree-driven propensity of both endpoints. We define the degree-referenced residual by standardizing the deviation of the training adjacency matrix A from this reference:

$$R_{ij} = \mathrm{clip}\left(\frac{A_{ij} - \Pi_{ij}}{\sqrt{\max\{\Pi_{ij}(1 - \Pi_{ij}), 10^{-6}\}}}, -12, 12 \right). \tag{2}$$

The Bernoulli variance normalizes deviations across reference probabilities, while the variance floor and clipping bounds stabilize residual magnitudes.

We convert the task-conditioned similarities into symmetric top-k graphs with normalized Laplacians L_m and L_c for miRNAs and circRNAs, respectively. Following dual-graph regularization (Shahid et al., 2016), we decompose the standardized residual into graph-smooth and complementary components. The graph-smooth component G minimizes

$$\mathcal{J}(G) = \frac{1}{2}\|G - R\|_F^2 + \frac{\lambda_m}{2}\operatorname{tr}(G^\top L_m G) + \frac{\lambda_c}{2}\operatorname{tr}(G L_c G^\top) \tag{3a}$$

$$+ \frac{\rho_r}{2n}\|G\mathbf{1}_n\|_2^2 + \frac{\rho_c}{2m}\|G^\top \mathbf{1}_m\|_2^2, \qquad E = R - G. \tag{3b}$$

Here, λ_m and λ_c control smoothness over the two RNA graphs, while ρ_r and ρ_c penalize row- and column-wise offsets. These penalties discourage G from reintroducing endpoint-specific effects already captured by the degree reference. For nonnegative regularization coefficients, the identity data-fitting term makes the objective strictly convex and ensures a unique minimizer. The resulting decomposition satisfies $R = G + E$, with G used for global prediction and E for candidate-specific refinement.

4.4 Global Prediction with Graph-Smooth Residual Evidence

We fit a graph-regularized logistic factor model on $\mathcal{P} \cup \mathcal{U}$, assigning weight 10 to positive examples and zero supervised weight outside these sets. The factor logits are $F_{ij} = u_i^\top v_j + b_i + c_j + b_0$. Appendix D.1 specifies the objective and graph construction. After model selection, we freeze the factor model and incorporate graph-smooth residual evidence:

$$Z^0 = F + \gamma \tanh(\operatorname{clip}(G/s_G, -8, 8)), \qquad s_G = \max\left\{\frac{\|G\|_F}{\sqrt{mn}}, 10^{-6}\right\}. \tag{4}$$

The scale s_G normalizes G by its root-mean-square magnitude, with a numerical floor, before the bounded transformation. The coefficient γ determines the residual contribution and is selected on validation data, then fixed for final evaluation. The base probabilities $B = \sigma(Z^0)$ define the row and column probability budgets for subsequent evidence reallocation.

4.5 Marginal-Constrained Local Evidence Reallocation

Global prediction sets the marginal probability budget for each RNA. Local refinement uses the complementary residual E to redistribute this budget across candidate partners. The row and column sums of the base probability matrix B serve as the target marginals.

For each biological view v, row-normalized neighborhood matrices $T_m^{(v)}$ and $T_c^{(v)}$ aggregate complementary residuals from the two RNA endpoints: $H^{(v)} = T_m^{(v)} E$ and $K^{(v)} = E(T_c^{(v)})^\top$. We construct the pair features

$$x_{ij}^{(v)} = \left[H_{ij}^{(v)}, K_{ij}^{(v)}, \frac{H_{ij}^{(v)} + K_{ij}^{(v)}}{2}, \left|H_{ij}^{(v)} - K_{ij}^{(v)}\right|\right]. \tag{5}$$

We standardize each feature using training pairs only and clip the resulting values to $[-12, 12]$, denoting them by $\widehat{x}_{ij}^{(v)}$. A candidate-specific gate weights the view-specific linear scores according to standardized endpoint disagreement:

$$\omega_{ijv} = \operatorname{softmax}_v\left(a_v - \operatorname{softplus}(r_v)\operatorname{softplus}\left(\widehat{x}_{ij,4}^{(v)}\right)\right), \tag{6a}$$

$$D_{ij} = \tau \tanh\left(\sum_{v=1}^{3} \omega_{ijv}\theta_v^\top \widehat{x}_{ij}^{(v)}\right), \qquad \tau = \operatorname{softplus}(s) + 10^{-4}. \tag{6b}$$

Here, a_v is a learned view bias, and $\mathrm{softplus}(r_v)$ determines the nonnegative disagreement penalty. The vector θ_v maps endpoint features to a local evidence score, while s parameterizes the positive scale τ. Holding other gate inputs fixed, greater endpoint disagreement decreases a view's normalized weight ω_{ijv}. The bounded correction D updates the base logits to $Z_{\mathrm{prop}} = Z^0 + D$, yielding the proposal $Q = \sigma(Z_{\mathrm{prop}})$.

The local correction D can change the row and column sums of Q relative to B. We define the refined probabilities through a Bernoulli information projection with target marginals given by B:

$$P^\star = \operatorname*{argmin}_{P \in \mathcal{C}(B)} \sum_{i,j} \mathrm{KL}(\mathrm{Ber}(P_{ij}) \,\|\, \mathrm{Ber}(Q_{ij}))\,. \tag{7}$$

The feasible set $\mathcal{C}(B)$ contains probability matrices whose row and column sums equal those of B. The projection therefore adjusts pairwise probabilities subject to the RNA-level budgets established by global prediction. Its solution takes the form

$$P^\star_{ij} = \sigma(Z_{\mathrm{prop},ij} + \eta_i + \xi_j)\,, \tag{8}$$

where η_i and ξ_j are dual offsets for the row and column constraints. We approximate these offsets through a fixed number of alternating Newton updates. The resulting probabilities are denoted by $P^{(T)}$, with marginal errors evaluated over the complete RNA-pair grid as defined in Appendix E.2.

Local evidence parameters are trained on positive pairs and a fixed sampled unlabeled set using binary cross-entropy with an evidence regularizer. Gradients propagate through the finite projection updates, while the factor representation and residual decomposition remain fixed. Appendix F.1 specifies the objective and staged training procedure.

5 Experiments

5.1 Experimental Setup

We evaluate TAPER on three circRNA-miRNA interaction datasets. CMI-9905, compiled by Wang et al. (2022), contains 9,905 recorded interactions between 962 miRNAs and 2,346 circRNAs. CMI-9589 is sourced from circBank (Liu et al., 2019) and contains 9,589 interactions between 821 miRNAs and 2,115 circRNAs. CMI-20208 contains 20,208 interactions between 1,152 miRNAs and 3,569 circRNAs (Cao et al., 2025).

We use five-fold pair-level cross-validation in the transductive setting. Within each outer training partition, 10% of the training positives are reserved for validation. All association-dependent quantities are rebuilt from the current training partition, including the task-conditioned view weights, degree-conditioned reference, residual decomposition, and graph-regularized factor model. Validation data select model checkpoints, the global residual scale, and the classification threshold. Test pairs are used only for final evaluation.

5.2 Comparison with Existing Methods

TAPER achieves the highest reported AUC, AUPR, and ACC among the listed methods on all three benchmarks. Table 1 summarizes results on CMI-9905, CMI-9589, and CMI-20208. All baseline values are taken from their original publications and reflect the evaluation protocols reported in those studies. TAPER results follow the five-fold protocol described above. Baseline method references are listed in Appendix G.

On CMI-9905, TAPER obtains an AUC of 0.9721 ± 0.0023 and an AUPR of 0.9794 ± 0.0014, exceeding the highest baseline values reported by DGCLCMI and CMAGN by 0.0076 and 0.0082, respectively. Its ACC reaches 93.23%, compared with 90.84% for MFERL, a difference of 2.39 percentage points. On CMI-9589, TAPER achieves 0.9827 ± 0.0012 AUC and 0.9860 ± 0.0007 AUPR, exceeding the reported CMAGN values by 0.0090 and 0.0084. Its ACC of 94.09% exceeds MFERL by 2.39 percentage points.

On CMI-20208, TAPER reaches 0.9691 ± 0.0020 AUC and 0.9732 ± 0.0020 AUPR. These values exceed the highest reported baseline values, DGCLCMI for AUC and LSAGHI for AUPR, by 0.0145 and 0.0253. Its ACC reaches 91.54%, exceeding LSAGHI by 2.32 percentage points.

Table 1: Comparison on three circRNA-miRNA interaction benchmarks. Baseline values are taken from the original publications; TAPER values are means from our five-fold evaluation. Bold and underlined values indicate the best and second-best reported results in each column, respectively. A dash indicates an unavailable result.

Method	CMI-9905			CMI-9589			CMI-20208		
	AUC	AUPR	ACC	AUC	AUPR	ACC	AUC	AUPR	ACC
BioDGW-CMI	0.9026	0.8962	0.8325	0.9476	0.9416	0.8851	—	—	—
iHofman	0.9095	0.9234	0.8359	0.9187	0.9269	0.8376	0.9173	0.9211	0.8392
BEROLECMI	0.9104	0.9086	0.8395	0.9491	0.9431	0.8777	—	—	—
Hither-CMI	0.9134	0.9127	0.8374	0.9517	0.9469	0.8743	0.9182	0.9164	0.8421
MRHRL	0.9276	0.9396	0.8851	0.9404	0.9447	0.8753	0.9264	0.9297	0.8656
MuGNet-CMI	0.9248	0.9278	0.8524	0.9568	0.9577	0.8927	0.9221	0.9187	0.8499
LSAGHI	0.9405	0.9435	0.8824	0.9552	0.9520	0.8988	0.9522	0.9479	0.8922
MFERL	0.9596	0.9610	0.9084	0.9669	0.9629	0.9170	0.9456	0.9443	0.8789
DGCLCMI	0.9645	0.9546	0.8719	0.9610	0.9501	0.8596	0.9546	0.9415	0.8417
CMAGN	0.9611	0.9712	—	0.9737	0.9776	—	—	—	—
TAPER	**0.9721**	**0.9794**	**0.9323**	**0.9827**	**0.9860**	**0.9409**	**0.9691**	**0.9732**	**0.9154**

5.3 Ablation Study

Removing the dual-graph residual decomposition produces the largest AUC decrease among the evaluated ablations. On CMI-9905, AUC falls from 0.9721 to 0.9383 and MCC from 86.63% to 76.26%. Table 2 evaluates each removal under the same five-fold protocol. Removing task-conditioned graph construction, local evidence reallocation, and Bernoulli marginal projection yields AUCs of 0.9486, 0.9520, and 0.9430, respectively. All four removals also reduce mean AUPR relative to the full model.

Table 2: Ablation study on CMI-9905.

Configuration	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
w/o task-conditioned graph construction	89.10 ± 0.31	86.01 ± 1.58	92.18 ± 1.05	90.77 ± 1.00	78.52 ± 0.59	0.9486 ± 0.0041	0.9566 ± 0.0030
w/o dual-graph residual decomposition	87.92 ± 0.67	84.34 ± 1.58	91.49 ± 1.69	89.72 ± 1.67	76.26 ± 1.35	0.9383 ± 0.0042	0.9467 ± 0.0038
w/o local evidence reallocation	89.82 ± 0.56	86.75 ± 1.41	92.89 ± 1.75	91.10 ± 1.83	80.07 ± 1.14	0.9520 ± 0.0035	0.9579 ± 0.0023
w/o Bernoulli marginal projection	88.22 ± 0.47	85.93 ± 2.60	90.50 ± 2.67	88.41 ± 2.64	76.71 ± 0.95	0.9430 ± 0.0040	0.9490 ± 0.0031
Full TAPER	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**

5.4 Mechanism-Level Comparisons

TAPER achieves the highest mean ACC, AUC, and AUPR in all three mechanism comparisons. Figure 2 evaluates one-component replacements on CMI-9905 under the same five-fold protocol, keeping the remaining pipeline fixed. TAPER obtains 93.23% ACC, 0.9721 AUC, and 0.9794 AUPR. Full seven-metric results, standard deviations, and references appear in Appendix G.3.

For residual modeling, we compare the degree-conditioned dual-graph formulation with GFHF, LGC, APPNP, and GDC. LGC is the strongest alternative, reaching 89.80% ACC, 0.9517 AUC, and 0.9600 AUPR. TAPER gains 3.43 percentage points in ACC and 0.0204 in AUC.

For view fusion, we compare RNA-specific weighting learned from interaction-recovery utility with SNF, ANF, NEMO, and UMKL using the same biological views. UMKL performs best among these alternatives, with 89.08% ACC, 0.9473 AUC, and 0.9561 AUPR. TAPER gains 4.15 percentage points in ACC and 0.0248 in AUC.

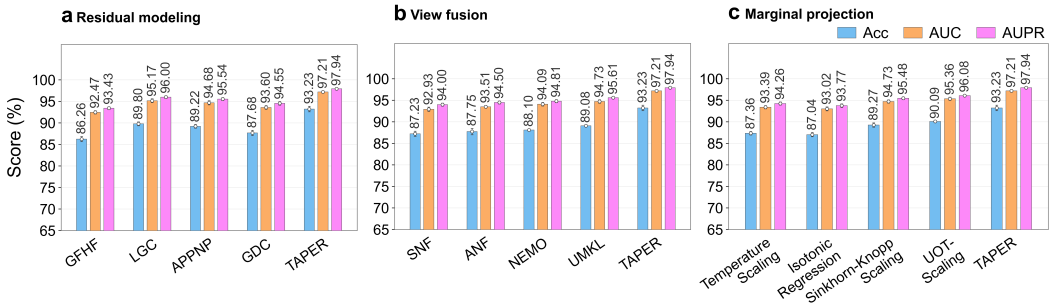

Figure 2: Mechanism-level comparisons on CMI-9905. (a) Degree-conditioned dual-graph residual modeling compared with GFHF, LGC, APPNP, and GDC. (b) Task-conditioned biological graph construction compared with SNF, ANF, NEMO, and UMKL. (c) Bernoulli marginal projection compared with temperature scaling, isotonic regression, Sinkhorn–Knopp scaling, and UOT-Scaling.

For local refinement, we compare Bernoulli projection with temperature scaling, isotonic regression, Sinkhorn–Knopp scaling, and UOT-Scaling. UOT-Scaling is the strongest alternative, reaching 90.09% ACC, 0.9536 AUC, and 0.9608 AUPR. TAPER improves these scores by 3.14 percentage points, 0.0185, and 0.0186, respectively.

5.5 Case Studies

We examine TAPER's top 20 circRNA-miRNA predictions through literature review and public expression analysis. P01–P20 follow descending prediction scores. Table 3 provides supporting publications for 19 pairs; P14 remains unconfirmed. GSE159980 and GSE159979 provide circRNA and miRNA expression profiles, respectively, for MCF-7, LCC2, and LCC9 breast cancer cells, with three biological replicates per cell line (Gao et al., 2020). The heatmaps in Figure 3 reveal distinct expression patterns across these cellular states. Row-standardized colors show relative variation within each RNA, while the adjacent contrasts quantify mean $\log_2$ expression differences from MCF-7. Panel c links the independently clustered profiles to the candidate pairs in Table 3.

Table 3: Top 20 circRNA-miRNA pairs predicted by TAPER and associated PubMed references.

ID	circRNA	miRNA	PMID	ID	circRNA	miRNA	PMID
P01	circ_0081065	miR-665	38195914	P11	circ_0009910	miR-145-5p	32588730
P02	circ_0002111	miR-432-5p	42020488	P12	circ_0014130	miR-376c-3p	32190005
P03	circ_0008305	miR-379-5p	39779892	P13	circ_0004771	miR-339-5p	32050790
P04	circ_0008234	miR-574-5p	34050132	P14	circ_0010653	miR-6860	—
P05	circ_0079929	miR-1184	39342367	P15	circ_0013561	miR-7-5p	38429830
P06	circ_0001946	miR-514a-3p	42561881	P16	circ_0044529	miR-4640-5p	30282067
P07	circ_0015004	miR-127-3p	33468140	P17	circ_0005729	miR-1301-3p	41104169
P08	circ_0006522	miR-769-3p	41158215	P18	circ_0001649	miR-4688	31137016
P09	circ_0093996	miR-494-3p	33108919	P19	circ_0000081	miR-423-5p	35302432
P10	circ_0004015	miR-1183	30509491	P20	circ_0025202	miR-182-5p	31153828

Published studies report the P13 and P20 interactions (Huang et al., 2020; Sang et al., 2019). The expression profiles in Figure 3 describe their cellular context. P13 illustrates opposing expression changes within a literature-supported pair. In esophageal squamous cell carcinoma, hsa_circ_0004771 was reported to regulate CDC25A through miR-339-5p, with rescue experiments supporting this regulatory relationship (Huang et al., 2020). In the breast cancer profiles examined here, hsa_circ_0004771 has higher mean expression in LCC9 than in MCF-7, with a $\log_2$ difference of $+0.674$, whereas miR-339-5p has lower mean expression, with a difference of -0.476.

The heatmaps therefore place this pair in opposing expression trajectories across the MCF-7–LCC9 comparison, extending its descriptive expression context to a different cancer type.

P20 illustrates concordant changes between interacting partners. Published experiments linked hsa_circ_0025202 and miR-182-5p to FOXO3a regulation and tamoxifen sensitivity in breast cancer (Sang et al., 2019). Both RNAs show lower mean expression in LCC9 than in MCF-7, indicating a coordinated decrease at the cell-line level. Together, P13 and P20 show that literature-supported pairs can exhibit either opposing or concordant expression changes across the same cellular contrast.

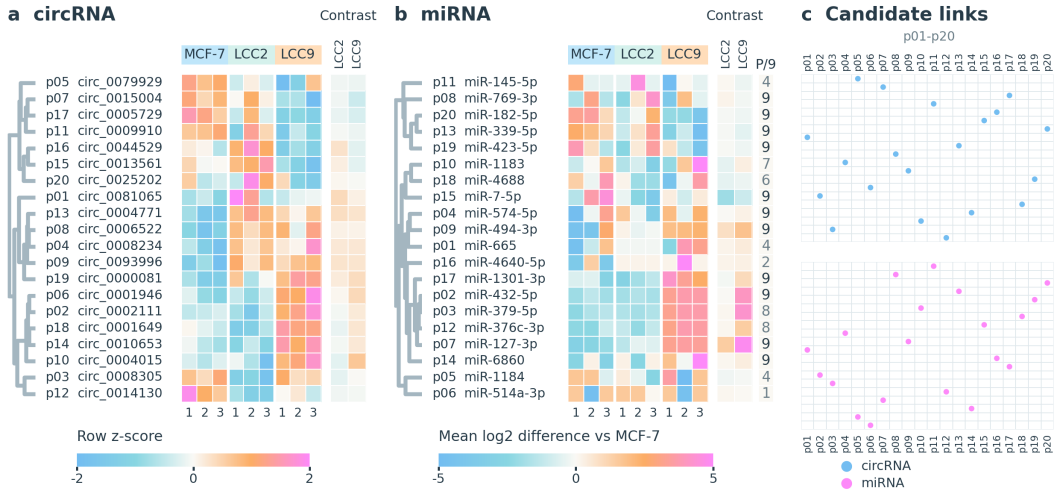

Figure 3: Expression context of TAPER's top 20 predicted circRNA-miRNA pairs. (a) Clustered circRNA expression profiles across MCF-7, LCC2, and LCC9 cells, with three biological replicates per cell line. (b) Clustered miRNA expression profiles across the same cell states. P/9 indicates the number of samples with a Present call among the nine samples. Both heatmaps show row-wise z-scores of $\log_2$ expression. Adjacent contrast columns report mean $\log_2$ expression differences for LCC2 and LCC9 relative to MCF-7. (c) Pair-membership matrices linking the independently clustered RNA profiles to predictions P01–P20. The upper and lower row orders follow panels a and b, respectively.

6 Conclusion

We presented TAPER for circRNA-miRNA interaction prediction. The model estimates RNA-specific view preferences from training-fold interaction-recovery utility to construct biological graphs. It then decomposes degree-referenced interaction residuals into graph-smooth and complementary components. These components support global prediction and candidate-specific refinement, with Bernoulli projection constraining local updates to the marginal budgets established by the global predictor. TAPER achieves the highest reported AUC, AUPR, and ACC among the compared methods on CMI-9905, CMI-9589, and CMI-20208. On CMI-9905, each evaluated component removal lowers mean AUC and AUPR, with the largest AUC decrease following removal of the dual-graph residual decomposition. Mechanism comparisons further favor the proposed formulations over the evaluated alternatives.

Future work will address TAPER's reliance on predefined RNA sets and fixed biological similarities by learning sequence-based representations for newly introduced RNAs. A further limitation is that local refinement treats the global predictor's marginal budgets as fixed, so errors in these budgets may constrain pairwise corrections. Incorporating uncertainty into the marginal constraints could make refinement more responsive to the reliability of global predictions. Experimental validation remains necessary to establish the biological roles of newly predicted interactions.

AI USE STATEMENT

Generative AI was used solely to check the formatting of the manuscript text and polish its language. All scientific content was developed and verified by the authors, who take full responsibility for the manuscript.

REFERENCES

Rachith Aiyappa, Xin Wang, Munjung Kim, Ozgur Can Seckin, Yong-Yeol Ahn, and Sadamori Kojaku. Implicit degree bias in the link prediction task. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 874–908. PMLR, 2025. URL https://proceedings.mlr.press/v267/aiyappa25a.html.

Chao Cao, Mengli Li, Chunyu Wang, Lei Xu, Quan Zou, Yansu Wang, and Wu Han. DGCLCMI: a deep graph collaboration learning method to predict circRNA-miRNA interactions. *BMC Biology*, 23(1):104, 2025. doi: 10.1186/s12915-025-02197-9.

Cledi A. Cerda-Jara, Flavia Scoyni, Grygoriy Zolotarov, Poojashree Bhaskar, Daniel León-Periñán, Jørgen Kjems, and Nikolaus Rajewsky. The history and function of a circular RNA. *Nature Communications*, 17(1):3862, 2026. doi: 10.1038/s41467-026-71822-0.

Hugh Chen, Ian C. Covert, Scott M. Lundberg, and Su-In Lee. Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence*, 5(6):590–601, 2023. doi: 10.1038/s42256-023-00657-x.

Lei Chen, Ying Chen, and Bo Zhou. HCLAMCMI: Prediction of circRNA–miRNA Interactions Based on Hypergraph Contrastive Learning and an Attention Mechanism. *Journal of Chemical Information and Modeling*, 65(21):12099–12115, 2025. doi: 10.1021/acs.jcim.5c01968.

Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018. doi: 10.1090/mcom/3303.

Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Kuzman Ganchev, João Graça, Jennifer Gillenwater, and Ben Taskar. Posterior Regularization for Structured Latent Variable Models. *Journal of Machine Learning Research*, 11(67):2001–2049, 2010. URL https://www.jmlr.org/papers/v11/ganchev10a.html.

Liang Gao, Kunwei Shen, Ni Yin, and Min Jiang. Comprehensive Transcriptomic Analysis Reveals Dysregulated Competing Endogenous RNA Network in Endocrine Resistant Breast Cancer Cells. *Frontiers in Oncology*, 10:600487, 2020. doi: 10.3389/fonc.2020.600487.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then Propagate: Graph Neural Networks meet Personalized PageRank. In *International Conference on Learning Representations*, 2019a. URL https://arxiv.org/abs/1810.05997.

Johannes Gasteiger, Stefan Weißenberger, and Stephan Günnemann. Diffusion Improves Graph Learning. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019b. URL https://proceedings.neurips.cc/paper/2019/hash/23c894276a2c5a16470e6a31f4618d73-Abstract.html.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Lu-Xiang Guo, Lei Wang, Zhu-Hong You, Chang-Qing Yu, Meng-Lei Hu, Bo-Wei Zhao, and Yang Li. Biolinguistic graph fusion model for circRNA–miRNA association prediction. *Briefings in Bioinformatics*, 25(2):bbae058, 2024. doi: 10.1093/bib/bbae058.

Enmin Huang, Junhui Fu, Qiuyan Yu, Peixin Xie, Zhongxian Yang, Huanlin Ji, Li Wang, Ganfeng Luo, Yanting Zhang, and Ke Li. CircRNA hsa_circ_0004771 promotes esophageal squamous cell cancer progression via miR-339-5p/CDC25A axis. *Epigenomics*, 12(7):587–603, 2020. doi: 10.2217/epi-2019-0404. PMID: 32050790.

Chen Jiang, Lei Wang, Chang-Qing Yu, Zhu-Hong You, Xin-Fei Wang, Meng-Meng Wei, Tai-Long Shi, Si-Zhe Liang, and Deng-Wu Wang. Hither-CMI: Prediction of circRNA-miRNA Interactions Based on a Hybrid Multimodal Network and Higher-Order Neighborhood Information via a Graph Convolutional Network. *Journal of Chemical Information and Modeling*, 65(1):446–459, 2025. doi: 10.1021/acs.jcim.4c01991.

Chen Jiang, Lei Wang, Changqing Yu, Zhuhong You, Xinfei Wang, Mengmeng Wei, and Mianshuo Lu. MuGNet-CMI: Multi-Head Hybrid Graph Neural Network for Predicting circRNA-miRNA Interactions With Global High-Order and Local Low-Order Information. *IEEE Transactions on Big Data*, 12(1):159–173, 2026. doi: 10.1109/tbdata.2025.3604175.

Karina Jouravleva and Phillip D. Zamore. A guide to the biogenesis and functions of endogenous small non-coding RNAs in animals. *Nature Reviews Molecular Cell Biology*, 26(5):347–370, 2025. doi: 10.1038/s41580-024-00818-9.

Ming Liu, Qian Wang, Jian Shen, Burton B. Yang, and Xiangming Ding. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biology*, 16(7):899–905, 2019. doi: 10.1080/15476286.2019.1600395.

Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1):26, 2011. doi: 10.1186/1748-7188-6-26.

Tianle Ma and Aidong Zhang. Affinity network fusion and semi-supervised learning for cancer patient clustering. *Methods*, 145:16–24, 2018. doi: 10.1016/j.ymeth.2018.05.020.

Jérôme Mariette and Nathalie Villa-Vialaneix. Unsupervised multiple kernel learning for heterogeneous data integration. *Bioinformatics*, 34(6):1009–1015, 2018. doi: 10.1093/bioinformatics/btx682.

Ya-Ru Miao, Wei Liu, Qiong Zhang, and An-Yuan Guo. lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Research*, 46(D1): D276–D280, 2018. doi: 10.1093/nar/gkx1004.

Li Peng, Wang Wang, Zongyi Yang, Xiangzheng Fu, Wei Liang, and Dongsheng Cao. Leveraging explainable multi-scale features for fine-grained circRNA-miRNA interaction prediction. *BMC Biology*, 23(1):121, 2025. doi: 10.1186/s12915-025-02227-6.

Nimrod Rappoport and Ron Shamir. NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*, 35(18):3348–3356, 2019. doi: 10.1093/bioinformatics/btz058.

Yuting Sang, Bing Chen, Xiaojin Song, Yaming Li, Yiran Liang, Dianwen Han, Ning Zhang, Hanwen Zhang, Ying Liu, Tong Chen, Chen Li, Lijuan Wang, Wenjing Zhao, and Qifeng Yang. circRNA_0025202 Regulates Tamoxifen Sensitivity and Tumor Progression via Regulating the miR-182-5p/FOXO3a Axis in Breast Cancer. *Molecular Therapy*, 27(9):1638–1652, 2019. doi: 10.1016/j.ymthe.2019.05.011. PMID: 31153828.

Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini. Randomizing bipartite networks: the case of the World Trade Web. *Scientific Reports*, 5(1):10595, 2015. doi: 10.1038/srep10595.

Nauman Shahid, Nathanael Perraudin, Vassilis Kalofolias, Gilles Puy, and Pierre Vandergheynst. Fast Robust PCA on Graphs. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):740–756, 2016. doi: 10.1109/jstsp.2016.2555239.

Bo Wang, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains, and Anna Goldenberg. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, 11(3):333–337, 2014. doi: 10.1038/nmeth.2810.

Xin-Fei Wang, Chang-Qing Yu, Li-Ping Li, Zhu-Hong You, Wen-Zhun Huang, Yue-Chao Li, Zhong-Hao Ren, and Yong-Jian Guan. KGDCMI: A New Approach for Predicting circRNA–miRNA Interactions From Multi-Source Information Extraction and Deep Learning. *Frontiers in Genetics*, 13:958096, 2022. doi: 10.3389/fgene.2022.958096.

Xin-Fei Wang, Chang-Qing Yu, Zhu-Hong You, Yan Qiao, Zheng-Wei Li, and Wen-Zhun Huang. An efficient circRNA-miRNA interaction prediction model by combining biological text mining and wavelet diffusion-based sparse network structure embedding. *Computers in Biology and Medicine*, 165:107421, 2023. doi: 10.1016/j.compbiomed.2023.107421.

Xin-Fei Wang, Chang-Qing Yu, Zhu-Hong You, Yan Wang, Lan Huang, Yan Qiao, Lei Wang, and Zheng-Wei Li. BEROLECMI: a novel prediction method to infer circRNA-miRNA interaction from the role definition of molecular attributes and biological networks. *BMC Bioinformatics*, 25 (1):264, 2024. doi: 10.1186/s12859-024-05891-7.

Mengmeng Wei, Lei Wang, Zhu-Hong You, Pengwei Hu, Bowei Zhao, Zhi-An Huang, Yu-An Huang, and Haicheng Yi. Dual-Channel Learning Framework for Zero-Shot CircRNA-miRNA Interaction Prediction via State Space Modeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(2):1258–1266, 2026. doi: 10.1609/aaai.v40i2.37098.

Anhui Yin, Lei Chen, Bo Zhou, and Yu-Dong Cai. CMAGN: circRNA–miRNA association prediction based on graph attention auto-encoder and network consistency projection. *BMC Bioinformatics*, 25(1):336, 2024a. doi: 10.1186/s12859-024-05959-4.

Wenjing Yin, Shudong Wang, Yuanyuan Zhang, Sibo Qiao, Wenhao Wu, and Hengxiao Li. Multirelational Hypergraph Representation Learning for Predicting circRNA-miRNA Associations. *Journal of Chemical Information and Modeling*, 64(21):8349–8360, 2024b. doi: 10.1021/acs.jcim.4c01436.

Wenjing Yin, Shudong Wang, Yuanyuan Zhang, Sibo Qiao, Shuqiang Wang, Fazlullah Khan, Ryan Altuki, and Zhihan Lyu. Autonomously Adjusting Multi-Relational Hypergraphs Structure for Predicting circRNA-MiRNA Associations. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–10, 2025. doi: 10.1109/jbhi.2025.3531427.

Chang-Qing Yu, Chen Jiang, Lei Wang, Zhu-Hong You, Xin-Fei Wang, Meng-Meng Wei, Tai-Long Shi, and Si-Zhe Liang. iHofman: a predictive model integrating high-order and low-order features with weighted attention mechanisms for circRNA-miRNA interactions. *BMC Biology*, 23(1):162, 2025. doi: 10.1186/s12915-025-02260-5.

Bianca Zadrozny and Charles Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 694–699, 2002. doi: 10.1145/775047.775151.

Muhan Zhang and Yixin Chen. Link Prediction Based on Graph Neural Networks. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/hash/53f0d7c537d99b3824f0f99d62ea2428-Abstract.html.

Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with Local and Global Consistency. In *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/hash/87682805257e619d49b8e0dfdc14affa-Abstract.html.

Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 912–919, 2003. URL https://www.cs.cmu.edu/~zhuxj/pub/zgl.pdf.

A Problem setting and biological inputs

A.1 Pair partitions and notation

Let $\mathcal{M} = \{\mu_1, \ldots, \mu_m\}$ and $\mathcal{C} = \{c_1, \ldots, c_n\}$ be the miRNA and circRNA sets. Rows index miRNAs and columns index circRNAs. For a fixed outer fold, $\mathcal{P}_{\mathrm{tr}}$ contains the permitted training positives and $\mathcal{U}_{\mathrm{tr}}$ contains the permitted unlabeled training pairs. The observed-positive matrix is

$$A_{ij} = \mathbf{1}\{(\mu_i, c_j) \in \mathcal{P}_{\mathrm{tr}}\}. \tag{A.1}$$

A zero indicates the absence of an observed training positive. Its contribution to supervised learning is determined by an explicit training mask. Unlabeled pairs receive label zero in the default binary cross-entropy objective. Label zero denotes an unlabeled pair under the sampling protocol. Validation and test pairs are excluded from training supervision. Quantities derived from associations use the training-positive matrix and the permitted training pool.

The default pairwise protocol uses five outer folds. Positive and unlabeled pairs are independently partitioned into five chunks. One chunk from each group supplies the outer test pool. Ten percent of the remaining positives are reserved for validation. A disjoint validation unlabeled pool is also reserved. Unused pairs from a reserved pool remain outside training. Validation controls checkpoint selection, the global evidence scale, and the classification threshold. Test labels enter only the final metric computation.

The default setting is transductive: the fixed RNA universe and its sequence-derived similarities are available during fitting. A node-held-out interaction split can retain this universe and its side information while withholding the node's training interactions. This evaluates missing interaction histories within the predefined RNA universe.

For RNA type $t \in \{m, c\}$, let $N_m = m$ and $N_c = n$. Each type has sequence, structural, and composition affinities $S_t^{\mathrm{seq}}, S_t^{\mathrm{str}}, S_t^{\mathrm{com}}$. We use $v \in \{1, 2, 3\}$ to index these views. All interaction-shaped quantities have dimension $m \times n$. The degree reference is Π; its standardized residual is $R = G + E$. The matrix F contains factor logits, and Z^0 contains the selected base logits. The base probabilities are $B = \sigma(Z^0)$. Local evidence D produces proposal logits Z^{prop} and probabilities Q. The exact constrained probabilities are $P^\star$; finite iterations produce $P^{(T)}$. The sigmoid is σ, $\mathbf{1}_q$ is a length-q vector of ones, and $\langle X, Y \rangle_F = \mathrm{tr}(X^\top Y)$.

A.2 Similarity definitions and preprocessing

For RNA sequence r_a^t of length ℓ_a^t, sequence similarity uses weighted edit distance with insertion and deletion costs one and substitution cost two. The Wagner–Fischer dynamic program starts with $D_{i,0} = i$ and $D_{0,j} = j$ and uses

$$\begin{aligned} D_{i,j} = \min\big\{ & D_{i-1,j} + 1,\ D_{i,j-1} + 1, \\ & D_{i-1,j-1} + 2\mathbf{1}[r_a^t(i) \neq r_b^t(j)] \big\}. \end{aligned} \tag{A.2}$$

For nonempty sequences, the resulting affinity is

$$S_t^{\mathrm{seq}}(a, b) = 1 - \frac{D_{\mathrm{Lev}}(r_a^t, r_b^t)}{\ell_a^t + \ell_b^t}. \tag{A.3}$$

Structural affinity is defined from RNAfold predictions in the ViennaRNA Package (Lorenz et al., 2011). Let b_a^t and b_b^t be the predicted base-pair counts and h_{ab}^t the count matched under the preprocessing coordinate mapping. For a positive denominator,

$$S_t^{\mathrm{str}}(a, b) = \frac{2h_{ab}^t}{b_a^t + b_b^t}. \tag{A.4}$$

Composition affinity uses the 16-dimensional dinucleotide-frequency vector over $\{A, C, G, U\}$. Let $d^t_{a,k}$ be the average rank of its kth component, with equal frequencies assigned their average rank. Set $\bar{d}^t_a = 16^{-1} \sum_{k=1}^{16} d^t_{a,k}$. For nonconstant ranked profiles,

$$S^{\mathrm{com}}_t(a, b) = \left| \frac{\sum_{k=1}^{16} (d^t_{a,k} - \bar{d}^t_a)(d^t_{b,k} - \bar{d}^t_b)}{\sqrt{\sum_{k=1}^{16} (d^t_{a,k} - \bar{d}^t_a)^2} \sqrt{\sum_{k=1}^{16} (d^t_{b,k} - \bar{d}^t_b)^2}} \right|. \tag{A.5}$$

The absolute correlation assigns equal affinity to rank relationships of equal magnitude. Reversed and matching rank orders can therefore have the same affinity. The average-rank definition handles tied frequencies explicitly.

The six affinity matrices are constructed independently of association labels. The loader checks squareness, finiteness, and consistent RNA identifiers. It symmetrizes each matrix as $(S + S^\top)/2$ and sets its diagonal to one. Graph routines subsequently remove this diagonal. The third view is consistently called composition similarity; the suffix exp in existing CSV filenames is a historical identifier. The nominal similarities lie in $[0, 1]$ and serve as graph weights.

B Task-conditioned biological graph construction

B.1 Inner evaluation and marginal view utility

Graph construction uses the three biological views and excludes relation-derived views in the default configuration. Let $K = \min(3, |\mathcal{P}_{\mathrm{tr}}|)$, with at least two training positives. For inner chunk h, remove its held-out positives to obtain $A^{(-h)}$. Evaluate those positives together with an unlabeled block drawn only from $\mathcal{U}_{\mathrm{tr}}$. These evaluation records are occurrences: the same unlabeled pair can be evaluated in multiple chunks.

For each type and view, clean the affinity by removing negative weights, symmetrizing, removing the diagonal, and dividing by its maximum with denominator floor 10^{-8}. Retain the largest $\min(k_g, N_t - 1)$ entries per row, where $k_g = 10$. If $\bar{S}^{(v)}_t$ denotes the retained matrix, the transition is

$$T^{(v)}_t(a, b) = \frac{\bar{S}^{(v)}_t(a, b)}{\max\{\sum_{b'} \bar{S}^{(v)}_t(a, b'), 10^{-8}\}}. \tag{B.1}$$

Zero rows remain zero. Define $\delta_h = \mathrm{clip}(\mathrm{mean}(A^{(-h)}), 10^{-5}, 1 - 10^{-5})$. The probability assigned by view v to an evaluated pair e is

$$p^{(v,h)}_e = \mathrm{clip}\left(\frac{[T^{(v)}_m A^{(-h)}]_e + [A^{(-h)}(T^{(v)}_c)^\top]_e}{4} + \frac{\delta_h}{2}, 10^{-5}, 1 - 10^{-5} \right). \tag{B.2}$$

All eight coalitions of the view set $\mathcal{V}_b = \{1, 2, 3\}$ are evaluated. A nonempty coalition S uses $p^{S,h}_e = |S|^{-1} \sum_{v \in S} p^{(v,h)}_e$; the empty coalition uses δ_h. With the evaluation label y_e, define

$$\ell_e(S) = -y_e \log p^{S,h}_e - (1 - y_e) \log(1 - p^{S,h}_e). \tag{B.3}$$

The Shapley contribution of a view is its weighted reduction in this loss:

$$\phi_{ev} = \sum_{S \subseteq \mathcal{V}_b \setminus \{v\}} \frac{|S|!(2 - |S|)!}{3!} \left[\ell_e(S) - \ell_e(S \cup \{v\}) \right]. \tag{B.4}$$

Positive contributions indicate a reduction in recovery loss on the evaluated occurrence. The utilities are computed from inner training and evaluation partitions contained entirely within the outer training data.

B.2 NODE PREFERENCES AND FUSED AFFINITIES

Let N_a count all evaluated occurrences incident to node a. Let $\bar{\phi}_{av}$ be their mean contribution when $N_a > 0$, and let $\bar{\phi}_v$ be the global mean over occurrences. Let T_w denote the view temperature T in the main text. With $\kappa = 30$ and $T_w = 0.10$, define

$$\widetilde{\phi}_{av} = \frac{N_a \bar{\phi}_{av} + \kappa \bar{\phi}_v}{N_a + \kappa}, \qquad w_{av} = \frac{\exp(\widetilde{\phi}_{av}/T_w)}{\sum_{u=1}^{3} \exp(\widetilde{\phi}_{au}/T_w)}. \tag{B.5}$$

For $N_a = 0$, set $\widetilde{\phi}_{av} = \bar{\phi}_v$ directly. Using the cleaned full affinities before top-k truncation, the fused edge weight is

$$W_{ab} = \sum_{v=1}^{3} \frac{w_{av} + w_{bv}}{2} S^{(v)}_{ab,\mathrm{clean}}. \tag{B.6}$$

Symmetrize the fused matrix and remove its diagonal. Apply the same construction to both RNA types. The effective number of views at a node is $\exp(-\sum_v w_{av} \log w_{av})$; its average is an implementation diagnostic of weight concentration.

Algorithm 1 Task-conditioned biological graph construction

Input: training positives $\mathcal{P}_{\mathrm{tr}}$, permitted unlabeled pool $\mathcal{U}_{\mathrm{tr}}$, three views $S_t^{(v)}$ for each RNA type, and fusion seed s_f. Default values are $K_{\mathrm{in}} = 3$, $k_g = 10$, $\kappa = 30$, and $T_w = 0.10$.

Output: fused affinities W_m, W_c and node specific view weights w_m, w_c.

Require at least two training positives and a nonempty permitted unlabeled pool.
Partition $\mathcal{P}_{\mathrm{tr}}$ into $K = \min(K_{\mathrm{in}}, |\mathcal{P}_{\mathrm{tr}}|)$ seeded chunks.
Initialize an empty list of evaluated pair occurrences and contributions.
For each inner chunk $h = 1, \ldots, K$:
  Build $A^{(-h)}$ using positive pairs outside the held-out chunk only.
  Select an unlabeled block from $\mathcal{U}_{\mathrm{tr}}$ using the seeded cursor rule.
  Concatenate the held-out positives and selected unlabeled pairs with labels y_e.
  Set $\delta_h = \mathrm{clip}(\mathrm{mean}(A^{(-h)}), 10^{-5}, 1 - 10^{-5})$.
  For each biological view v:
    Clean each affinity and form its top-k_g row transition $T_t^{(v)}$.
    Compute $L = T_m^{(v)} A^{(-h)}$ and $H = A^{(-h)} (T_c^{(v)})^{\top}$.
    For each evaluated pair, set $p_e^{(v)} = \mathrm{clip}((L_e + H_e)/4 + \delta_h/2, 10^{-5}, 1 - 10^{-5})$.
  Set the empty-coalition probability to δ_h.
  For every nonempty coalition, average its member-view probabilities.
  Compute elementwise binary cross-entropy for all eight coalitions.
  Compute each pair-view contribution ϕ_{ev} using the Shapley formula in B.1.
  Append the evaluated pair occurrences and their contributions to the list.
Compute global mean contributions $\bar{\phi}_v$ over all recorded occurrences.
For each RNA node a on either side:
  Count its occurrences N_a and average their contributions when $N_a > 0$.
  Set $\widetilde{\phi}_{av} = (N_a \bar{\phi}_{av} + \kappa \bar{\phi}_v)/(N_a + \kappa)$; use $\bar{\phi}_v$ when $N_a = 0$.
  Set $w_{av} = \mathrm{softmax}_v(\widetilde{\phi}_{av}/T_w)$.
For each RNA type, set $W_{ab} = \sum_v (w_{av} + w_{bv}) S^{(v)}_{ab,\mathrm{clean}}/2$.
Symmetrize W, set its diagonal to zero, and return both RNA graphs and weights.

The permitted unlabeled pool is randomly permuted using a fixed seed. Under the default 1:1 ratio, each chunk selects a consecutive block whose size is the smaller of the held-out positive count and the pool size. The cursor advances by the block size. If the next block exceeds the pool boundary, the cursor resets to zero and the permutation is circularly shifted by the one-based chunk index. Each occurrence contributes separately to the utility estimates. Seed offsets are specified in Appendix F.1.

C DEGREE REFERENCE AND DUAL-GRAPH RESIDUAL DECOMPOSITION

C.1 DEGREE REFERENCE AND STANDARDIZED RESIDUAL

Let $d_i^m = \sum_j A_{ij}$ and $d_j^c = \sum_i A_{ij}$ be the training degrees. The degree-conditioned Bernoulli reference has the additive-logit form $\Pi_{ij} = \sigma(\alpha_i + \beta_j)$, with row and column targets d_i^m and d_j^c. This form corresponds to maximum Bernoulli entropy under expected-degree constraints when a finite interior solution exists. Finite numerical fitting approximates boundary targets such as zero or full degrees.

Initialize $g_0 = \mathrm{logit}(\mathrm{clip}(\mathrm{mean}(A), 10^{-7}, 1 - 10^{-7}))$ and

$$u_i = \mathrm{clip}\left(\frac{d_i^m + 0.5}{n+1}, 10^{-7}, 1 - 10^{-7}\right), \qquad \alpha_i = \mathrm{logit}(u_i) - g_0, \qquad \beta_j = g_0. \tag{C.1}$$

Each iteration first computes $P_{ij} = \sigma(\mathrm{clip}(\alpha_i + \beta_j, -30, 30))$ and updates rows:

$$\begin{aligned} \Delta\alpha_i &= \mathrm{clip}\left(\frac{d_i^m - \sum_j P_{ij}}{\max\{\sum_j P_{ij}(1 - P_{ij}), 10^{-8}\}}, -3, 3\right), \\ \alpha_i &\leftarrow \mathrm{clip}(\alpha_i + \Delta\alpha_i, -30, 30). \end{aligned} \tag{C.2}$$

Recompute P using the updated rows before applying the analogous column update to β_j. After both updates, set $g = \mathrm{mean}(\alpha)$ and replace $\alpha \leftarrow \alpha - g$, $\beta \leftarrow \beta + g$. This gauge shift preserves the summed logits. Fit for at most 100 iterations; check maximum absolute row and column degree errors at iteration 1, every fifth iteration, and the final iteration. Stop when both errors are at most 10^{-4}. The returned reference is the sigmoid of the clipped summed logits, and achieved errors are recorded.

The standardized degree-referenced residual is

$$R_{ij} = \mathrm{clip}\left(\frac{A_{ij} - \Pi_{ij}}{\sqrt{\max\{\Pi_{ij}(1 - \Pi_{ij}), 10^{-6}\}}}, -12, 12\right). \tag{C.3}$$

The variance floor bounds the standardization factor. Clipping bounds the residual magnitude. The same training-positive matrix supplies both degree targets and the observed term in this residual.

C.2 QUADRATIC TARGET AND NUMERICAL SOLUTION

For each fused affinity W_t, retain $\min(k_r, N_t - 1)$ neighbors per row with $k_r = 20$ and take the elementwise maximum with the transpose. Let A_t^g be this symmetric adjacency and $d_a = \sum_b (A_t^g)_{ab}$. Define $\widehat{D}_t = \mathrm{diag}(\max(d_a, 10^{-8}))$ and

$$\widehat{A}_t = \widehat{D}_t^{-1/2} A_t^g \widehat{D}_t^{-1/2}, \qquad L_t = I - \widehat{A}_t. \tag{C.4}$$

An isolated node has a zero normalized-adjacency row and Laplacian diagonal one. The graph-smooth residual minimizes

$$\begin{aligned} \mathcal{J}(G) = \frac{1}{2}\|G - R\|_F^2 + \frac{\lambda_m}{2}\,\mathrm{tr}(G^\top L_m G) + \frac{\lambda_c}{2}\,\mathrm{tr}(G L_c G^\top) \\ + \frac{\rho_r}{2n}\|G\mathbf{1}_n\|_2^2 + \frac{\rho_c}{2m}\|G^\top \mathbf{1}_m\|_2^2. \end{aligned} \tag{C.5}$$

The default coefficients are $\lambda_m = \lambda_c = 0.50$ and $\rho_r = \rho_c = 0.05$. Set $J_q = \mathbf{1}_q \mathbf{1}_q^\top / q$. Differentiation gives the linear system

$$\mathcal{A}(G) = R, \qquad \mathcal{A}(X) = X + \lambda_m L_m X + \lambda_c X L_c + \rho_r X J_n + \rho_c J_m X. \tag{C.6}$$

For nonnegative coefficients, the Laplacian and mean-penalty terms are positive semidefinite. Consequently, $\langle X, \mathcal{A}(X)\rangle_F \geq \|X\|_F^2$ for every X. The identity term makes the operator positive definite, so the fixed-graph objective has a unique minimizer. The complementary residual is $E = R - G$. The components reconstruct R exactly by definition; their Frobenius inner product can be nonzero.

Algorithm 2 solves the system with matrix conjugate gradients initialized at zero. Sparse adjacency products and row/column means implement the operator. The maximum iteration count is 40 and the relative recursive-residual tolerance is 10^{-5}.

Algorithm 2 Degree-referenced dual-graph residual decomposition

DegreeReference denotes the numerical fit in Appendix C.1. The operator $\mathcal{A}$ is defined in Eq. (C.6). All inner products are Frobenius inner products.

Input: observed-positive matrix A, fused affinities W_m, W_c, graph coefficients, $k_r = 20$, maximum CG iterations $T_{\mathrm{CG}} = 40$, and tolerance 10^{-5}.

Output: reference Π, standardized residual R, graph-smooth residual G, complementary residual E, and solver diagnostics.

Fit $\Pi \leftarrow \mathrm{DegreeReference}(A)$ using only the training-positive matrix.
Set $R = \mathrm{clip}((A - \Pi)/\sqrt{\max(\Pi(1-\Pi), 10^{-6})}, -12, 12)$ elementwise.
For each RNA type, clean W_t and retain $\min(k_r, N_t - 1)$ neighbors per row.
Take the elementwise maximum with the transpose to obtain a symmetric adjacency.
Normalize with degree floor 10^{-8} and form the sparse Laplacian operator L_t.
Initialize $G = 0$, residual $r = R$, search direction $p = r$, and $s = \langle r, r\rangle_F$.
Set $b = \max(\|R\|_F, 10^{-12})$, $\epsilon = \sqrt{s}/b$, and $q = 0$.
While $q < T_{\mathrm{CG}}$ and $\epsilon > 10^{-5}$:
  Compute $v = \mathcal{A}(p)$ through sparse adjacency products and row/column means.
  Set $a = s/\max(\langle p, v\rangle_F, 10^{-20})$.
  Update $G \leftarrow G + ap$ and $r \leftarrow r - av$.
  Set $s_{\mathrm{new}} = \langle r, r\rangle_F$, $\epsilon = \sqrt{s_{\mathrm{new}}}/b$, and $q \leftarrow q + 1$.
  If $\epsilon \leq 10^{-5}$, exit the loop.
  Update $p \leftarrow r + [s_{\mathrm{new}}/\max(s, 10^{-20})]p$ and $s \leftarrow s_{\mathrm{new}}$.
Set $E = R - G$.
Compute reconstruction error $\max_{i,j} |R_{ij} - G_{ij} - E_{ij}|$ and graph energies.
Return Π, R, G, E, completed iteration count q, and recursive residual ratio ϵ.

The solver uses float32 arithmetic and tests its recursive residual. Decomposition diagnostics include the reconstruction error and normalized graph energies:

$$\epsilon_{\mathrm{rec}} = \max_{i,j} |R_{ij} - G_{ij} - E_{ij}|, \qquad \mathcal{S}_m(X) = \frac{\langle X, L_m X\rangle_F}{\max(\|X\|_F^2, 10^{-12})}. \tag{C.7}$$

The circRNA energy replaces $L_m X$ by $X L_c$. Row and column drift are the root-mean-square row and column means of G. Algebraic reconstruction and linear-system convergence assess different properties: solver accuracy is assessed using the residual of Eq. (C.6).

D Base prediction and local evidence

D.1 Factor fitting and global residual integration

The rank-r factor model uses $r = 64$ and logits

$$F_{ij} = u_i^\top v_j + b_i + c_j + b_0. \tag{D.1}$$

Factor entries are initialized from a zero-mean Gaussian with standard deviation 0.05. Node biases start at zero and $b_0 = -3$. The default path uses this initialization. Define $M_{ij} = 10$ on training positives, $M_{ij} = 1$ on the full permitted unlabeled training pool, and $M_{ij} = 0$ elsewhere. With $\mathrm{BCELogit}(z, y) = \mathrm{softplus}(z) - yz$,

$$\mathcal{L}_{\text{data}} = \frac{\sum_{i,j} M_{ij}\,\text{BCELogit}(F_{ij}, A_{ij})}{\max(\sum_{i,j} M_{ij}, 1)}. \tag{D.2}$$

The factor regularizer uses fused affinities with $k_f = 10$. The supplied implementation selects top-k_f neighbors per row and keeps selected directed entries whose source index is smaller than the target index. An edge selected only in the opposite direction is omitted. This is the implemented factor-edge rule; the residual graph in Appendix C.2 uses a symmetric union. Let $\mathcal{E}_m^{\text{fac}}$ and $\mathcal{E}_c^{\text{fac}}$ denote the resulting factor edge sets. Define

$$\mathcal{R}_m(U) = \frac{\sum_{(i,i')\in\mathcal{E}_m^{\text{fac}}} w_{ii'}\|u_i - u_{i'}\|_2^2}{\max(\sum_{(i,i')\in\mathcal{E}_m^{\text{fac}}} w_{ii'}, 10^{-8})}. \tag{D.3}$$

Define $\mathcal{R}_c(V)$ analogously. The optimized factor objective is

$$\begin{aligned}\mathcal{L}_{\text{fac}} = \mathcal{L}_{\text{data}} &+ 0.003[\mathcal{R}_m(U) + \mathcal{R}_c(V)] \\ &+ 10^{-4}[\text{mean}(U^2) + \text{mean}(V^2) + \text{mean}(b^2) + \text{mean}(c^2)].\end{aligned} \tag{D.4}$$

Adam uses learning rate 0.03 for at most 300 epochs. Validation average precision (AP) is checked every 10 epochs and at the final epoch. A checkpoint improves the incumbent when AP increases by more than 10^{-5}. Training stops after eight consecutive checks without improvement. Restore the best checkpoint and freeze the factor parameters.

Scale the graph-smooth component using

$$s_G = \max\left(\sqrt{\frac{1}{mn}\sum_{i,j} G_{ij}^2}, 10^{-6}\right), \qquad \widetilde{G} = \tanh(\text{clip}(G/s_G, -8, 8)). \tag{D.5}$$

Select γ from $\{0, 0.05, 0.10, 0.20, 0.30, 0.50, 0.75, 1.00\}$ by validation AP. Traverse candidates in this order and replace the incumbent only for an improvement exceeding 10^{-8}. Then fix

$$Z^0 = F + \gamma\widetilde{G}, \qquad B = \sigma(Z^0). \tag{D.6}$$

The row and column totals of B supply the probability budgets for the projection in Appendix E.

D.2 Neighborhood features and reliability gating

Local features use E and the original biological views. For each view, remove the diagonal and retain $\min(k_n, N_t - 1)$ neighbors, with $k_n = 50$. Clamp retained affinities below at zero and multiply the entry of descending rank q by 0.9^q, starting at $q = 0$. Divide by the row sum with floor 10^{-8}. This yields local transitions $T_{m,\text{loc}}^{(v)}$ and $T_{c,\text{loc}}^{(v)}$; a zero row remains zero. These transitions are distinct from the graph-utility transitions in Appendix B.1.

For each view, compute $H^{(v)} = T_{m,\text{loc}}^{(v)}E$ and $K^{(v)} = E(T_{c,\text{loc}}^{(v)})^\top$. The four pair features are

$$x_{ij}^{(v)} = \left(H_{ij}^{(v)}, K_{ij}^{(v)}, \frac{H_{ij}^{(v)} + K_{ij}^{(v)}}{2}, |H_{ij}^{(v)} - K_{ij}^{(v)}|\right). \tag{D.7}$$

Sample $\widetilde{\mathcal{U}}_{\text{tr}}$ once from $\mathcal{U}_{\text{tr}}$. The default target size is $|\mathcal{P}_{\text{tr}}|$, capped by pool availability. Define $\mathcal{T} = \mathcal{P}_{\text{tr}} \cup \widetilde{\mathcal{U}}_{\text{tr}}$. For each feature coordinate, estimate the mean and population standard deviation from $\mathcal{T}$ and apply

$$\widehat{x}_{ij,q}^{(v)} = \text{clip}\left(\frac{x_{ij,q}^{(v)} - \mu_{v,q}}{\max(\sigma_{v,q}, 10^{-6})}, -12, 12\right). \tag{D.8}$$

The statistics remain fixed during local optimization. For view-specific parameters $\theta_v \in \mathbb{R}^4$, a_v, and r_v, define

$$e_{ij}^{(v)} = \theta_v^\top \widehat{x}_{ij}^{(v)}, \qquad \ell_{ij}^{(v)} = a_v - \operatorname{softplus}(r_v)\operatorname{softplus}(\widehat{x}_{ij,4}^{(v)}). \tag{D.9}$$

Normalize the gate across views and bound the combined evidence:

$$\begin{aligned} \omega_{ijv} &= \frac{\exp(\ell_{ij}^{(v)})}{\sum_{u=1}^{3}\exp(\ell_{ij}^{(u)})}, \\ D_{ij} &= [\operatorname{softplus}(s) + 10^{-4}]\tanh\left(\sum_{v=1}^{3}\omega_{ijv}e_{ij}^{(v)}\right). \end{aligned} \tag{D.10}$$

Initialize θ_v, a_v, and r_v at zero and $s = -1$, yielding zero initial evidence. There are 19 local parameters. The proposal logits and probabilities are $Z^{\text{prop}} = Z^0 + D$ and $Q = \sigma(Z^{\text{prop}})$.

E Bernoulli marginal projection

E.1 Constrained target and characterization

The proposal is projected onto probabilities with the base predictor's row and column totals. For interior base and proposal probabilities, define

$$\mathcal{C}(B) = \{P \in [0,1]^{m\times n} : P\mathbf{1}_n = B\mathbf{1}_n,\ P^\top\mathbf{1}_m = B^\top\mathbf{1}_m\}. \tag{E.1}$$

The mathematical target is

$$P^\star = \arg\min_{P\in\mathcal{C}(B)}\sum_{i,j}\left[P_{ij}\log\frac{P_{ij}}{Q_{ij}} + (1-P_{ij})\log\frac{1-P_{ij}}{1-Q_{ij}}\right]. \tag{E.2}$$

The matrix B is feasible. For interior probabilities, each summand has second derivative $1/[p(1-p)] > 0$, so the objective is strictly convex and the constrained probability minimizer is unique. Feasibility of an interior point and the boundary behavior of the derivative place the optimum in the interior. The Lagrangian first-order condition yields

$$\operatorname{logit}(P_{ij}^\star) = Z_{ij}^{\text{prop}} + \eta_i + \xi_j. \tag{E.3}$$

The signs of η and ξ absorb the multiplier convention. The offsets have one gauge degree of freedom: adding a constant to every row offset and subtracting it from every column offset preserves all logits. The degree reference Π used to construct R is distinct from the base probabilities B that define these constraints.

E.2 Finite updates and marginal diagnostics

Algorithm 3 performs alternating row and column updates. It starts each call from zero offsets and recomputes probabilities after the row update. Training differentiates through four iterations. Validation and final prediction use 60 iterations.

Logits are clipped to $[-30, 30]$ when computing probabilities within an iteration. The returned logits retain the full additive offsets without this clipping. The finite output is $P^{(T)} = \sigma(Z^{(T)})$, which is distinguished from the exact solution $P^\star$. Its achieved marginal errors are

$$\epsilon_r = \max_i\left|\sum_j(P_{ij}^{(T)} - B_{ij})\right|, \qquad \epsilon_c = \max_j\left|\sum_i(P_{ij}^{(T)} - B_{ij})\right|. \tag{E.4}$$

Algorithm 3 Finite Bernoulli marginal projection

All divisions and clipping operations act elementwise. Row and column totals use the complete fixed RNA-pair grid.

Input: proposal logits Z^{prop}, base probability matrix $B = \sigma(Z^0)$, and iteration count T.

Output: numerical projected logits $Z^{(T)}$. Default values are $T = 4$ for training and $T = 60$ for validation and final prediction.

Set row targets $r = B\mathbf{1}_n$ and column targets $c = B^\top \mathbf{1}_m$.
Initialize row offsets $\eta = 0_m$ and column offsets $\xi = 0_n$ for this call.
For $t = 1, \ldots, T$:
  Set $Z = Z^{\mathrm{prop}} + \eta \mathbf{1}_n^\top + \mathbf{1}_m \xi^\top$.
  Set $P = \sigma(\mathrm{clip}(Z, -30, 30))$.
  Compute $v_r = \max([P \odot (1 - P)]\mathbf{1}_n, 10^{-8})$.
  Update $\eta \leftarrow \eta + \mathrm{clip}((r - P\mathbf{1}_n)/v_r, -3, 3)$.
  Recompute $Z = Z^{\mathrm{prop}} + \eta \mathbf{1}_n^\top + \mathbf{1}_m \xi^\top$.
  Recompute $P = \sigma(\mathrm{clip}(Z, -30, 30))$.
  Compute $v_c = \max([P \odot (1 - P)]^\top \mathbf{1}_m, 10^{-8})$.
  Update $\xi \leftarrow \xi + \mathrm{clip}((c - P^\top \mathbf{1}_m)/v_c, -3, 3)$.
  Set $g = \mathrm{mean}(\eta)$; update $\eta \leftarrow \eta - g$ and $\xi \leftarrow \xi + g$.
Return $Z^{(T)} = Z^{\mathrm{prop}} + \eta \mathbf{1}_n^\top + \mathbf{1}_m \xi^\top$.

These errors are evaluated over the complete fixed RNA-pair grid against the base predictor's probability totals. A fixed iteration budget yields an approximate constraint solution. The theoretical characterization in Appendix E.1 assumes interior probabilities; finite-precision sigmoid saturation and logit clipping belong to the numerical implementation.

F Staged optimization and computational cost

F.1 Local objective and complete fold procedure

Factor fitting, global-scale selection, and local fitting occur sequentially. The base logits remain fixed during local optimization. Let $z^{(4)}$ be the logits returned by four projection iterations. On the fixed local training set $\mathcal{T}$, minimize

$$\mathcal{L}_{\mathrm{local}} = \frac{1}{|\mathcal{T}|} \sum_{e \in \mathcal{T}} [\mathrm{softplus}(z_e^{(4)}) - y_e z_e^{(4)}] + \frac{10^{-4}}{|\mathcal{T}|} \sum_{e \in \mathcal{T}} D_e^2. \tag{F.1}$$

AdamW uses learning rate 0.03 and weight decay 10^{-4}, applied separately from the evidence penalty. Clip the gradient norm at 5. Run at most 300 epochs, checking validation AP every fifth epoch and at the final epoch with 60 projection iterations. Require an improvement greater than 10^{-5} and stop after 20 unsuccessful checks. Include the zero-evidence initialization as a candidate, scored by the base validation predictions. Restore the best local parameters before final prediction.

Factor fitting uses the full permitted unlabeled training pool. Local fitting uses the subset sampled once. For experiment seed s and fold index f, the supplied workflow uses $s_{\mathrm{fold}} = 1009s + 9176f$. The fusion seed is $s_{\mathrm{fold}} + 1200$. Positive inner chunks and the initial unlabeled permutation use fusion-seed offsets 3301 and 4409. Local sampling uses $s_{\mathrm{fold}} + 3000$. The positive chunks and unlabeled permutation are generated independently.

After model selection, compute the final logits on the full grid and cache $C = Z^{(60)} - Z^0$. A query returns $\sigma(Z_{ij}^0 + C_{ij})$. Chunked pair scoring reads this cache. Select the classification threshold by maximizing validation F1 along the precision–recall curve, then freeze it for test evaluation. Threshold-free metrics use scores, while threshold-dependent metrics use the fixed validation-selected threshold.

Learned quantities comprise the factor parameters and the local evidence parameters. View utilities are computed from inner evaluations. The degree reference and projection offsets are fitted numeri-

cally, and the graph residual solves a fixed quadratic problem. Their different optimization roles are reflected in the order of Algorithm 4.

Algorithm 4 Training within one outer fold

Input: training positives $\mathcal{P}_{\mathrm{tr}}$, permitted unlabeled pool $\mathcal{U}_{\mathrm{tr}}$, labeled validation set $\mathcal{V}$, labeled test set $\mathcal{E}_{\mathrm{test}}$, six biological affinities, and fold seed s_{fold}. Test labels enter only the final evaluation.

Output: frozen base logits Z^0, cached correction C, validation-selected threshold τ_{cls}, and test metrics. The notation Θ collects only the local evidence parameters defined in Appendix D.2.

Audit split overlap and construct A from $\mathcal{P}_{\mathrm{tr}}$ only.
Compute W_m, W_c with Algorithm 1 using fusion seed $s_{\mathrm{fold}} + 1200$. Compute Π, R, G, E with Algorithm 2.
Fit the factor model on the explicit mask in Appendix D.1, using the edge rule in Appendix D.1. Restore its best validation-AP checkpoint according to Appendix D.1 and freeze the factors.
Transform G as specified in Appendix D.1; select γ by validation AP on its fixed grid. Set $Z^0 = F + \gamma\widetilde{G}$ and $B = \sigma(Z^0)$; keep both fixed below.
Build all four endpoint features from E through each original biological view. Sample $\widetilde{\mathcal{U}}_{\mathrm{tr}}$ once using seed $s_{\mathrm{fold}} + 3000$; set $\mathcal{T} = \mathcal{P}_{\mathrm{tr}} \cup \widetilde{\mathcal{U}}_{\mathrm{tr}}$. Fit means and population standard deviations on $\mathcal{T}$; standardize the full feature grid.
Initialize Θ as in Appendix D.2, giving zero local evidence; create the AdamW optimizer. Save the initial state as Θ_{best}; set $s_{\mathrm{best}} = \mathrm{AP}(B_{\mathcal{V}})$ and wait = 0.
For epoch = 1 to 300:
  Form evidence D_Θ from the fixed standardized features using Appendix D.2. Set $Z_{\mathrm{train}} = \mathrm{Algorithm3}(Z^0 + D_\Theta, B, 4)$.
  Evaluate the mean BCE and evidence penalty in Eq. (F.1) on $\mathcal{T}$. Backpropagate through the four updates into Θ; clip gradient norm at 5. Update Θ with AdamW at learning rate 0.03 and weight decay 10^{-4}.
  Every fifth epoch, or at the final epoch:
    Recompute D_Θ and $Z_{\mathrm{val}} = \mathrm{Algorithm3}(Z^0 + D_\Theta, B, 60)$ without gradients. Set $s = \mathrm{AP}(\sigma(Z_{\mathrm{val}})_{\mathcal{V}})$.
    If $s > s_{\mathrm{best}} + 10^{-5}$, set $\Theta_{\mathrm{best}} \leftarrow \Theta$ and $s_{\mathrm{best}} \leftarrow s$; reset wait to 0.
    Otherwise, increment wait by one. If wait reaches 20, exit the training loop.
Restore Θ_{best} and recompute its evidence D. Set $Z^{(60)} = \mathrm{Algorithm3}(Z^0 + D, B, 60)$ and cache $C = Z^{(60)} - Z^0$. Report row and column marginal errors from $\sigma(Z^{(60)})$ against B.
Select τ_{cls} by validation F1 using the frozen final predictions. Evaluate $\mathcal{E}_{\mathrm{test}}$ using $\sigma(Z^0 + C)$ and the fixed threshold; return the saved outputs.

F.2 Computational complexity and memory

Let $V = 3$ be the number of views, K the number of inner chunks, and N_{ev} the total number of inner evaluation occurrences. The graph-utility code materializes dense transition matrices. Its propagation therefore costs $O(KV(m^2n + mn^2))$. Coalition evaluation adds $O(2^V V N_{\mathrm{ev}})$ arithmetic, and node aggregation adds $O(VN_{\mathrm{ev}})$. Full affinity fusion costs $O(V(m^2 + n^2))$, in addition to top-k selection and normalization. A sparse transition implementation would reduce propagation to $O(KVk_gmn)$; the supplied graph-utility path uses dense multiplication.

Degree-reference fitting costs $O(T_{\mathrm{null}}mn)$ for T_{null} iterations. The residual solver stores normalized adjacencies sparsely. If their nonzero counts are s_m and s_c, its cost is

$$O\left(T_{\mathrm{CG}}[mn + s_m n + s_c m]\right). \tag{F.2}$$

A factor epoch costs $O(mnr)$ for the full logit matrix, plus $O((|\mathcal{E}_m^{\mathrm{fac}}| + |\mathcal{E}_c^{\mathrm{fac}}|)r)$ for graph regularization. The full masked loss also scans mn positions. Local feature construction uses dense transitions and costs $O(V(m^2n+mn^2))$ once. Each local epoch evaluates the evidence in $O(Vmn)$ and performs projection in $O(T_pmn)$, with $T_p = 4$ during training and 60 during validation or final prediction. Reverse differentiation through the fixed iteration sequence has the same leading arithmetic order.

The six affinity matrices occupy $O(V(m^2 + n^2))$ memory. The local feature grid contains $4Vmn$ values. Factor storage is $O((m + n)r)$, and the sparse residual graphs occupy $O(s_m + s_c)$. Dense transitions add $O(V(m^2 + n^2))$ storage. Interaction-shaped intermediates require $O(mn)$ each; the

unrolled projection can retain $O(T_p mn)$ activations for differentiation. The implementation retains quadratic affinity storage and full-grid operations despite sparse residual graphs. Runtime and peak memory depend on the actual RNA sizes, hardware, and selected checkpoints.

G Extended benchmark results

This section provides the numerical records supporting the benchmark and mechanism comparisons. Table 4 summarizes the three primary datasets. ACC, sensitivity, specificity, precision, and MCC use a percentage scale. AUC and AUPR use a 0–1 scale. TAPER entries report means and standard deviations across five outer folds. Baseline values in the cross-method benchmark tables are taken from the original publications.

Table 4: Dataset sizes for the three primary interaction benchmarks.

Dataset	miRNAs	circRNAs	Interactions
CMI-9905	962	2,346	9,905
CMI-9589	821	2,115	9,589
CMI-20208	1,152	3,569	20,208

Baseline sources. The methods in Table 1 are BioDGW-CMI (Wang et al., 2023), iHofman (Yu et al., 2025), BEROLECMI (Wang et al., 2024), Hither-CMI (Jiang et al., 2025), MRHRL (Yin et al., 2024b), MuGNet-CMI (Jiang et al., 2026), LSAGHI (Yin et al., 2025), MFERL (Peng et al., 2025), DGCLCMI (Cao et al., 2025), and CMAGN (Yin et al., 2024a).

G.1 Results for individual evaluation folds

Tables 5–7 report the individual test-fold results. The mean AUC values, rounded to four decimal places, are 0.9721, 0.9827, and 0.9691 for CMI-9905, CMI-9589, and CMI-20208, respectively. These agree with Table 1. The fold records also show variation in threshold-dependent metrics under validation-selected thresholds.

Table 5: Individual outer-fold results on CMI-9905. Each row corresponds to one test fold. Bold and underlined values mark the highest and second-highest values in each metric column, respectively.

Fold	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
Fold 1	93.24	<u>91.57</u>	94.90	94.73	86.52	0.9698	0.9782
Fold 2	92.15	86.77	**97.53**	**97.23**	84.79	<u>0.9733</u>	0.9787
Fold 3	**93.72**	90.86	96.57	96.36	<u>87.57</u>	0.9722	<u>0.9799</u>
Fold 4	<u>93.69</u>	90.16	<u>97.22</u>	<u>97.01</u>	**87.60**	0.9699	0.9784
Fold 5	93.34	**93.39**	93.29	93.29	86.67	**0.9751**	**0.9816**

Table 6: Individual outer-fold results on CMI-9589. Each row corresponds to one test fold. Bold and underlined values mark the highest and second-highest values in each metric column, respectively.

Fold	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
Fold 1	93.85	93.01	94.68	94.59	87.71	0.9823	0.9857
Fold 2	94.08	91.35	**96.82**	**96.64**	88.30	**0.9847**	**0.9870**
Fold 3	**94.26**	91.87	<u>96.66</u>	<u>96.50</u>	**88.63**	0.9818	0.9852
Fold 4	94.08	<u>93.07</u>	95.10	95.00	88.18	0.9818	0.9857
Fold 5	<u>94.18</u>	**94.11**	94.26	94.25	<u>88.37</u>	<u>0.9831</u>	<u>0.9865</u>

Table 7: Individual outer-fold results on CMI-20208. Each row corresponds to one test fold. Bold and underlined values mark the highest and second-highest values in each metric column, respectively.

Fold	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
Fold 1	90.98	89.66	92.31	92.10	81.99	0.9675	0.9705
Fold 2	91.49	<u>90.82</u>	92.16	92.05	82.99	0.9666	0.9718
Fold 3	91.58	88.05	**95.10**	**94.73**	<u>83.36</u>	**0.9713**	0.9741
Fold 4	**92.01**	90.52	<u>93.49</u>	<u>93.29</u>	**84.05**	0.9695	<u>0.9745</u>
Fold 5	<u>91.65</u>	**90.84**	92.45	92.33	83.31	<u>0.9707</u>	**0.9751**

G.2 An additional lncRNA–miRNA interaction benchmark

The additional benchmark was obtained from lncRNASNP2 (Miao et al., 2018) and contains 4,966 experimentally supported interactions involving 770 lncRNAs and 275 miRNAs. Its RNA type, network composition, and interaction scale differ from those of the primary datasets. TAPER is trained on the corresponding training partitions and evaluated with five-fold cross-validation. This experiment assesses the architecture on a related RNA interaction task after task-specific training.

Table 8 reports an AUC of 0.9113 ± 0.0038 and an AUPR of 0.9248 ± 0.0026 for TAPER. These exceed the reported MGCAT values by 0.0103 and 0.0056. TAPER also has the highest reported accuracy, at $85.58 \pm 0.61\%$. MGCAT has higher precision and F1, while LncMirNet has the highest recall. The strongest comparative results for TAPER are therefore in the ranking metrics and accuracy.

The baseline methods, linked to their original publications, are LncMirNet, RNAI-FRID, GCLMI, GATv2, CFANE, NDALMA, preMLI, and MGCAT.

Table 8: Results on the lncRNASNP2-derived benchmark. Baseline entries are literature-reported results. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Method	AUC	AUPR	ACC(%)	Pre(%)	Rec(%)	F1(%)
LncMirNet	0.8066 ± 0.0110	0.8017 ± 0.0083	72.98 ± 1.17	70.19 ± 1.12	**79.88 ± 1.64**	74.72 ± 1.14
RNAI-FRID	0.8292 ± 0.0069	0.8468 ± 0.0066	75.29 ± 0.36	76.16 ± 0.98	73.68 ± 1.68	74.88 ± 0.56
GCLMI	0.8544 ± 0.0166	0.8678 ± 0.0097	78.62 ± 1.82	80.69 ± 1.34	75.23 ± 2.84	77.86 ± 2.09
GATv2	0.8651 ± 0.0124	0.8693 ± 0.0130	80.54 ± 1.13	82.07 ± 1.65	78.21 ± 2.79	80.06 ± 1.38
CFANE	0.8733 ± 0.0104	0.8880 ± 0.0098	79.91 ± 0.88	83.01 ± 1.11	75.25 ± 2.29	78.92 ± 1.15
NDALMA	0.8389 ± 0.0105	0.8516 ± 0.0090	77.41 ± 1.24	77.61 ± 1.93	77.13 ± 2.11	77.34 ± 1.20
preMLI	0.8104 ± 0.0100	0.8028 ± 0.0104	73.48 ± 1.63	71.46 ± 2.96	<u>78.66 ± 4.00</u>	74.78 ± 1.22
MGCAT	<u>0.9010 ± 0.0053</u>	<u>0.9192 ± 0.0064</u>	<u>85.02 ± 0.72</u>	**92.34 ± 0.86**	76.38 ± 1.15	**83.60 ± 0.90**
TAPER	**0.9113 ± 0.0038**	**0.9248 ± 0.0026**	**85.58 ± 0.61**	<u>91.65 ± 3.64</u>	75.33 ± 3.24	<u>82.59 ± 0.90</u>

G.3 Numerical results for mechanism comparisons

Tables 9–11 report the numerical results for Figure 2. Each comparison uses CMI-9905 and replaces one mechanism while retaining the surrounding model structure and five-fold protocol. AUC and AUPR retain the 0–1 scale here; the figure displays them on a percentage scale.

The graph-propagation alternatives are GFHF (Zhu et al., 2003), LGC (Zhou et al., 2003), APPNP (Gasteiger et al., 2019a), and GDC (Gasteiger et al., 2019b). The view-fusion comparison includes SNF (Wang et al., 2014), ANF (Ma & Zhang, 2018), NEMO (Rappoport & Shamir, 2019), and the STATIS-UMKL variant of Mariette & Villa-Vialaneix (2018). The marginal-adjustment comparison uses temperature scaling (Guo et al., 2017), isotonic regression (Zadrozny & Elkan, 2002), Sinkhorn scaling (Cuturi, 2013), and unbalanced transport scaling (Chizat et al., 2018).

For residual modeling, TAPER exceeds LGC by 0.0204 AUC and 0.0194 AUPR. The corresponding differences against UMKL for view fusion are 0.0248 and 0.0233. For marginal projection, the differences against UOT-Scaling are 0.0185 and 0.0186. These results compare the implemented alternatives at each stage.

H Random-seed and sampling analyses

H.1 Variation across random seeds

Table 12 reports five runs with seeds 0–4 using the same outer folds and validation partitions. The experiment seed controls model initialization and stochastic sampling within the training procedure. Each run completes all five outer folds. The across-seed summary therefore measures variation in model fitting under fixed evaluation partitions.

Each seed row reports the mean and standard deviation across its five outer folds. The final row summarizes the five seed-level means using their sample standard deviation. The seed-level AUC means range from 0.9690 to 0.9721, with an across-seed summary of 0.9705 ± 0.0012. AUPR ranges from 0.9768 to 0.9794, with a summary of 0.9782 ± 0.0010. Accuracy averages $92.81 \pm 0.33\%$ across seeds.

H.2 Larger test candidate pools

This experiment varies the positive-to-unlabeled ratio in the test pool while retaining the training and validation procedure. Table 13 reports ratios from 1:1 to 1:20 and an ALL setting.

For each outer fold, ALL includes every unlabeled pair assigned to that fold's reserved test pool. These pairs are excluded from training and validation. The positive test pairs are unchanged across the evaluated ratios. AUC remains between 0.9707 and 0.9721. AUPR decreases from 0.9794 at 1:1 to 0.8848 at 1:20 and 0.7102 under ALL. Precision decreases from 95.72% to 8.96%, while sensitivity stays at 90.55%. The increase in accuracy to 95.29% accompanies a decrease in MCC to 27.18%. These values describe how prediction performance changes with the size and class composition of the candidate pool.

I Sensitivity to model settings

Each sweep changes the indicated setting on CMI-9905 and retains the other reference settings. Tables report means and standard deviations across five folds. Bold and underlined entries identify the highest and second-highest mean in each metric column. Bold setting labels mark the reference configuration.

I.1 Graph smoothness and view-fusion temperature

Table 14 varies the circRNA and miRNA smoothness coefficients jointly. The reference setting $(\lambda_c, \lambda_m) = (0.5, 0.5)$ gives 0.9721 AUC and 0.9794 AUPR. Among the remaining configurations, $(0.7, 0.3)$ has the highest AUC and AUPR, at 0.9526 and 0.9604. The gaps from the reference are 0.0195 and 0.0190, respectively.

Table 15 evaluates the view-fusion temperature T_w. The reference value 0.10 has the highest mean in each metric. AUC and AUPR are 0.9494 and 0.9573 at 0.05, and 0.9270 and 0.9321 at 0.50. Both more concentrated and more diffuse view weights produce lower ranking performance at the tested settings.

Table 9: Residual-modeling alternatives on CMI-9905, corresponding to Figure 2(a). Values are means ± standard deviations across five folds. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Method	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
GFHF	86.26 ± 0.60	82.80 ± 3.59	89.41 ± 3.43	87.85 ± 3.05	72.63 ± 1.12	0.9247 ± 0.0038	0.9343 ± 0.0033
LGC	<u>89.80 ± 0.46</u>	<u>86.92 ± 2.49</u>	<u>92.42 ± 2.18</u>	<u>91.33 ± 2.11</u>	<u>79.67 ± 0.90</u>	<u>0.9517 ± 0.0043</u>	<u>0.9600 ± 0.0033</u>
APPNP	89.22 ± 0.43	86.12 ± 2.09	92.04 ± 1.50	90.82 ± 1.42	78.47 ± 0.80	0.9468 ± 0.0048	0.9554 ± 0.0035
GDC	87.68 ± 0.60	84.81 ± 2.47	90.29 ± 2.46	88.91 ± 2.24	75.40 ± 1.22	0.9360 ± 0.0042	0.9455 ± 0.0042
TAPER	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**

Table 10: Biological view-fusion alternatives on CMI-9905, corresponding to Figure 2(b). Values are means ± standard deviations across five folds. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Method	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
SNF	87.23 ± 0.60	83.91 ± 1.72	90.12 ± 2.02	88.14 ± 2.05	74.37 ± 1.29	0.9293 ± 0.0048	0.9400 ± 0.0035
ANF	87.75 ± 0.75	83.45 ± 1.88	91.51 ± 1.45	89.64 ± 1.47	75.47 ± 1.49	0.9351 ± 0.0031	0.9450 ± 0.0034
NEMO	88.10 ± 0.24	85.80 ± 2.36	90.10 ± 1.74	88.34 ± 1.59	76.12 ± 0.43	0.9409 ± 0.0041	0.9481 ± 0.0035
UMKL	<u>89.08 ± 0.16</u>	<u>86.03 ± 2.08</u>	<u>91.86 ± 1.78</u>	<u>90.63 ± 1.71</u>	<u>78.20 ± 0.27</u>	<u>0.9473 ± 0.0041</u>	<u>0.9561 ± 0.0024</u>
TAPER	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**

Table 11: Marginal-adjustment alternatives on CMI-9905, corresponding to Figure 2(c). Values are means ± standard deviations across five folds. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Method	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
Temperature scaling	87.36 ± 0.40	83.69 ± 1.39	90.61 ± 1.28	88.77 ± 1.22	74.67 ± 0.81	0.9339 ± 0.0038	0.9426 ± 0.0036
Isotonic regression	87.04 ± 0.48	85.58 ± 1.69	88.50 ± 1.43	88.18 ± 1.14	74.14 ± 0.94	0.9302 ± 0.0045	0.9377 ± 0.0042
Sinkhorn–Knopp	89.27 ± 0.50	85.07 ± 1.78	92.98 ± 1.30	91.50 ± 1.34	78.56 ± 0.96	0.9473 ± 0.0038	0.9548 ± 0.0031
UOT-Scaling	<u>90.09 ± 0.12</u>	<u>86.70 ± 2.27</u>	<u>93.09 ± 2.08</u>	<u>91.81 ± 2.07</u>	<u>80.21 ± 0.29</u>	<u>0.9536 ± 0.0037</u>	<u>0.9608 ± 0.0027</u>
TAPER	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**

Table 12: Results across random seeds on CMI-9905. Seed rows report fold means and standard deviations. The last row reports the mean and sample standard deviation of the five seed-level means. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively. The across-seed summary is excluded from ranking.

Seed	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
0	**93.23 ± 0.64**	90.55 ± 2.43	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
1	92.80 ± 0.33	90.33 ± 1.35	95.28 ± 0.80	95.04 ± 0.75	85.72 ± 0.59	0.9712 ± 0.0021	0.9784 ± 0.0019
2	92.34 ± 0.48	90.85 ± 0.79	93.82 ± 1.03	93.64 ± 0.98	84.72 ± 0.98	0.9690 ± 0.0050	0.9768 ± 0.0030
3	92.72 ± 0.57	89.69 ± 1.77	95.75 ± 1.83	95.52 ± 1.74	85.64 ± 1.13	0.9700 ± 0.0038	0.9776 ± 0.0025
4	92.95 ± 0.52	**91.05 ± 1.34**	94.86 ± 2.00	94.71 ± 1.92	86.01 ± 1.12	0.9704 ± 0.0022	0.9786 ± 0.0015
Across seeds	92.81 ± 0.33	90.49 ± 0.53	95.12 ± 0.83	94.93 ± 0.82	85.74 ± 0.69	0.9705 ± 0.0012	0.9782 ± 0.0010

Table 13: Effect of the positive-to-unlabeled test ratio on CMI-9905. Values are means ± standard deviations across five folds. ALL uses the full eligible unlabeled test pool. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively. Tied values share the same format.

Test ratio	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
1:1	93.23 ± 0.64	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
1:5	94.51 ± 1.18	90.55 ± 2.43	95.30 ± 1.88	79.92 ± 6.22	81.77 ± 2.78	0.9707 ± 0.0025	0.9394 ± 0.0040
1:10	94.94 ± 1.59	90.55 ± 2.43	95.39 ± 1.99	67.46 ± 9.39	75.43 ± 4.98	0.9710 ± 0.0024	0.9156 ± 0.0038
1:20	95.09 ± 1.82	90.55 ± 2.43	95.32 ± 2.02	51.14 ± 10.91	65.64 ± 6.77	0.9709 ± 0.0024	0.8848 ± 0.0040
ALL	**95.29 ± 1.92**	90.55 ± 2.43	95.32 ± 1.94	8.96 ± 3.89	27.18 ± 5.55	0.9710 ± 0.0025	0.7102 ± 0.0089

Table 14: Sensitivity to the graph smoothness coefficients on CMI-9905. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

(λ_c, λ_m)	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
(0.3, 0.3)	87.16 ± 0.44	84.41 ± 2.36	89.55 ± 1.51	87.58 ± 1.35	74.22 ± 0.81	0.9316 ± 0.0054	0.9397 ± 0.0051
(0.3, 0.7)	88.94 ± 0.53	84.22 ± 2.57	93.04 ± 1.45	91.37 ± 1.42	77.90 ± 0.95	0.9450 ± 0.0033	0.9520 ± 0.0029
(0.5, 0.5)	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
(0.7, 0.3)	89.67 ± 0.32	86.78 ± 1.32	92.29 ± 1.50	91.13 ± 1.46	79.35 ± 0.68	0.9526 ± 0.0036	0.9604 ± 0.0023
(0.7, 0.7)	87.41 ± 0.48	85.10 ± 1.24	89.62 ± 0.92	88.65 ± 0.80	74.86 ± 0.94	0.9338 ± 0.0027	0.9455 ± 0.0029

Table 15: Sensitivity to the view-fusion temperature on CMI-9905. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

T_w	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
0.03	87.79 ± 0.98	83.58 ± 3.96	91.29 ± 2.52	89.02 ± 2.49	75.47 ± 1.79	0.9335 ± 0.0060	0.9403 ± 0.0059
0.05	89.76 ± 0.52	85.91 ± 2.11	93.14 ± 1.22	91.69 ± 1.20	79.53 ± 0.98	0.9494 ± 0.0047	0.9573 ± 0.0039
0.10	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
0.20	88.89 ± 0.47	83.23 ± 1.05	93.60 ± 1.00	91.57 ± 1.16	77.70 ± 0.98	0.9430 ± 0.0038	0.9484 ± 0.0038
0.50	86.63 ± 0.59	82.16 ± 3.02	90.36 ± 2.77	87.79 ± 2.61	73.12 ± 1.22	0.9270 ± 0.0035	0.9321 ± 0.0037

I.2 Local neighborhood size and factor rank

Table 16 varies the neighborhood count used for local residual features in Appendix D.2. The reference $k_n = 50$ gives 0.9721 AUC and 0.9794 AUPR. The corresponding values are 0.9240 and 0.9300 for $k_n = 10$, and 0.9411 and 0.9521 for $k_n = 100$. The highest ranking performance occurs at an intermediate neighborhood size in this sweep.

Table 16: Sensitivity to the local-feature neighborhood size on CMI-9905. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

k_n	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
10	86.37 ± 0.53	82.99 ± 1.44	89.19 ± 2.14	86.55 ± 2.03	72.51 ± 1.04	0.9240 ± 0.0030	0.9300 ± 0.0032
30	89.95 ± 0.31	87.11 ± 2.20	92.53 ± 2.11	91.45 ± 1.98	79.95 ± 0.63	0.9535 ± 0.0039	0.9613 ± 0.0027
50	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
75	90.58 ± 0.11	88.28 ± 2.12	92.67 ± 2.07	91.70 ± 1.97	81.19 ± 0.25	0.9566 ± 0.0042	0.9644 ± 0.0027
100	88.43 ± 0.32	86.20 ± 1.42	90.55 ± 0.98	89.69 ± 0.82	76.89 ± 0.61	0.9411 ± 0.0054	0.9521 ± 0.0036

Table 17 evaluates factor ranks from 16 to 256. Rank 64 has the highest mean AUC and AUPR. The next-highest values occur at rank 128, with 0.9503 AUC and 0.9581 AUPR. Rank 256 gives 0.9375 AUC and 0.9445 AUPR. Increasing the factor dimension beyond the reference reduces ranking performance in this sweep.

Table 17: Sensitivity to the latent factor rank on CMI-9905. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Rank	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
16	86.12 ± 0.58	84.53 ± 3.15	87.63 ± 3.48	86.83 ± 2.75	72.35 ± 1.08	0.9278 ± 0.0044	0.9404 ± 0.0041
32	88.66 ± 0.39	85.40 ± 1.40	91.59 ± 1.07	90.16 ± 1.03	77.30 ± 0.77	0.9436 ± 0.0048	0.9521 ± 0.0038
64	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**
128	89.51 ± 0.56	85.18 ± 1.19	93.42 ± 0.94	92.11 ± 0.99	79.10 ± 1.11	0.9503 ± 0.0045	0.9581 ± 0.0036
256	88.21 ± 0.46	84.89 ± 2.06	91.09 ± 1.93	89.29 ± 1.86	76.35 ± 0.90	0.9375 ± 0.0045	0.9445 ± 0.0043

J RECORDED COMPUTATIONAL COST

Table 18 reports the recorded runtime per fold and parameter counts. Mean runtime is 52.67 s on CMI-9589, 118.64 s on CMI-9905, and 206.62 s on CMI-20208. The storage column converts the reported parameter count N into $4N/2^{20}$ MiB, assuming four bytes per parameter. The affinity matrices, feature tensors, and projection intermediates have the additional memory costs described in Appendix F.2.

Table 18: Recorded computational cost. Runtime is mean ± standard deviation per fold. Storage is calculated from those reported counts.

Dataset	Runtime/fold (s)	Reported parameters	Four-byte parameter storage (MiB)
CMI-9905	118.64 ± 40.13	215,040	0.820
CMI-9589	52.67 ± 11.37	190,860	0.728
CMI-20208	206.62 ± 60.32	306,885	1.171

K NODE-HELD-OUT EVALUATION

We evaluate two node-held-out settings on CMI-9905: circRNA-cold and miRNA-cold. In each setting, nodes of the corresponding RNA type are partitioned into five outer folds. All interactions incident to the test nodes are excluded from training and validation. Validation nodes are selected from the remaining nodes, and their incident interactions are excluded from training. Biological similarities remain available for all nodes in the predefined RNA universe. The pair split provides the standard interaction-level reference.

Table 19 reports 0.9242 ± 0.0078 AUC and 0.9290 ± 0.0072 AUPR for circRNA-cold evaluation. The miRNA-cold results are 0.8923 ± 0.0043 AUC and 0.8969 ± 0.0044 AUPR. Relative to the pair-split AUC of 0.9721, the decreases are 0.0479 and 0.0798. Both settings yield lower ranking performance, with the larger decrease under miRNA-cold evaluation. These results characterize prediction for RNAs whose interaction histories are withheld while their biological similarities remain available.

Table 19: Node-held-out evaluation on CMI-9905. Values are means ± standard deviations across five folds. Pair split is the interaction-level reference. Bold and underlined entries mark the highest and second-highest means in each metric column, respectively.

Setting	ACC(%)	Sen(%)	Spe(%)	Pre(%)	MCC(%)	AUC	AUPR
circRNA-cold	<u>87.01 ± 1.08</u>	78.50 ± 3.72	<u>93.77 ± 2.28</u>	<u>91.04 ± 2.76</u>	<u>73.97 ± 2.14</u>	<u>0.9242 ± 0.0078</u>	<u>0.9290 ± 0.0072</u>
miRNA-cold	83.04 ± 0.59	<u>79.74 ± 2.86</u>	85.79 ± 2.68	82.49 ± 2.29	65.81 ± 1.16	0.8923 ± 0.0043	0.8969 ± 0.0044
Pair split	**93.23 ± 0.64**	**90.55 ± 2.43**	**95.90 ± 1.78**	**95.72 ± 1.68**	**86.63 ± 1.14**	**0.9721 ± 0.0023**	**0.9794 ± 0.0014**