# INCOMPLETE GROMOV-WASSERSTEIN ALIGNMENT FOR MULTI-VIEW CLUSTERING

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

## ABSTRACT

Incomplete multi-view clustering (IMVC) struggles with heterogeneous feature spaces and severe instance missingness. Current methods typically project disparate views into a shared Euclidean space and rely on static graphs to preserve local structures. However, these point-to-point embeddings ignore intrinsic metric discrepancies across modalities and suffer from topological distortion when samples are missing. To address this, we propose **I**ncomplete **G**romov-**W**asserstein **A**lignment (**IGWA**), a novel optimal transport framework for IMVC. IGWA abandons forced Euclidean consensus representations; instead, it directly couples heterogeneous intra-view metric spaces by learning a probabilistic transport plan. Building on established partial Gromov–Wasserstein transport, we incorporate view-specific partial couplings into a transport-aware global consensus framework, allowing reliable structural mass to be aligned without explicit data imputation. By matching intra-view distance matrices rather than raw features, IGWA preserves geometric invariants and maintains robustness against missing nodes. Furthermore, we establish relabeling invariance and balanced recovery of the partial alignment, together with uniqueness, boundedness, and perturbation stability of the consensus update under fixed transport plans and view weights. Extensive experiments on diverse benchmarks demonstrate that IGWA consistently outperforms state-of-the-art IMVC methods.

**Code** : `https://anonymous.4open.science/r/8CB6`.

## 1 INTRODUCTION

Incomplete multi-view clustering (IMVC)Zhong et al. (2026); Wang et al. (2026); Xu et al. (2026) aims to discover the underlying group structure from multiple views when only a subset of instances is observed in each view. The missing-view setting is challenging because the available modalities may differ substantially in both feature space and sample coverage. As illustrated in Fig. 1, a reliable IMVC method simultaneously address the following two issues:

- ***Metric Incompatibility across Heterogeneous Spaces.*** Different views may have different feature dimensions, scales, and semantic meanings. Directly projecting them into a shared Euclidean space and enforcing point-to-point correspondence implicitly assumes that all views share a compatible metric structure, which is generally unrealisticDing et al. (2026).
- ***Topological Distortion under Structural Incompleteness.*** Missing instances break the local neighborhood relationships within individual views. Consequently, graphs constructed from incomplete observations may contain fragmented or unreliable manifoldsLiu et al. (2026); Gu et al. (2026b). Rigidly aligning such graphs can propagate structural noise and produce sub-optimal clustering boundariesJiang et al. (2025); Dong et al. (2025).

**Graph-Regularized Incomplete Multi-View Clustering:** A dominant line of IMVC research combines matrix factorization or subspace learning with graph regularization to preserve local manifold structures Li et al. (2014); Zhao et al. (2016); Wen et al. (2020); Wang et al. (2019); Li et al. (2021); Xu et al. (2018); Liu et al. (2022); Gu et al. (2024); Wen et al. (2025). These methods typically construct a view-specific $K$-nearest-neighbor graph and learn view-specific latent representations together with a global consensus representation. Specifically, let the observed data in the $v$-th view be denoted by $\mathbf{Y}^{(v)} \in \mathbb{R}^{d_v \times N_v}$, where $N_v$ is the number of observed instances. Let $\mathbf{Z}^{(v)} \in \mathbb{R}^{k \times N_v}$



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.

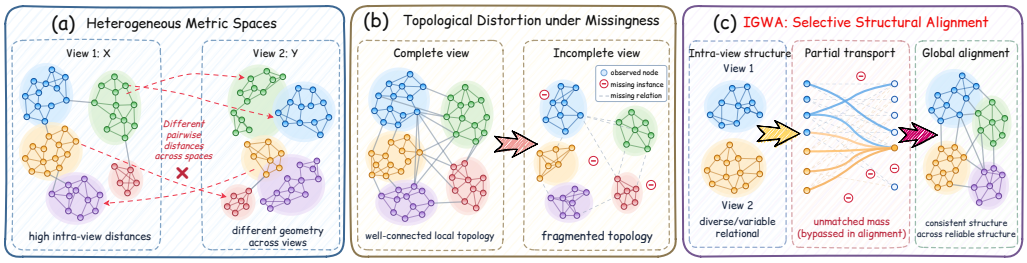



Figure 1: Motivation of IGWA for structure-aware alignment under heterogeneous and incomplete multi-view data.

be the view-specific representation and $\mathbf{H} \in \mathbb{R}^{k\times N}$ be the consensus representation over all $N$ unique instances. Given a missing-indicator matrix $\mathbf{M}^{(v)} \in \{0,1\}^{N_v\times N}$ and a graph Laplacian $\mathbf{L}^{(v)} \in \mathbb{R}^{N_v\times N_v}$, a generalized objective of this family can be written as

$$\min_{\{\mathbf{Z}^{(v)},\mathbf{W}^{(v)}\},\mathbf{H}} \sum_{v=1}^{V} \theta_v \left( \underbrace{\|\mathbf{Y}^{(v)} - \mathbf{W}^{(v)}\mathbf{Z}^{(v)}\|_{\mathcal{F}}^2}_{\text{Reconstruction}} + \alpha \underbrace{\operatorname{Tr}\left(\mathbf{Z}^{(v)}\mathbf{L}^{(v)}\mathbf{Z}^{(v)\top}\right)}_{\text{Manifold Preservation}} + \beta \underbrace{\|\mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{H}\|_{\mathcal{F}}^2}_{\text{Euclidean Consensus Alignment}} \right), \tag{1}$$

where $\mathbf{W}^{(v)}$ is a view-specific basis matrix, $\theta_v$ is an adaptive view weight, and $\alpha, \beta > 0$ are trade-off parameters. However, the Euclidean consensus and graph regularization terms assume shared metrics and reliable neighborhoods, assumptions violated by heterogeneous and incomplete views, thereby motivating a metric-free structural alignment mechanism.

**Gromov-Wasserstein Optimal Transport:** Optimal transport provides a principled framework for comparing probability measures and learning soft correspondences Xu et al. (2021; 2019a). Standard Wasserstein distance, however, requires the compared distributions to be embedded in the same metric space. Gromov-Wasserstein (GW) distance removes this requirement by comparing pairwise distances within each space Luo et al. (2022); Klein et al. (2024). Let $(\mathcal{X}, d_{\mathcal{X}}, \mathbf{p})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mathbf{q})$ be two metric-measure spaces, with intra-space distance matrices $\mathbf{C}^{(x)}$ and $\mathbf{C}^{(y)}$, and empirical marginal distributions $\mathbf{p}$ and $\mathbf{q}$. The entropically regularized discrete GW discrepancy is defined as

$$\begin{aligned} \mathcal{GW}_\epsilon\left(\mathbf{C}^{(x)}, \mathbf{C}^{(y)}, \mathbf{p}, \mathbf{q}\right) &= \min_{\mathbf{T}\in\Pi(\mathbf{p},\mathbf{q})} \sum_{i,j=1}^{N_x}\sum_{k,l=1}^{N_y} \mathcal{L}\left(\mathbf{C}^{(x)}_{i,j}, \mathbf{C}^{(y)}_{k,l}\right)\mathbf{T}_{i,k}\mathbf{T}_{j,l} - \epsilon\mathcal{H}(\mathbf{T}), \\ \text{s.t.} \quad \Pi(\mathbf{p},\mathbf{q}) &= \left\{\mathbf{T} \in \mathbb{R}_+^{N_x\times N_y} \mid \mathbf{T}\mathbf{1}_{N_y} = \mathbf{p},\ \mathbf{T}^\top\mathbf{1}_{N_x} = \mathbf{q}\right\} \end{aligned} \tag{2}$$

where $\mathbf{W}^{(v)}$ is a view-specific basis matrix, $\theta_v$ is an adaptive view weight, and $\alpha, \beta > 0$ are trade-off parameters. However, the Euclidean consensus and graph regularization terms assume shared metrics and reliable neighborhoods, assumptions violated by heterogeneous and incomplete views, thereby motivating a metric-free structural alignment mechanism.

To address these limitations, we propose **I**ncomplete **G**romov-**W**asserstein **A**lignment (**IGWA**). ***First*,** IGWA constructs an intra-view metric matrix from the observed samples of each view, avoiding the need to project heterogeneous modalities into a common Euclidean space. ***Then*,** building on established partial Gromov–Wasserstein transport, it aligns each intra-view metric matrix with a learnable global consensus structure through a partial Gromov–Wasserstein transport plan. ***Finally*,** by relaxing the marginal constraints, the transport plan selectively matches reliable structural mass while ignoring fragments caused by missing instances. Consequently, IGWA addresses both metric incompatibility across modalities and topological distortion under incomplete observations without explicitly imputing missing features.

The main contributions of this work are summarized as follows:

- ***Unified Partial Transport Framework.*** Building on established partial GW transport, IGWA formulates incomplete multi-view clustering as a cohesive optimal transport prob-

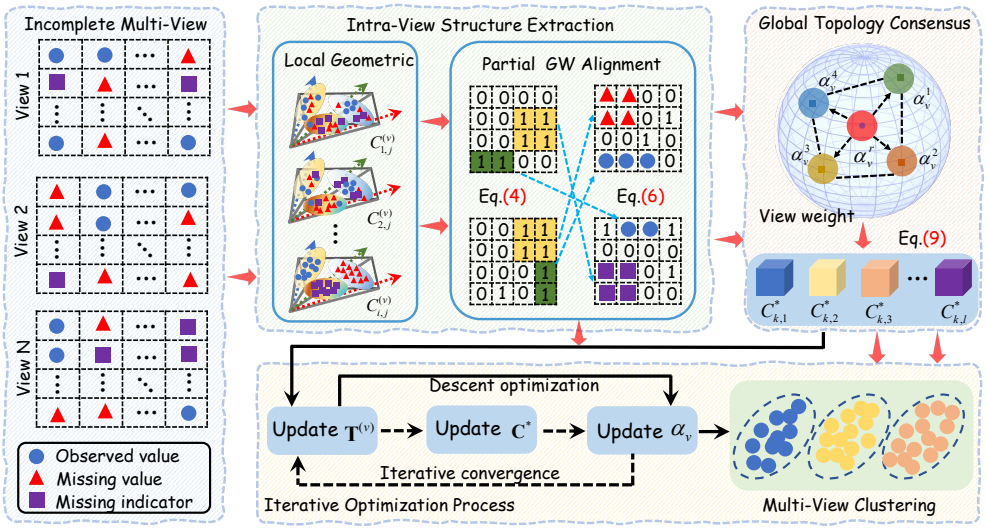



Figure 2: Overview of the proposed IGWA framework. Incomplete multi-view data are encoded into intra-view distance matrices, partially aligned via Gromov–Wasserstein transport, and iteratively integrated into a global topology consensus.

lem, seamlessly distilling reliable view-specific partial couplings into a dynamically weighted global topological consensus.

- ***Transport-Support-Aware Consensus.*** We characterize how partial transport and consensus regularization jointly determine the contribution of each aligned relation. The closed-form update applies stronger shrinkage to entries with less transport support, and we establish uniqueness, boundedness, and stability of the consensus update for fixed plans and weights.
- ***Imputation-Free Topological Robustness.*** We learn probabilistic transport plans under relaxed marginal constraints, thereby achieving robustness to missing instances and fragmented local topologies without explicit data imputation.

## 2 Related Work

### 2.1 Optimal Transport and Gromov–Wasserstein Alignment

Optimal transport (OT) compares probability measures by minimizing a ground cost under marginal constraints. Entropic regularization made discrete OT scalable through Sinkhorn or Bregman projections Cuturi (2013); Benamou et al. (2015), enabling applications such as domain adaptation Courty et al. (2017). Unbalanced OT further replaces exact mass preservation with divergence penalties Chizat et al. (2018). These formulations nevertheless require a meaningful cross-domain ground cost and are therefore unsuitable when views have incomparable feature coordinates. Gromov–Wasserstein (GW) transport instead compares within-domain pairwise relations and is invariant to measure-preserving isometries Mémoli (2011). Entropic GW supports efficient soft correspondence Solomon et al. (2016), while GW barycenters aggregate distance or kernel matrices Peyré et al. (2016). Subsequent work fused structural and feature costs Vayer et al. (2019) and developed GW objectives for graph matching, node embedding, and scalable graph partitioning Xu et al. (2019b;a). Most relevantly, GWMAC learns a kernel-based GW barycenter and uses modularity to cluster multimodal data with unknown cross-modal correspondence Gong et al. (2022).

GWMAC jointly performs multimodal alignment and clustering through a balanced GW barycenter with adaptive modality weights Gong et al. (2022). IGWA uses partial couplings with view-specific transported masses to integrate observed relational structures into an $N \times N$ consensus matrix. These couplings permit some mass to remain unmatched, while consensus regularization controls the

magnitude of the aggregated relations. The resulting formulation addresses incomplete observations without explicitly reconstructing missing feature vectors.

### 2.2 Partial and Relaxed Gromov–Wasserstein Transport

Partial optimal transport replaces full mass conservation by transporting a prescribed mass; its mathematical foundations were developed by Caffarelli and McCann and by Figalli Caffarelli & McCann (2010); Figalli (2010). Chapel et alChapel et al. (2020) extended partial transport to Wasserstein and GW objectives, proposed exact procedures, and demonstrated partial GW for positive–unlabeled learning across heterogeneous domains. Related relaxations treat mass differently: unbalanced GW penalizes marginal deviations, whereas semi-relaxed GW fixes one marginal and optimizes the other Séjourné et al. (2021); Vincent-Cuaz et al. (2022). Bai et alBai et al. (2025b) established a partial GW metric, analyzed its barycenters, and derived Frank–Wolfe solvers; their linear embedding reduces repeated pairwise PGW comparisons Bai et al. (2025a). Multi-marginal GW uses a higher-order joint coupling, while distributional reduction learns compact GW prototypes for clustering or dimensionality reduction Beier et al. (2023); Assel et al. (2025).

IGWA builds on established partial GW formulations Chapel et al. (2020); Bai et al. (2025b) and incorporates them into a joint incomplete multi-view consensus objective with adaptive view weighting and quadratic consensus regularization. For fixed plans and weights, Eq. 9 reduces to the unregularized barycenter update when $\gamma = 0$ and its denominator is positive. With $\gamma > 0$, relations receiving less transported support undergo stronger shrinkage. Our contribution lies in coupling view-specific partial transport with a support-aware global consensus, rather than applying partial GW as an isolated alignment module.

## 3 Methodology

Fig. 2 illustrates the proposed IGWA pipeline, including intra-view structure extraction, partial Gromov–Wasserstein coupling, global topology consensus learning and iterative optimization.

### 3.1 Partial Gromov-Wasserstein Coupling

**Remark 1.** *Raw-space imputation may amplify noise and impose artificial correspondence across heterogeneous views. We instead define the relational lifting* $\mathbf{Y}^{(v)} \in \mathbb{R}^{d_v \times N_v} \longmapsto \mathbf{C}^{(v)} \in \mathbb{R}^{N_v \times N_v}$ *by* $\mathbf{C}^{(v)}_{i,j} \propto \exp\Big(-\|\mathbf{y}^{(v)}_i - \mathbf{y}^{(v)}_j\|_2^2/(2\sigma_v^2)\Big)$*, with* $\sum_{i=1}^{N_v}\sum_{j=1}^{N_v}\mathbf{C}^{(v)}_{i,j} = 1$*. Thus, variations in* $d_v$*,* $N_v$*, and feature scales are absorbed into normalized intra-view geometry, reducing scale-induced bias and enabling relational coupling without direct raw-space comparison:*

$$\mathbf{C}^{(v)}_{i,j} = \frac{\exp\left(-\|\mathbf{y}^{(v)}_i - \mathbf{y}^{(v)}_j\|_2^2/2\sigma_v^2\right)}{\sum_{k=1}^{N_v}\sum_{l=1}^{N_v}\exp\left(-\|\mathbf{y}^{(v)}_k - \mathbf{y}^{(v)}_l\|_2^2/2\sigma_v^2\right)} \tag{3}$$

where $\sigma_v$ controls the kernel scale of view $v$. Standard Gromov–Wasserstein transport requires $\mathbf{T}^{(v)}\mathbf{1}_N = \mathbf{p}^{(v)}$ and $\mathbf{T}^{(v)\top}\mathbf{1}_{N_v} = \mathbf{q}$, implying $\mathbf{1}_{N_v}^\top \mathbf{T}^{(v)}\mathbf{1}_N = \|\mathbf{p}^{(v)}\|_1 = \|\mathbf{q}\|_1$ and thus complete mass participation. Under instance missingness, typically $N_v < N$, such balanced coupling cannot discard unmatched or fragmented relations. We instead optimize $\mathbf{T}^{(v)} \in \mathbb{R}_+^{N_v \times N}$ under $\mathbf{T}^{(v)}\mathbf{1}_N \le \mathbf{p}^{(v)}$ and $\mathbf{T}^{(v)\top}\mathbf{1}_{N_v} \le \mathbf{q}$, while retaining $\mathbf{1}_{N_v}^\top \mathbf{T}^{(v)}\mathbf{1}_N = s^{(v)}$. Hence, only mass $s^{(v)}$ contributes to aligning $\mathbf{C}^{(v)} \in \mathbb{R}^{N_v \times N_v}$ with the learnable consensus $\mathbf{C}^* \in \mathbb{R}^{N \times N}$. The resulting alignment loss is defined as:

$$\mathcal{L}_{\text{IGW}}\big(\mathbf{C}^{(v)}, \mathbf{C}^*, \mathbf{T}^{(v)}\big) = \sum_{i,j=1}^{N_v}\sum_{k,l=1}^{N}\Big(\mathbf{C}^{(v)}_{i,j} - \mathbf{C}^*_{k,l}\Big)^2 \mathbf{T}^{(v)}_{i,k}\mathbf{T}^{(v)}_{j,l} + \eta\mathcal{H}\big(\mathbf{T}^{(v)}\big) \tag{4}$$

where $\mathcal{H}(\mathbf{T}^{(v)}) = \sum_{i,k}\mathbf{T}^{(v)}_{i,k}(\log \mathbf{T}^{(v)}_{i,k} - 1)$ is the entropy regularization parameterized by $\eta > 0$ to promote computational tractability. The feasible set of the transport plan is relaxed to $\Pi_{\le}\big(\mathbf{p}^{(v)}, \mathbf{q}\big)$:

$$\Pi_{\le}\big(\mathbf{p}^{(v)}, \mathbf{q}\big) = \Big\{\mathbf{T}^{(v)} \in \mathbb{R}_+^{N_v \times N} \;\Big|\; \mathbf{T}^{(v)}\mathbf{1}_N \le \mathbf{p}^{(v)},\ \mathbf{T}^{(v)\top}\mathbf{1}_{N_v} \le \mathbf{q},\ \mathbf{1}_{N_v}^\top \mathbf{T}^{(v)}\mathbf{1}_N = s^{(v)}\Big\} \tag{5}$$

Here, $\mathbf{p}^{(v)} \in \Delta_{N_v}$ and $\mathbf{q} \in \Delta_N$ are prior marginals, typically $\mathbf{p}^{(v)} = \frac{1}{N_v}\mathbf{1}_{N_v}$ and $\mathbf{q} = \frac{1}{N}\mathbf{1}_N$. The admissible range $0 < s^{(v)} \leq \min\big(\|\mathbf{p}^{(v)}\|_1, \|\mathbf{q}\|_1\big)$ determines the matched mass, leaving $\|\mathbf{p}^{(v)}\|_1 - s^{(v)}$ and $\|\mathbf{q}\|_1 - s^{(v)}$ untransported.

### 3.2 Unified Topology Integration

**Remark 2.** *Rather than separating local alignment from consensus recovery, IGWA jointly optimizes $\big(\{\mathbf{T}^{(v)}\}_{v=1}^V, \mathbf{C}^*, \boldsymbol{\alpha}\big)$ within a unified variational objective. Each view contributes $\alpha_v^r \mathcal{L}_{\mathrm{IGW}}\big(\mathbf{C}^{(v)}, \mathbf{C}^*, \mathbf{T}^{(v)}\big)$, while $\boldsymbol{\alpha}^\top \mathbf{1} = 1$ and $\alpha_v \geq 0$ constrain adaptive view weighting. Thus, $\sum_{v=1}^V \alpha_v^r \mathcal{L}_{\mathrm{IGW}}$ aggregates localized couplings, whereas $\gamma\|\mathbf{C}^*\|_F^2$ controls global consensus complexity. This yields $\mathbf{C}^*$ for final clustering:*

$$\min_{\{\mathbf{T}^{(v)}\},\mathbf{C}^*,\boldsymbol{\alpha}} \mathcal{J} = \sum_{v=1}^V \alpha_v^r \mathcal{L}_{\mathrm{IGW}}\big(\mathbf{C}^{(v)}, \mathbf{C}^*, \mathbf{T}^{(v)}\big) + \gamma\|\mathbf{C}^*\|_F^2, \quad s.t. \quad \boldsymbol{\alpha}^T\mathbf{1} = 1,\ \alpha_v \geq 0 \tag{6}$$

where $r > 1$ controls the weight distribution and $\gamma > 0$ regularizes $\mathbf{C}^*$.

**Theorem 1** (**Relabeling invariance and balanced recovery**). *For permutation matrices $\mathbf{P}_v, \mathbf{P}$, set $\widetilde{\mathbf{C}}^{(v)} = \mathbf{P}_v\mathbf{C}^{(v)}\mathbf{P}_v^\top$, $\widetilde{\mathbf{C}}^* = \mathbf{P}\mathbf{C}^*\mathbf{P}^\top$, and $\widetilde{\mathbf{T}}^{(v)} = \mathbf{P}_v\mathbf{T}^{(v)}\mathbf{P}^\top$. Then feasibility and $\mathcal{L}_{\mathrm{IGW}}$ are preserved under $\widetilde{\mathbf{p}}^{(v)} = \mathbf{P}_v\mathbf{p}^{(v)}$ and $\widetilde{\mathbf{q}} = \mathbf{P}\mathbf{q}$. If $s^{(v)} = \|\mathbf{p}^{(v)}\|_1 = \|\mathbf{q}\|_1$, both marginal inequalities saturate, so Eq. 4 recovers Eq. 2 for squared loss and the same entropy convention. The Euclidean term in Eq. 1 lacks invariance to independent view transforms: if $\mathbf{Z}^{(v)}\mathbf{M}^{(v)} = \mathbf{H} \neq \mathbf{0}$, its penalty changes from 0 to $4\|\mathbf{H}\|_{\mathcal{F}}^2$ under $\mathbf{Z}^{(v)} \mapsto -\mathbf{Z}^{(v)}$.* ***(Proof see appendix §D)***

For fixed plans, let $\mathbf{q}^{(v)} = \mathbf{T}^{(v)\top}\mathbf{1}_{N_v}$ denote the induced consensus mass. Expanding the quadratic loss and removing terms independent of $\mathbf{C}^*$ gives:

$$\mathcal{J}(\mathbf{C}^*) \Leftrightarrow \min_{\mathbf{C}^*} \sum_{v=1}^V \alpha_v^r \left( \sum_{i=1}^{N_v}\sum_{j=1}^{N_v}\sum_{k=1}^{N}\sum_{l=1}^{N} \Big( \big(\mathbf{C}_{i,j}^{(v)}\big)^2 + \big(\mathbf{C}_{k,l}^*\big)^2 - 2\mathbf{C}_{i,j}^{(v)}\mathbf{C}_{k,l}^* \Big) \mathbf{T}_{i,k}^{(v)}\mathbf{T}_{j,l}^{(v)} \right) + \gamma\|\mathbf{C}^*\|_F^2 \tag{7}$$

To accurately obtain the closed-form solution, we observe that the objective function is element-wise separable across each entry $(k, l)$ of the matrix $\mathbf{C}^*$. By rewriting the trace operations back into their scalar summation forms, we can decouple the global matrix optimization into $N \times N$ independent scalar optimization subproblems:

$$\mathcal{J}(\mathbf{C}_{k,l}^*) = \min_{\mathbf{C}_{k,l}^*} \sum_{k,l=1}^{N} \left[ \sum_{v=1}^V \alpha_v^r \left( \big(\mathbf{C}_{k,l}^*\big)^2 \mathbf{q}_k^{(v)}\mathbf{q}_l^{(v)} - 2\mathbf{C}_{k,l}^* \big(\mathbf{T}^{(v)T}\mathbf{C}^{(v)}\mathbf{T}^{(v)}\big)_{k,l} \right) + \gamma\big(\mathbf{C}_{k,l}^*\big)^2 \right] \tag{8}$$

By taking the partial derivative of the decoupled polynomial objective w.r.t. a specific element $\mathbf{C}_{k,l}^*$ and setting $\frac{\partial \mathcal{J}}{\partial \mathbf{C}_{k,l}^*} = 0$, we obtain the explicit closed-form update:

$$\mathbf{C}_{k,l}^* = \frac{\sum_{v=1}^V \alpha_v^r \Big(\mathbf{T}^{(v)T}\mathbf{C}^{(v)}\mathbf{T}^{(v)}\Big)_{k,l}}{\gamma + \sum_{v=1}^V \alpha_v^r \mathbf{q}_k^{(v)}\mathbf{q}_l^{(v)}} \tag{9}$$

**Theorem 2** (**Unique, bounded, and stable consensus**). *In the setting of Theorem 1, fix feasible $\{\mathbf{T}^{(v)}\}$ and $\boldsymbol{\alpha}$. If $\gamma > 0$ and $0 \leq C_{ij}^{(v)} \leq M$ with $M > 0$, Eqs. 7–8 have diagonal Hessian entries $2(\gamma + a_{kl}) > 0$; hence Eq. 9 is the unique minimizer. Moreover, $0 \leq b_{kl} \leq Ma_{kl}$ implies $0 \leq C_{kl}^* \leq Ma_{kl}/(\gamma + a_{kl}) < M$. For perturbed relations $\widehat{\mathbf{C}}^{(v)}$ with the same plans and weights, $\|\mathbf{C}^* - \widehat{\mathbf{C}}^*\|_\infty \leq \max_v \|\mathbf{C}^{(v)} - \widehat{\mathbf{C}}^{(v)}\|_\infty$. The relabeling in Theorem 1 also yields $\widetilde{\mathbf{C}}^* = \mathbf{P}\mathbf{C}^*\mathbf{P}^\top$.*

### 3.3 Optimization Strategy

To optimize the non-convex joint objective $\mathcal{J}$ in Eq.6 , we employ Block Coordinate Descent (BCD), iteratively updating one variable block while fixing the others.

***Step 1: Update the*** $\mathbf{T}^{(v)}$**.** Fixing the consensus structure $\mathbf{C}^*$ and view weights $\boldsymbol{\alpha}$, the objective function decouples into $V$ independent partial Gromov-Wasserstein optimization problems. For the $v$-th view, utilizing the properties of matrix traces, we can perform a multi-step equivalent expansion of the quartic GW discrepancy:

$$\min_{\mathbf{T}^{(v)}\in\Pi_{\leq}} \sum_{i,j,k,l} \Big(\mathbf{C}^{(v)}_{i,j} - \mathbf{C}^{*}_{k,l}\Big)^2 \mathbf{T}^{(v)}_{i,k}\mathbf{T}^{(v)}_{j,l} + \eta\mathcal{H}(\mathbf{T}^{(v)}), \quad \text{where} \quad \mathbf{M}^{(v)}_{(t)} = \mathbf{L}\Big(\mathbf{T}^{(v)}_{(t-1)}\Big) \tag{10}$$

where $\odot$ denotes the Hadamard product. Since the objective remains non-convex w.r.t. $\mathbf{T}^{(v)}$, we adopt a safeguarded successive linearization strategy. At iteration $t$, we compute a local linear approximation of the cost tensor based on the previous plan $\mathbf{T}^{(v)}_{(t-1)}$. By taking the gradient of the trace formulation, the local cost matrix $\mathbf{M}^{(v)} \in \mathbb{R}^{N_v\times N}$ is derived as:

$$\mathbf{M}^{(v)} = 2\Big[(\mathbf{C}^{(v)}\odot\mathbf{C}^{(v)})\mathbf{T}^{(v)}_{(t-1)}\mathbf{1}_N\mathbf{1}_N^\top + \mathbf{1}_{N_v}\mathbf{1}_{N_v}^\top\mathbf{T}^{(v)}_{(t-1)}(\mathbf{C}^*\odot\mathbf{C}^*) - 2\mathbf{C}^{(v)}\mathbf{T}^{(v)}_{(t-1)}\mathbf{C}^{*\top}\Big]. \tag{11}$$

Given $\mathbf{M}^{(v)}$, the subproblem reduces to a standard Entropic Partial Optimal Transport problem, which is efficiently solved using generalized Sinkhorn scaling algorithms.

***Step 2: Update the*** $\mathbf{C}^*$**.** Fixing the transport plans $\{\mathbf{T}^{(v)}\}$ and weights $\boldsymbol{\alpha}$, the objective function w.r.t. $\mathbf{C}^*$ is strictly convex. This subproblem aligns identically with the topological barycenter learning addressed in Sec 3.2. We directly apply the derived element-wise closed-form solution:

$$\mathbf{C}^*_{k,l} = \frac{\sum_{v=1}^V \alpha_v^r \Big(\mathbf{T}^{(v)T}\mathbf{C}^{(v)}\mathbf{T}^{(v)}\Big)_{k,l}}{\gamma + \sum_{v=1}^V \alpha_v^r \mathbf{q}_k^{(v)}\mathbf{q}_l^{(v)}} \tag{12}$$

This closed-form update ensures a monotonic decrease of the structural loss, completely avoiding gradient-based iterative approximations.

***Step 3: Update*** $\boldsymbol{\alpha}$. Fixing $\{\mathbf{T}^{(v)}\}$ and $\mathbf{C}^*$, let $E_v = \mathcal{L}_{\text{IGW}}\big(\mathbf{C}^{(v)}, \mathbf{C}^*, \mathbf{T}^{(v)}\big)$ denote the converged structural alignment loss for the $v$-th view. We introduce a Lagrangian multiplier $\lambda$ to formulate the unconstrained objective $\mathcal{L}(\boldsymbol{\alpha},\lambda) = \sum_{v=1}^V \alpha_v^r E_v - \lambda(\sum_{v=1}^V \alpha_v - 1)$. By setting the partial derivative to zero and substituting it into the simplex constraint, we derive the optimal weights through the following deduction:

$$\alpha_v = \begin{cases} \frac{E_v^{-1/(r-1)}}{\sum_{m=1}^V E_m^{-1/(r-1)}}, & \min_m E_m > 0, \\ \mathbf{1}_{\{v=j^*\}}, & \min_m E_m \leq 0, \end{cases} \quad j^* = \min\left(\arg\min_{1\leq m\leq V} E_m\right). \tag{13}$$

Where, Eq. 13 favors reliable views with smaller errors $E_v$, isolating the global barycenter from degraded modalities.

# 4 Experiments

## 4.1 Experimental Settings

**Metrics.** We evaluate clustering performance using three widely used metrics: Accuracy (ACC), Normalized Mutual Information (NMI), and Purity (PUR).

Table 1: Detailed statistics of the six multi-view datasets.

| Dataset | # Samples | # Views | # Classes | # Features |
|---|---|---|---|---|
| YouTubeFace | 101499 | 5 | 31 | 640 / 944 / 576 / 512 / 640 |
| Hdigit | 10000 | 2 | 10 | 274 / 784 |
| Caltech7 | 1474 | 6 | 7 | 48 / 40 / 254 / 1984 / 512 |
| Fashion | 10000 | 3 | 10 | 784 / 512 / 512 |
| Animal | 10158 | 2 | 50 | 4096 / 4096 |
| CIFAR 10 | 10000 | 3 | 10 | 512 / 2000 / 512 |

**Datasets.** We use six benchmark datasets in Table 1. The missing ratio $\rho \in \{0.1, 0.3, 0.5, 0.7\}$. At each $\rho$, all methods use the same 30 masks; masks are independently sampled across $\rho$, so performance need not be monotonic.

**Baselines.** We compare IGWA with eight representative IMVC methods: DSIMVC Tang & Liu (2022); ICMVC Chao et al. (2024) and RPCIC Yuan et al. (2024); SIIHPC Yu et al. (2025), FIML Qin et al. (2025), CPMN Wang et al., and GUITAR Zhong et al. (2025); and CAMEL Gu et al. (2026a).

Table 2: Clustering performance under different missing rates on six datasets.

| Dataset | Method | Source | 0.1 ACC | 0.1 NMI | 0.1 PUR | 0.3 ACC | 0.3 NMI | 0.3 PUR | 0.5 ACC | 0.5 NMI | 0.5 PUR | 0.7 ACC | 0.7 NMI | 0.7 PUR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YouTubeFace | DSIMVC | ICML '22 | 36.47±0.48 | 42.20±0.63 | 55.63±0.55 | 42.51±1.69 | 48.29±2.03 | **66.67±1.48** | 34.90±0.67 | 49.88±1.01 | 46.11±0.43 | 36.44±1.54 | 50.99±1.80 | 58.30±1.36 |
| | ICMVC | AAAI '24 | 33.23±1.08 | 45.82±0.56 | 54.40±1.15 | 38.91±1.05 | 50.87±1.43 | 49.03±0.53 | 44.72±1.60 | 50.69±2.14 | **61.86±0.49** | 39.89±1.02 | 64.94±0.79 | 56.88±1.17 |
| | RPCIC | ACM MM '24 | 28.40±0.53 | 34.54±2.04 | **60.43±2.04** | 48.56±2.11 | **72.79±1.10** | 51.83±0.64 | 42.06±0.52 | 49.50±0.97 | 51.05±1.36 | 41.73±0.74 | **69.98±1.55** | 55.88±1.12 |
| | SIIHPC | ICLR '25 | 38.12±0.50 | **50.00±2.35** | 44.37±1.95 | 25.10±1.50 | 50.30±1.99 | 36.77±1.40 | 37.25±0.61 | 40.68±2.22 | 46.45±1.11 | **65.08±2.17** | 61.44±1.38 | **69.66±1.22** |
| | FIML | ICML '25 | 38.78±2.06 | 27.86±1.09 | 42.66±1.82 | 48.68±1.31 | 42.90±1.21 | 52.88±2.30 | 47.11±1.54 | 59.93±1.78 | 50.43±2.17 | 40.38±1.12 | 66.84±0.42 | 55.49±0.51 |
| | CPMN | NeurIPS '25 | **49.70±0.84** | 45.22±2.13 | 55.09±1.04 | 36.21±0.36 | 43.07±1.66 | 65.04±1.55 | 36.32±0.80 | **73.45±0.71** | 40.42±2.07 | 48.50±1.27 | 44.50±0.91 | 50.12±0.39 |
| | GUITAR | NeurIPS '25 | 37.69±0.94 | 33.75±2.34 | 39.58±1.74 | 46.61±2.36 | 59.69±0.58 | 53.75±0.39 | **48.58±1.17** | 55.46±2.02 | 56.33±1.10 | 52.03±2.18 | 46.64±2.22 | 68.02±2.08 |
| | CAMEL | ICLR '26 | 43.03±1.38 | 30.72±0.56 | 53.57±0.82 | **55.96±1.75** | 57.16±1.56 | 61.07±1.28 | 46.13±0.96 | 40.44±2.24 | 51.15±1.78 | 41.20±1.35 | 41.50±2.13 | 44.89±0.47 |
| | **IGWA (Ours)** | – | **56.12±0.67** | **53.56±2.23** | **67.54±1.06** | **62.28±2.29** | **78.29±1.44** | **74.69±1.08** | **53.39±1.11** | **70.72±0.69** | **69.09±0.63** | **75.11±1.54** | **76.63±2.07** | **75.36±0.73** |
| | *Improvement* | – | +12.92% | +7.12% | +11.77% | +11.29% | +7.56% | +12.03% | +9.90% | – | +11.69% | +15.41% | +9.50% | +8.18% |
| Hdigit | DSIMVC | ICML '22 | 25.82±1.55 | 40.85±0.97 | 53.89±0.91 | 55.60±0.53 | 49.32±1.47 | 56.56±2.34 | 37.27±0.94 | 52.04±1.77 | 40.32±2.16 | 32.80±2.23 | 44.87±1.47 | 51.57±1.77 |
| | ICMVC | AAAI '24 | **48.53±1.52** | 45.13±1.04 | 52.87±0.73 | **63.98±0.58** | 40.61±0.84 | 69.33±0.95 | 30.77±1.53 | 61.28±1.81 | 58.56±0.53 | 28.48±2.30 | 43.36±1.78 | 48.79±2.10 |
| | RPCIC | ACM MM '24 | 35.60±2.36 | 47.98±0.32 | 45.94±1.23 | 49.63±1.36 | 54.26±0.88 | 61.34±0.73 | 30.81±1.74 | 65.58±1.49 | 35.18±0.39 | 46.94±2.16 | 56.94±0.95 | 69.57±1.80 |
| | SIIHPC | ICLR '25 | 39.00±0.99 | 29.16±0.96 | 56.52±2.02 | 34.36±0.62 | 41.77±0.42 | 57.14±0.59 | **54.57±2.20** | 44.33±1.02 | 55.03±1.80 | 46.43±2.10 | 53.03±0.88 | 50.25±2.29 |
| | FIML | ICML '25 | 23.62±1.57 | **56.69±1.98** | 28.40±0.84 | 37.53±1.89 | 42.83±1.98 | 51.85±0.43 | 40.65±1.53 | **72.80±2.19** | 45.58±0.68 | 25.49±1.32 | 59.10±2.11 | **70.69±0.35** |
| | CPMN | NeurIPS '25 | 40.81±1.11 | 37.96±1.20 | **58.77±1.54** | 43.79±1.95 | 40.97±1.13 | 45.01±1.76 | 53.62±2.15 | 48.09±2.08 | **60.68±2.00** | **49.50±0.58** | 63.78±1.65 | 66.67±0.74 |
| | GUITAR | NeurIPS '25 | 18.04±1.00 | 50.32±1.21 | 53.47±0.66 | 37.79±0.46 | 47.03±1.69 | **74.37±0.53** | 32.27±1.83 | 56.77±1.20 | 53.71±1.51 | 27.18±0.58 | **70.74±1.51** | 48.90±2.06 |
| | CAMEL | ICLR '26 | 21.04±0.52 | 50.96±0.51 | 49.46±1.42 | 55.05±1.01 | **56.56±0.59** | 62.19±0.82 | 33.94±1.65 | 45.95±2.26 | 42.65±0.40 | 47.92±2.28 | 69.37±0.59 | 52.94±1.65 |
| | **IGWA (Ours)** | – | **56.30±1.09** | **62.95±1.68** | **66.92±1.34** | **72.56±1.81** | **62.17±0.65** | **84.31±2.17** | **61.55±0.37** | **84.33±0.84** | **65.87±1.73** | **54.52±1.54** | **81.83±2.34** | **76.40±2.04** |
| | *Improvement* | – | +16.01% | +11.04% | +13.87% | +13.41% | +9.92% | +13.37% | +12.79% | +15.84% | +8.55% | +10.14% | +15.68% | +8.08% |
| Caltech7 | DSIMVC | ICML '22 | 36.40±1.23 | 58.07±0.95 | 65.32±2.03 | 40.53±0.73 | **60.28±0.85** | 54.27±0.65 | 48.57±1.64 | 37.57±2.35 | **74.75±0.95** | 37.62±2.19 | 46.00±1.83 | 48.99±0.42 |
| | ICMVC | AAAI '24 | 40.19±0.50 | 46.96±1.68 | **74.19±0.97** | 41.86±1.25 | 51.60±0.74 | **75.55±0.56** | 45.55±1.68 | 44.89±1.58 | 55.67±1.31 | 32.51±2.29 | 39.73±1.08 | **64.68±1.41** |
| | RPCIC | ACM MM '24 | 40.46±0.82 | 54.14±1.19 | 71.79±1.19 | **52.02±1.21** | 31.23±2.02 | 68.68±0.91 | 36.08±1.15 | 36.44±1.35 | 72.23±0.47 | 36.82±1.44 | 46.95±1.16 | 63.48±1.66 |
| | SIIHPC | ICLR '25 | 39.04±2.04 | 48.28±0.78 | 70.38±2.01 | 43.57±1.63 | 36.32±0.64 | 44.88±1.52 | 45.09±1.05 | 50.15±0.95 | 52.54±2.01 | 25.48±0.62 | 55.45±2.12 | 48.08±0.56 |
| | FIML | ICML '25 | 48.99±0.80 | 47.40±1.01 | 67.98±2.29 | 45.46±0.76 | 53.19±1.94 | 47.11±1.89 | **59.73±1.65** | 30.87±2.35 | 72.85±1.49 | **56.04±1.75** | 39.43±1.37 | 56.37±0.56 |
| | CPMN | NeurIPS '25 | 43.64±1.77 | 47.59±1.17 | 47.27±0.91 | 33.34±1.17 | 43.99±0.71 | 68.93±0.52 | 47.63±1.57 | 34.64±1.75 | 63.61±0.40 | 40.36±1.88 | 36.12±1.47 | 61.08±1.77 |
| | GUITAR | NeurIPS '25 | **54.70±0.34** | 51.75±0.61 | 66.23±2.30 | 29.06±1.06 | 55.36±1.45 | 47.43±0.94 | 35.47±0.53 | **56.00±0.53** | 69.94±0.87 | 46.97±0.52 | 61.29±0.93 | 53.47±1.09 |
| | CAMEL | ICLR '26 | 38.57±1.13 | **72.39±0.78** | 57.37±0.34 | 48.08±0.58 | 54.67±1.10 | 70.56±1.28 | 35.20±1.97 | 35.56±0.56 | 53.81±0.50 | 34.90±0.88 | **63.36±2.10** | 38.47±1.77 |
| | **IGWA (Ours)** | – | **50.14±0.88** | **84.07±1.53** | **81.74±1.67** | **57.25±0.88** | **66.53±1.25** | **82.04±0.41** | **66.02±1.36** | **61.14±1.45** | **86.18±1.23** | **62.50±0.42** | **72.86±0.87** | **69.80±1.16** |
| | *Improvement* | – | – | +16.13% | +10.18% | +10.05% | +10.37% | +8.59% | +10.53% | +9.18% | +15.29% | +11.53% | +14.99% | +7.92% |
| Fashion | DSIMVC | ICML '22 | 31.90±0.34 | **52.68±0.38** | 48.35±0.50 | 32.18±0.57 | 35.31±0.36 | 43.99±1.87 | 53.85±1.99 | 23.35±0.67 | 61.58±2.34 | 43.50±1.22 | 40.99±0.82 | 52.41±1.83 |
| | ICMVC | AAAI '24 | 44.87±1.82 | 39.54±0.99 | 55.05±1.61 | 20.13±0.51 | **55.81±1.49** | 48.40±1.14 | 31.61±2.23 | **53.08±0.69** | 40.93±1.61 | **50.17±1.52** | 32.98±1.88 | **60.90±0.37** |
| | RPCIC | ACM MM '24 | 29.05±2.35 | 23.05±1.20 | 51.54±2.21 | 29.38±1.21 | 37.48±2.02 | 32.55±1.74 | 57.19±0.96 | 36.33±0.72 | 59.34±2.18 | 30.10±0.95 | 38.97±1.13 | 47.91±0.73 |
| | SIIHPC | ICLR '25 | 39.59±0.78 | 30.71±1.19 | **73.45±1.21** | 28.39±1.23 | 28.68±1.81 | 35.36±0.94 | 38.06±2.02 | 42.02±1.29 | 42.34±0.40 | 39.13±2.35 | 36.93±1.92 | 51.53±0.75 |
| | FIML | ICML '25 | 45.22±1.50 | 30.49±1.86 | 59.75±2.08 | **32.69±1.37** | 51.16±1.13 | 45.12±0.79 | **57.91±1.03** | 38.74±0.51 | 58.63±2.35 | 35.25±1.75 | 34.09±1.89 | 38.87±1.49 |
| | CPMN | NeurIPS '25 | **56.50±2.26** | 31.26±0.45 | 61.69±2.28 | 26.56±0.70 | 41.90±1.97 | 45.27±2.31 | 52.45±1.92 | 37.36±2.11 | 69.25±0.51 | 44.62±2.19 | 56.45±1.42 | 48.05±0.78 |
| | GUITAR | NeurIPS '25 | 29.80±1.51 | 34.66±1.30 | 64.78±1.71 | 24.63±1.69 | 25.24±1.71 | 32.38±2.12 | 33.97±0.59 | 49.49±0.99 | 47.54±1.55 | 40.56±2.26 | **60.24±1.19** | 46.14±1.37 |
| | CAMEL | ICLR '26 | 44.06±1.34 | 31.09±0.96 | 58.23±1.34 | 19.62±0.37 | 47.22±2.17 | **50.52±1.38** | 34.66±0.33 | 34.06±0.37 | **70.72±1.15** | 37.42±1.24 | 37.72±1.36 | 51.22±2.29 |
| | **IGWA (Ours)** | – | **62.06±0.32** | **60.73±1.32** | **79.62±2.32** | **35.94±1.83** | **62.07±1.57** | **57.55±1.36** | **62.47±1.81** | **61.60±1.39** | **80.28±1.07** | **57.07±0.47** | **67.52±1.82** | **67.48±1.00** |
| | *Improvement* | – | +9.84% | +15.28% | +8.40% | +9.94% | +11.22% | +13.92% | +7.87% | +16.05% | +13.52% | +13.75% | +12.08% | +10.80% |
| Animal | DSIMVC | ICML '22 | 40.14±0.33 | 50.71±0.71 | 50.00±0.79 | 40.70±0.91 | **65.47±1.21** | 49.44±2.09 | 43.49±1.78 | **67.82±0.81** | 46.46±2.25 | 35.00±1.34 | 51.88±0.55 | 51.36±2.11 |
| | ICMVC | AAAI '24 | 48.90±1.14 | 57.19±0.69 | 73.25±0.53 | 38.22±1.34 | 57.24±0.66 | 57.71±2.30 | 49.04±2.35 | 40.76±0.33 | 70.27±2.16 | 41.73±1.91 | 44.67±1.66 | 45.12±0.86 |
| | RPCIC | ACM MM '24 | **60.10±1.23** | 37.85±0.88 | 61.73±2.03 | 27.67±0.61 | 49.80±0.52 | 68.60±2.00 | **53.76±0.71** | 52.44±0.85 | 57.06±1.81 | 54.17±0.55 | 43.42±0.36 | 58.69±2.36 |
| | SIIHPC | ICLR '25 | 47.89±0.58 | 61.05±0.37 | 49.00±1.24 | 44.38±0.45 | 58.45±1.34 | **74.34±0.95** | 27.62±2.07 | 48.90±1.25 | 45.08±0.90 | 46.24±0.65 | 46.30±0.58 | 51.53±0.50 |
| | FIML | ICML '25 | 38.57±0.91 | 43.15±0.66 | 74.58±0.71 | 32.37±1.84 | 53.29±0.49 | 59.77±1.62 | 42.84±1.05 | 42.38±2.16 | 43.00±2.13 | **64.08±0.38** | 43.76±2.33 | **71.45±1.15** |
| | CPMN | NeurIPS '25 | 35.00±1.17 | 62.08±2.04 | **77.74±2.31** | 41.78±2.33 | 40.49±0.38 | 60.00±1.56 | 48.44±0.77 | 62.09±1.34 | **72.42±1.31** | 34.75±0.43 | **70.95±1.81** | 56.96±1.93 |
| | GUITAR | NeurIPS '25 | 43.53±2.37 | **63.35±1.27** | 60.26±2.19 | 26.82±1.24 | 54.71±2.03 | 71.13±2.07 | 47.29±1.57 | 37.83±0.75 | 51.29±1.05 | 55.68±1.36 | 54.48±0.54 | 58.02±1.30 |
| | CAMEL | ICLR '26 | 30.61±2.22 | 55.81±0.50 | 53.98±1.60 | **46.88±1.94** | 57.27±0.86 | 70.69±1.36 | 34.05±0.45 | 51.17±0.51 | 46.00±1.18 | 45.48±1.68 | 50.14±1.16 | 51.47±1.26 |
| | **IGWA (Ours)** | – | **67.59±0.40** | **68.88±0.35** | **87.53±2.37** | **51.55±1.42** | **72.75±0.70** | **81.02±1.89** | **61.22±0.55** | **72.99±0.80** | **82.86±0.77** | **69.26±2.34** | **77.71±1.73** | **70.55±1.68** |
| | *Improvement* | – | +12.46% | +8.73% | +12.59% | +9.96% | +11.12% | +8.99% | +13.88% | +7.62% | +14.42% | +8.08% | +9.53% | – |
| CIFAR-10 | DSIMVC | ICML '22 | 49.82±0.64 | 66.26±2.19 | 63.43±2.02 | 59.09±2.03 | 66.95±0.75 | 69.30±1.04 | 24.72±1.39 | 29.46±0.55 | 36.13±1.77 | 37.38±1.10 | 19.04±1.22 | 62.38±1.74 |
| | ICMVC | AAAI '24 | 40.22±1.27 | 68.83±1.34 | 64.95±1.04 | 52.92±0.51 | 52.69±2.05 | 57.69±2.11 | 31.22±1.55 | 50.00±0.67 | 59.76±0.47 | 40.81±1.67 | 41.36±1.56 | 52.75±0.37 |
| | RPCIC | ACM MM '24 | 40.03±1.98 | 69.75±1.89 | 57.00±1.74 | 60.39±0.64 | 49.73±2.04 | 70.03±0.95 | 27.83±1.79 | **56.13±1.78** | 42.74±0.51 | 42.98±2.12 | 26.96±1.91 | **63.87±0.59** |
| | SIIHPC | ICLR '25 | **58.83±2.09** | 69.97±2.29 | **70.86±1.98** | 57.63±1.03 | 45.54±0.83 | **72.88±1.89** | **45.06±0.72** | 39.52±1.96 | 55.05±1.83 | **50.68±1.13** | 27.72±2.22 | 56.98±1.95 |
| | FIML | ICML '25 | 41.66±1.11 | **73.76±0.89** | 62.06±1.88 | 52.50±0.53 | 57.32±1.83 | 61.75±1.67 | 21.18±0.59 | 28.33±0.40 | 43.48±1.27 | 40.39±2.36 | 34.89±1.48 | 50.55±1.81 |
| | CPMN | NeurIPS '25 | 44.40±0.34 | 50.02±2.21 | 65.18±1.23 | 42.72±1.32 | 48.21±1.66 | 69.17±1.27 | 29.69±0.34 | 38.97±0.54 | 44.16±1.22 | 41.62±2.02 | 19.71±0.46 | 44.14±0.55 |
| | GUITAR | NeurIPS '25 | 45.43±1.10 | 43.28±1.31 | 55.22±1.35 | **64.98±1.96** | **70.23±0.54** | 67.97±0.65 | 20.63±1.00 | 37.19±1.61 | **63.88±1.37** | 35.57±1.95 | **45.31±2.27** | 44.06±1.13 |
| | CAMEL | ICLR '26 | 53.91±2.30 | 50.82±1.69 | 56.98±1.89 | 60.07±2.08 | 43.07±1.73 | 65.13±1.12 | 18.11±2.35 | 36.10±1.46 | 59.30±1.77 | 37.29±2.36 | 21.24±0.63 | 40.08±1.91 |
| | **IGWA (Ours)** | – | **68.14±0.66** | **84.12±0.89** | **79.66±2.11** | **71.61±1.13** | **76.86±0.42** | **82.26±1.54** | **50.36±0.47** | **60.42±1.46** | **68.86±0.38** | **56.53±1.19** | **50.30±1.77** | **70.68±0.62** |
| | *Improvement* | – | +15.83% | +14.05% | +12.42% | +10.20% | +9.44% | +12.87% | +11.76% | +7.64% | +7.80% | +11.54% | +11.01% | +10.66% |

## 4.2 Main Performance

Table 2 compares IGWA with eight representative methods on six datasets under four missing rates. We analyze the results from three perspectives:

- **Broad Superiority.** IGWA performs strongly across all six datasets despite their different scales, views, and class structures. Normalized intra-view relations reduce scale discrepancies, while partial structural alignment incorporates only reliable topology into the consensus, supporting consistent generalization.
- **Substantial Performance Margins.** The advantage over competing methods is sizeable rather than marginal. Relative to the runner-up, representative improvements reach $14.99\%$, $15.28\%$, $15.29\%$, $15.41\%$, $15.68\%$, $15.83\%$, $15.84\%$, $16.01\%$, $16.05\%$, and $16.13\%$. These margins appear across different datasets, metrics, and missing rates, showing that selective mass transport and adaptive weighting consistently prevent corrupted relations or weak views from distorting the learned consensus.
- **Consistent Metric Gains.** IGWA reports consistently strong ACC, NMI, and PUR values, indicating that its advantage is not tied to one criterion. With $\mathbf{T}^{(v)} \in \Pi_{\leq}(\mathbf{p}^{(v)}, \mathbf{q})$, $\mathbf{T}^{(v)}\mathbf{1}_N \leq \mathbf{p}^{(v)}$, $\mathbf{T}^{(v)\top}\mathbf{1}_{N_v} \leq \mathbf{q}$, and $\mathbf{1}_{N_v}^{\top}\mathbf{T}^{(v)}\mathbf{1}_N = s^{(v)}$, only reliable mass is matched; $\mathbf{C}^*$ then integrates aligned structures, while $\alpha_v \propto E_v^{1/(1-r)}$ reduces the influence of views with large $E_v$.

## 4.3 Ablation Study

**Effect of Partial Transport.** Table 3 shows that the full IGWA consistently outperforms balanced, semi-relaxed, unbalanced, and fixed-mass alternatives. The gains of $18.55\%$, $21.30\%$, and $17.33\%$ in ACC, NMI, and PUR indicate that observation-aware partial transport is more effective than

Table 3: Transport-constraint ablation on Caltech7 at a missing rate of 0.7.

| Variant | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| Balanced GW ($s^{(v)} = 1$) | 46.71 | 51.54 | 51.63 |
| Semi-relaxed GW | 49.52 | 55.51 | 55.28 |
| Unbalanced GW | 50.85 | 57.34 | 57.21 |
| Partial GW ($s^{(v)} = 0.7$) | 51.98 | **60.07** | **59.49** |
| Partial GW ($s^{(v)} = 0.5$) | **52.72** | 59.13 | 58.12 |
| **IGWA ($s^{(v)} = N_v/N$)** | **62.50** | **72.86** | **69.80** |
| **Improvement** | +18.55% | +21.30% | +17.33% |

Table 4: Transport-alignment ablation on Caltech7 at a missing rate of 0.7.

| Variant | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| Random mapping (fixed) | 33.21 | 36.81 | 35.48 |
| Nearest-neighbor mapping (fixed) | 44.13 | 47.51 | 49.12 |
| Identity mapping (fixed) | 50.05 | 56.12 | 56.02 |
| Learned $\mathbf{T}^{(v)}$ (random init.) | **53.91** | **58.80** | **58.74** |
| **IGWA (identity init.)** | **62.50** | **72.86** | **69.80** |
| **Improvement** | +15.93% | +23.92% | +18.83% |

enforcing complete or predetermined mass participation. By leaving unreliable structural mass unmatched, IGWA prevents missing-view fragments from contaminating the learned consensus.

**Effect of Learned Alignment.** Table 4 demonstrates that learning $\mathbf{T}^{(v)}$ is crucial for recovering cross-view structural correspondence. Fixed random, nearest-neighbor, and identity mappings cannot adapt to topology disrupted by missing observations, whereas identity-initialized IGWA progressively refines the coupling. Its gains of 15.93%, 23.92%, and 18.83% in ACC, NMI, and PUR confirm that learned transport repairs structural gaps while retaining a reliable connection between consensus indices and original samples. **(More ablation studies see §E)**

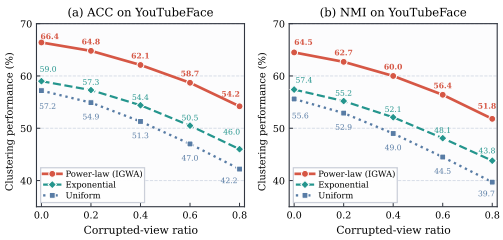



Figure 3: Ablation of view-weighting strategies on YouTubeFace at missing rate $\rho = 0.5$.

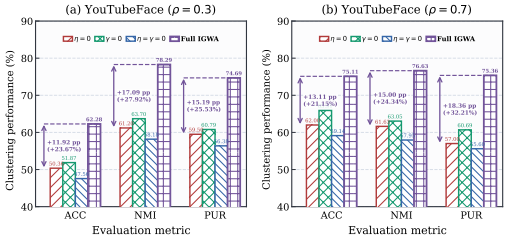



Figure 4: Ablation of the regularization components on YouTubeFace under missing rates $\rho = 0.3$ and $\rho = 0.7$.

**View-Weighting Strategies.** Fig 3 compares Power-law, Exponential and Uniform weighting as the corrupted-view ratio increases. Power-law weighting consistently gives the highest ACC and NMI, with its advantage widening under stronger corruption; at ratio 0.8, it improves ACC and NMI over exponential weighting by 17.83% and 18.26%, respectively. This confirms that Eq. (13) suppresses unreliable views without prematurely discarding their complementary structure.

**Regularization Components.** Fig. 4 shows that Full IGWA outperforms every reduced variant at both missing rates. Relative to $\eta = 0$, its gains range from 23.67% to 27.92% at $\rho = 0.3$ and from 21.15% to 32.21% at $\rho = 0.7$. Removing $\gamma$ also causes clear degradation, while removing both terms produces the weakest results. Thus, $\eta$ stabilizes partial coupling, whereas $\gamma$ shrinks weakly supported consensus entries.

**Relational Construction.** Table 5 compares six intra-view relational constructions on YouTubeFace. Normalized Gaussian similarity achieves the best ACC, NMI, and PUR, surpassing the strongest alternative by 19.68%, 18.92%, and 20.06%, respectively. Euclidean-distance variants remain sensitive to heterogeneous feature scales, whereas binary $k$-NN adjacency discards graded neighborhood information. Normalization further limits scale-driven dominance across views. These results identify relational lifting as a key source of improvement and confirm that normalized Gaussian relations preserve local topology more reliably for partial alignment and consensus learning.

Table 5: Ablation of relational construction on YouTubeFace at $\rho = 0.5$.

| Relational Construction | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| Raw Euclidean distance | 34.15 | 45.20 | 38.60 |
| Normalized Euclidean distance | 38.56 | 50.12 | 45.34 |
| Unnormalized Gaussian similarity | 42.34 | 55.30 | 51.60 |
| Binary $k$-NN adjacency | 40.18 | 53.45 | 49.12 |
| Cosine similarity | **44.61** | **59.47** | **57.55** |
| **Normalized Gaussian (IGWA)** | **53.39** | **70.72** | **69.09** |

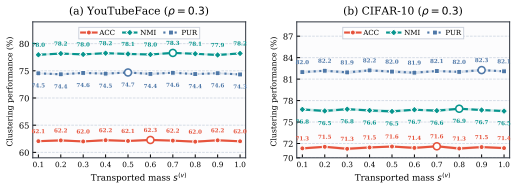



Figure 5: Sensitivity to the transported mass $s^{(v)}$ on YouTubeFace and CIFAR-10, $\rho = 0.3$.

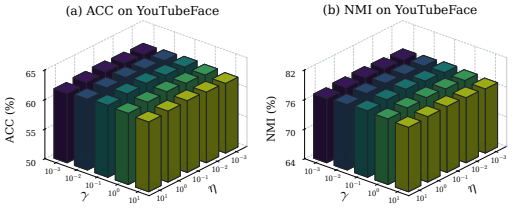



Figure 6: Joint sensitivity to $\eta$ and $\gamma$ on YouTubeFace at missing rate $\rho = 0.3$.

### 4.4 Sensitivity Analysis

**Transported Mass:** We vary the transported mass $s^{(v)}$ from 0.1 to 1.0. Fig. 5 shows that ACC, NMI, and PUR remain consistently high on both YouTubeFace and CIFAR-10, with only minor fluctuations throughout this interval. The broad performance plateau demonstrates that IGWA is insensitive to mass selection.

**Joint Regularization:** We jointly vary $\eta$ and $\gamma \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$. As shown in Fig 6, ACC and NMI change only slightly across all combinations on YouTubeFace. This smooth response shows that both regularizers remain balanced, preventing either from dominating optimization and allowing IGWA to maintain strong clustering performance without narrow parameter tuning.

Table 6: Sensitivity to the view-weight exponent $r$ on YouTubeFace at missing rate $\rho = 0.3$.

| $r$ | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| 1.1 | $61.85 \pm 0.43$ | $77.80 \pm 0.58$ | $74.25 \pm 0.36$ |
| 1.2 | $62.01 \pm 0.57$ | $77.95 \pm 0.34$ | $74.40 \pm 0.49$ |
| 1.5 | $\mathbf{62.15 \pm 0.39}$ | $\mathbf{78.14 \pm 0.46}$ | $\mathbf{74.56 \pm 0.61}$ |
| **2** | $\mathbf{62.28 \pm 0.52}$ | $\mathbf{78.29 \pm 0.31}$ | $\mathbf{74.69 \pm 0.44}$ |
| 3 | $62.11 \pm 0.35$ | $78.08 \pm 0.55$ | $74.51 \pm 0.40$ |
| 5 | $61.94 \pm 0.48$ | $77.90 \pm 0.37$ | $74.34 \pm 0.59$ |
| 10 | $61.78 \pm 0.61$ | $77.71 \pm 0.49$ | $74.17 \pm 0.33$ |

Table 7: Sensitivity to the Gaussian bandwidth $\sigma_v$ on YouTubeFace at missing rate $\rho = 0.3$.

| $c$ | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| 0.25 | $61.75 \pm 0.56$ | $77.70 \pm 0.38$ | $74.18 \pm 0.47$ |
| 0.50 | $61.93 \pm 0.41$ | $77.92 \pm 0.63$ | $74.35 \pm 0.35$ |
| 0.75 | $62.12 \pm 0.59$ | $78.15 \pm 0.44$ | $74.54 \pm 0.52$ |
| **1.00** | $\mathbf{62.28 \pm 0.36}$ | $\mathbf{78.29 \pm 0.51}$ | $\mathbf{74.69 \pm 0.43}$ |
| 1.50 | $\mathbf{62.18 \pm 0.47}$ | $\mathbf{78.18 \pm 0.32}$ | $\mathbf{74.58 \pm 0.60}$ |
| 2.00 | $61.96 \pm 0.33$ | $77.95 \pm 0.57$ | $74.38 \pm 0.39$ |
| 4.00 | $61.72 \pm 0.62$ | $77.67 \pm 0.40$ | $74.14 \pm 0.54$ |

**View-Weight Exponent:** We vary $r$ in $\{1.1, 1.2, 1.5, 2, 3, 5, 10\}$ for Eqs. (6) and (13). As shown in Table 6, IGWA maintains strong performance throughout this range and achieves its best results at $r = 2$. Compared with the lowest scores obtained at $r = 10$, ACC, NMI, and PUR improve from 61.78, 77.71, and 74.17 to 62.28, 78.29, and 74.69, respectively. These correspond to absolute maximum–minimum gaps of only 0.50, 0.58, and 0.52 percentage points, or relative gains of 0.81%, 0.75%, and 0.70%. Such limited variation demonstrates that IGWA is highly insensitive to the choice of $r$, while an intermediate value effectively balances reliable-view emphasis and complementary information.

**Gaussian Kernel Bandwidth:** We set $\sigma_v = c\sigma_{v,\mathrm{ref}}$ in Eq. (3) and vary $c$ in $\{0.25, 0.5, 0.75, 1, 1.5, 2, 4\}$. Table 7 shows that IGWA performs best at $c = 1$, while the interval $[0.75, 1.5]$ forms an extremely stable high-performance region. This indicates that normalized Gaussian relations preserve reliable neighborhood geometry under moderate bandwidth changes, substantially reducing dependence on precise kernel-scale selection.

## 5 Conclusion and Future Work

We presented IGWA, an imputation-free framework for incomplete multi-view clustering. By aligning normalized intra-view relations through partial Gromov–Wasserstein plans, IGWA avoids forced raw-space correspondence and mitigates missing-structure contamination. A jointly learned consensus with adaptive view weights integrates complementary geometry, while theoretical analysis establishes relabeling invariance and stable consensus recovery. Experiments on six benchmarks show consistent gains across missing rates. ***Future work*** will explore scalable low-rank and multi-scale solvers, learn relational kernels, and automatically estimate transported mass under structured missingness.

## REFERENCES


Hugues Van Assel, Cédric Vincent-Cuaz, Nicolas Courty, Rémi Flamary, Pascal Frossard, and Titouan Vayer. Distributional reduction: Unifying dimensionality reduction and clustering with gromov–wasserstein. *Transactions on Machine Learning Research*, 2025. 4, 19, 20

Yikun Bai, Abihith Kothapalli, Hengrong Du, Rocio Diaz Martin, and Soheil Kolouri. Linear partial gromov–wasserstein embedding. In *International Conference on Learning Representations*, 2025a. 4, 28

Yikun Bai, Rocio Diaz Martin, Abihith Kothapalli, Hengrong Du, Xinran Liu, and Soheil Kolouri. Partial gromov–wasserstein metric. In *International Conference on Learning Representations*, 2025b. 4, 16, 20, 28

Florian Beier, Robert Beinert, and Gabriele Steidl. Multi-marginal gromov–wasserstein transport and barycentres. *Information and Inference: A Journal of the IMA*, 12(4):2753–2781, 2023. 4

Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015. 3

Luis A. Caffarelli and Robert J. McCann. Free boundaries in optimal transport and Monge–Ampère obstacle problems. *Annals of Mathematics*, 171(2):673–730, 2010. 4

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11221–11229, 2024. 6

Laetitia Chapel, Mokhtar Z. Alaya, and Gilles Gasso. Partial optimal transport with applications on positive–unlabeled learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2903–2913, 2020. 4, 16, 20, 28

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. 3

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. 3

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013. 3

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *Advances in Neural Information Processing Systems*, 38:52175–52195, 2026. 1

Zhibin Dong, Dayu Hu, Jiaqi Jin, Siwei Wang, Xinwang Liu, and En Zhu. Selective cross-view topology for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:4792–4805, 2025. 1

Alessio Figalli. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2):533–560, 2010. 4

Fengjiao Gong, Yuzhou Nie, and Hongteng Xu. Gromov–wasserstein multi-modal alignment and clustering. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pp. 603–613, 2022. 3, 15, 18, 20, 28

Zhibin Gu, Zhendong Li, and Songhe Feng. Edison: enhanced dictionary-induced tensorized incomplete multi-view clustering with gaussian error rank minimization. In *Forty-first international conference on machine learning*, 2024. 1

Zhibin Gu, Zhenhao Zhong, Xi Zhang, and Bing Li. Constant degree matrix-driven incomplete multi-view clustering via connectivity-structure and embedding tensor learning. In *International Conference on Learning Representations*, 2026a. 6

Zhibin Gu, Zhenhao Zhong, Xi Zhang, and Bing Li. Constant degree matrix-driven incomplete multi-view clustering via connectivity-structure and embedding tensor learning. In *International Conference on Learning Representations*, volume 2026, pp. 84649–84671, 2026b. 1

Xiaorui Jiang, Buyun He, Peng Yuan Zhou, Xinyue Chen, Jingcai Guo, Jie Xu, and Yong Liao. A unified framework to bridge complete and incomplete deep multi-view clustering under non-iid missing patterns. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 594–603. IEEE, 2025. 1

Dominik Klein, Théo Uscidda, Fabian Theis, and Marco Cuturi. Genot: Entropic (gromov) wasserstein flow matching with applications to single-cell genomics. *Advances in Neural Information Processing Systems*, 37:103897–103944, 2024. 2

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014. 1

Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu. Consensus graph learning for multi-view clustering. *IEEE Transactions on Multimedia*, 24:2461–2472, 2021. 1

Chengliang Liu, Zhihao Wu, Jie Wen, Yong Xu, and Chao Huang. Localized sparse incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 25:5539–5551, 2022. 1

Jiyuan Liu, Xinwang Liu, Xinhang Wan, Ke Liang, Weixuan Liang, Huijun Wu, Kehua Guo, et al. Incomplete multi-view deep clustering with data imputation and alignment. *Advances in Neural Information Processing Systems*, 38:89119–89142, 2026. 1

Dixin Luo, Hongteng Xu, and Lawrence Carin. Differentiable hierarchical optimal transport for robust multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (6):7293–7307, 2022. 2

Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011. 3

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov–wasserstein averaging of kernel and distance matrices. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2664–2672, 2016. 3

Yalan Qin, Guorui Feng, and Xinpeng Zhang. Fast incomplete multi-view clustering by flexible anchor learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 50235–50245. PMLR, 2025. 6

Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov–wasserstein distance: Conic formulation and relaxation. In *Advances in Neural Information Processing Systems*, volume 34, 2021. 4, 28

Justin Solomon, Gabriel Peyré, Vladimir G. Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics*, 35(4):72:1–72:13, 2016. 3

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International conference on machine learning*, pp. 21090–21110. PMLR, 2022. 6

Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Optimal transport for structured data with application on graphs. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6275–6284, 2019. 3

Cédric Vincent-Cuaz, Rémi Flamary, Marco Corneli, Titouan Vayer, and Nicolas Courty. Semi-relaxed gromov–wasserstein divergence with applications on graphs. In *International Conference on Learning Representations*, 2022. 4

Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral perturbation meets incomplete multi-view data. *arXiv preprint arXiv:1906.00098*, 2019. 1

Qianqian Wang, Haiming Xu, Wei Feng, and Quanxue Gao. Adversarial fair incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 10011–10019, 2026. 1

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, Fu Xiao, and Yao Zhao. Learning from disjoint views: A contrastive prototype matching network for fully incomplete multi-view clustering. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 6

Jie Wen, Zheng Zhang, Zhao Zhang, Lunke Fei, and Meng Wang. Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE transactions on cybernetics*, 51(1):101–114, 2020. 1, 28

Jie Wen, Yadong Liu, Zhanyan Tang, Yuting He, Yulong Chen, Mu Li, and Chengliang Liu. Learning compact semantic information for incomplete multi-view missing multi-label classification. In *Forty-second International Conference on Machine Learning*, 2025. 1

Hongteng Xu, Dixin Luo, and Lawrence Carin. Scalable gromov-wasserstein learning for graph partitioning and matching. *Advances in neural information processing systems*, 32, 2019a. 2, 3, 28

Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin. Gromov–wasserstein learning for graph matching and node embedding. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6932–6941, 2019b. 3, 17, 20

Hongteng Xu, Dixin Luo, Lawrence Carin, and Hongyuan Zha. Learning graphons via structured gromov-wasserstein barycenters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 10505–10513, 2021. 2

Jie Xu, Wenyuan Yang, Yazhou Ren, Lifang He, Philip S Yu, and Xiaofeng Zhu. Missing pattern tree based decision grouping and ensemble for enhancing pair utilization in deep incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. 1

Nan Xu, Yanqing Guo, Xin Zheng, Qianyu Wang, and Xiangyang Luo. Partial multi-view subspace clustering. In *Proceedings of the 26th ACM International conference on multimedia*, pp. 1794–1801, 2018. 1

Shengju Yu, Zhibin Dong, Siwei Wang, Pei Zhang, Yi Zhang, Xinwang Liu, Thomas Guan, Tiejun Li, and Yiu-Ming Cheung. Simple yet effective incomplete multi-view clustering: Similarity-level imputation and intra-view hybrid-group prototype construction. In *International Conference on Learning Representations*, 2025. 6

Honglin Yuan, Shiyun Lai, Xingfeng Li, Jian Dai, Yuan Sun, and Zhenwen Ren. Robust prototype completion for incomplete multi-view clustering. In *Proceedings of the 32nd ACM international conference on multimedia*, pp. 10402–10411, 2024. 6

Handong Zhao, Hongfu Liu, Yun Fu, et al. Incomplete multi-modal visual data grouping. In *Ijcai*, volume 16, pp. 2392–2398, 2016. 1

Zhenhao Zhong, Zhibin Gu, Pengpeng Yang, Yaqian Zhou, and Ruiqiang Guo. Gaussian regression-driven tensorized incomplete multi-view clustering with dual manifold regularization. In *Advances in Neural Information Processing Systems*, volume 38, 2025. 6

Zhenhao Zhong, Zhibin Gu, Pengpeng Yang, Ruiqiang Guo, et al. Gaussian regression-driven tensorized incomplete multi-view clustering with dual manifold regularization. *Advances in Neural Information Processing Systems*, 38:72102–72131, 2026. 1

# APPENDIX

## A LLM USAGE

We employ LLMs solely to refine the language and wording for improved fluency and precision,with no further application beyond this purpose.

## B LIST OF SYMBOLS

Table 8 concisely lists the symbols, excluding unnecessary subscripts for clarity.

| Notation | Description | Index |
|---|---|---|
| $V$ | Number of views | §1; Eq. (1) |
| $N$ | Number of total unique instances | §1; Eq. (1) |
| $N_v$ | Number of observed instances in view $v$ | §1; Eq. (1) |
| $d_v$ | Feature dimensions of the $v$-th view | §1; Eq. (1) |
| $\mathbf{Y}^{(v)}$ | Observed data in the $v$-th view | §1; Eq. (1) |
| $\mathbf{C}^{(v)}$ | intra-view relational matrix for view $v$ | §3.1; Eq. (3) |
| $\mathbf{C}^*$ | Learnable global consensus structure | §3.1; Eq. (4)&(6) |
| $\mathbf{T}^{(v)}$ | Partial Gromov-Wasserstein transport plan | §3.1; Eq. (4)&(5) |
| $\mathbf{p}^{(v)}, \mathbf{q}$ | Prior marginal distributions | §3.1; Eq. (5) |
| $s^{(v)}$ | Transported matched mass | §3.1; Eq. (5) |
| $\mathcal{L}_{\text{IGW}}$ | Incomplete Gromov-Wasserstein discrepancy | §3.1; Eq. (4) |
| $\Pi_{\leq}$ | Relaxed feasible set of the transport plan | §3.1; Eq. (5) |
| $\eta$ | Entropy regularization parameter | §3.1; Eq. (4) |
| $\boldsymbol{\alpha}$ | Adaptive view weights vector | §3.2; Eq. (6) |
| $\alpha_v$ | Adaptive weight for the $v$-th view | §3.2; Eq. (13) |
| $\gamma$ | Global consensus complexity regularizer | §3.2; Eq. (6) |
| $\mathcal{J}$ | Unified variational objective function | §3.2; Eq. (6) |
| $\mathbf{q}^{(v)}$ | Induced consensus mass from transport plan | §3.2; Eq. (8) |
| $\mathbf{M}^{(v)}$ | Local cost matrix at iteration $t$ | §3.3; Eq. (11) |
| $E_v$ | Converged structural alignment loss for view $v$ | §3.3; Eq. (13) |

Table 8: Notation and Description of Key Symbols.

## C OPTIMIZATION ALGORITHM

Algo 1 summarizes the complete optimization procedure of IGWA. Starting from the normalized intra-view relations, IGWA alternately updates the partial Gromov–Wasserstein transport plans, the global consensus relation, and the adaptive view weights within a block coordinate optimization framework. For each view, the transport plan is refined through generalized Sinkhorn iterations using the local cost matrix in Eq. (11). The resulting transported masses are then used to update the consensus relation by Eq. (12), followed by adaptive view reweighting according to Eq. (13). These steps are repeated until convergence, after which the learned consensus relation $\mathbf{C}^*$ is used to obtain the final clustering assignments.

## D PROOF OF RELABELING INVARIANCE AND BALANCED RECOVERY

### D.1 PROOF OF THEOREM 1

This section provides a detailed proof of Theorem 1. We prove, in order, the preservation of the relaxed transport constraints, invariance of the structural objective, recovery of balanced GW as a limiting case, and the difference between relational alignment and direct Euclidean consensus matching.

**Algorithm 1: Optimization Procedure of IGWA**

**Input:** $\{\mathbf{Y}^{(v)}\}_{v=1}^{V}$, number of clusters $K$, $\eta, \gamma, r$, transported masses $\{s^{(v)}\}_{v=1}^{V}$, backtracking factor $\tau \in (0,1)$.
**Output:** Clustering assignments.

```
// Initialization
for v = 1 to V do
    Construct C^(v) by Eq. (3);
    Initialize feasible T^(v) ∈ Π_≤(p^(v), q);
Initialize C* and α = [α_1, ..., α_V]^⊤;
// Block Coordinate Optimization
while not converged do
    // Safeguarded Partial GW Alignment
    for v = 1 to V do
        while T^(v) not converged do
            Set T_0 ← T^(v);
            Compute M^(v) by Eq. (11);
            Obtain candidate T̃^(v) by generalized Sinkhorn scaling for Eq. (10);
            Set β ← 1 and T_β ← (1 − β)T_0 + βT̃^(v);
            while L_IGW(C^(v), C*, T_β) > L_IGW(C^(v), C*, T_0) do
                β ← τβ;
                T_β ← (1 − β)T_0 + βT̃^(v);
            T^(v) ← T_β;
    // Global Topology Consensus
    for v = 1 to V do
        q^(v) ← T^(v)⊤ 1_{N_v};
    Update C* by Eq. (12);
    // Adaptive View Reweighting
    for v = 1 to V do
        E_v ← L_IGW(C^(v), C*, T^(v));
    Update α by Eq. (13);
// Final Clustering
Apply spectral clustering with K clusters to (C* + C*^⊤)/2;
return Clustering assignments;
```

*Proof.* For the $v$-th view, let $\mathbf{P}_v \in \mathbb{R}^{N_v \times N_v}$ and $\mathbf{P} \in \mathbb{R}^{N \times N}$ be arbitrary permutation matrices. Following Theorem 1, define

$$\widetilde{\mathbf{C}}^{(v)} = \mathbf{P}_v \mathbf{C}^{(v)} \mathbf{P}_v^\top, \qquad \widetilde{\mathbf{C}}^{*} = \mathbf{P} \mathbf{C}^{*} \mathbf{P}^\top, \qquad \widetilde{\mathbf{T}}^{(v)} = \mathbf{P}_v \mathbf{T}^{(v)} \mathbf{P}^\top. \tag{14}$$

Because a permutation matrix is orthogonal and preserves the all-one vector, we have

$$\mathbf{P}_v^\top \mathbf{P}_v = \mathbf{I}_{N_v}, \qquad \mathbf{P}^\top \mathbf{P} = \mathbf{I}_N, \qquad \mathbf{P}_v^\top \mathbf{1}_{N_v} = \mathbf{1}_{N_v}, \qquad \mathbf{P}^\top \mathbf{1}_N = \mathbf{1}_N. \tag{15}$$

Consider first the row marginal of the transformed transport plan:

$$\widetilde{\mathbf{T}}^{(v)} \mathbf{1}_N = \mathbf{P}_v \mathbf{T}^{(v)} \mathbf{P}^\top \mathbf{1}_N = \mathbf{P}_v \mathbf{T}^{(v)} \mathbf{1}_N. \tag{16}$$

Using the original partial marginal constraint $\mathbf{T}^{(v)} \mathbf{1}_N \le \mathbf{p}^{(v)}$, the previous identity gives the following three-step implication:

$$\mathbf{T}^{(v)} \mathbf{1}_N \le \mathbf{p}^{(v)} \Longleftrightarrow \mathbf{P}_v \mathbf{T}^{(v)} \mathbf{1}_N \le \mathbf{P}_v \mathbf{p}^{(v)} \Longleftrightarrow \widetilde{\mathbf{T}}^{(v)} \mathbf{1}_N \le \widetilde{\mathbf{p}}^{(v)}, \qquad \widetilde{\mathbf{p}}^{(v)} = \mathbf{P}_v \mathbf{p}^{(v)}. \tag{17}$$

Similarly, the column marginal satisfies $\widetilde{\mathbf{T}}^{(v)\top} \mathbf{1}_{N_v} = \mathbf{P} \mathbf{T}^{(v)\top} \mathbf{P}_v^\top \mathbf{1}_{N_v} = \mathbf{P} \mathbf{T}^{(v)\top} \mathbf{1}_{N_v}$. Hence,

$$\mathbf{T}^{(v)\top} \mathbf{1}_{N_v} \le \mathbf{q} \Longleftrightarrow \mathbf{P} \mathbf{T}^{(v)\top} \mathbf{1}_{N_v} \le \mathbf{P} \mathbf{q} \Longleftrightarrow \widetilde{\mathbf{T}}^{(v)\top} \mathbf{1}_{N_v} \le \widetilde{\mathbf{q}}, \qquad \widetilde{\mathbf{q}} = \mathbf{P} \mathbf{q}. \tag{18}$$

The transported mass is also invariant. Indeed,

$$\mathbf{1}_{N_v}^\top \widetilde{\mathbf{T}}^{(v)} \mathbf{1}_N = \mathbf{1}_{N_v}^\top \mathbf{P}_v \mathbf{T}^{(v)} \mathbf{P}^\top \mathbf{1}_N = \mathbf{1}_{N_v}^\top \mathbf{T}^{(v)} \mathbf{1}_N = s^{(v)}. \tag{19}$$

Consequently, feasibility before and after relabeling is equivalent: $\mathbf{T}^{(v)} \in \Pi_{\leq}\big(\mathbf{p}^{(v)}, \mathbf{q}\big)$. We next consider the quadratic structural component of Eq. (4). For compactness of the derivation, write it explicitly as $\mathcal{S}_v = \sum_{i,j=1}^{N_v} \sum_{k,l=1}^{N} \left(C_{ij}^{(v)} - C_{kl}^{*}\right)^2 T_{ik}^{(v)} T_{jl}^{(v)}$. After applying the permutations in Eq. (14), the same term becomes $\widetilde{\mathcal{S}}_v = \sum_{i,j=1}^{N_v} \sum_{k,l=1}^{N} \left(\widetilde{C}_{ij}^{(v)} - \widetilde{C}_{kl}^{*}\right)^2 \widetilde{T}_{ik}^{(v)} \widetilde{T}_{jl}^{(v)}$. Since $\mathbf{P}_v$ and $\mathbf{P}$ only bijectively reorder row and column indices, a change of summation indices yields

$$\widetilde{\mathcal{S}}_v = \sum_{i,j,k,l} \left(C_{\pi_v(i)\pi_v(j)}^{(v)} - C_{\pi(k)\pi(l)}^{*}\right)^2 T_{\pi_v(i)\pi(k)}^{(v)} T_{\pi_v(j)\pi(l)}^{(v)} = \mathcal{S}_v, \tag{20}$$

where $\pi_v$ and $\pi$ denote the index permutations represented by $\mathbf{P}_v$ and $\mathbf{P}$, respectively.

The entropy term is invariant for the same reason:

$$\mathcal{H}\big(\widetilde{\mathbf{T}}^{(v)}\big) = \sum_{i,k} \widetilde{T}_{ik}^{(v)} \left(\log \widetilde{T}_{ik}^{(v)} - 1\right) \Longleftrightarrow \sum_{i,k} T_{ik}^{(v)} \left(\log T_{ik}^{(v)} - 1\right) = \mathcal{H}\big(\mathbf{T}^{(v)}\big). \tag{21}$$

Combining Eqs. (20) and (21) therefore gives

$$\mathcal{L}_{\mathrm{IGW}}\left(\widetilde{\mathbf{C}}^{(v)}, \widetilde{\mathbf{C}}^{*}, \widetilde{\mathbf{T}}^{(v)}\right) = \mathcal{L}_{\mathrm{IGW}}\left(\mathbf{C}^{(v)}, \mathbf{C}^{*}, \mathbf{T}^{(v)}\right). \tag{22}$$

The consensus regularizer is also unchanged:

$$\begin{aligned} \left\|\mathbf{P}\mathbf{C}^{*}\mathbf{P}^\top\right\|_F^2 &= \mathrm{Tr}\left(\mathbf{P}\mathbf{C}^{*}\mathbf{P}^\top\mathbf{P}\mathbf{C}^{*\top}\mathbf{P}^\top\right) \\ &= \mathrm{Tr}\left(\mathbf{P}\mathbf{C}^{*}\mathbf{C}^{*\top}\mathbf{P}^\top\right) = \mathrm{Tr}\left(\mathbf{C}^{*}\mathbf{C}^{*\top}\right) = \left\|\mathbf{C}^{*}\right\|_F^2. \end{aligned} \tag{23}$$

Therefore, with the same view weights, the complete IGWA objective in Eq. (6) satisfies

$$\mathcal{J}\left(\{\widetilde{\mathbf{T}}^{(v)}\}_{v=1}^{V}, \widetilde{\mathbf{C}}^{*}, \boldsymbol{\alpha}\right) = \mathcal{J}\left(\{\mathbf{T}^{(v)}\}_{v=1}^{V}, \mathbf{C}^{*}, \boldsymbol{\alpha}\right). \tag{24}$$

**Remark 3.** *The invariance established above concerns the complete IGWA construction, including partial feasibility, the $N \times N$ consensus relation, and its regularization. GWMAC Gong et al. (2022) also exploits the permutation invariance of GW barycenters, but its barycenter is built using standard balanced couplings of the form $\mathbf{T}_m \in \Pi(\bar{\mathbf{p}}, \widehat{\mathbf{p}}_m)$ and is coupled to a clustering objective. IGWA instead permits $\mathbf{T}^{(v)}\mathbf{1}_N \leq \mathbf{p}^{(v)}$ and $\mathbf{T}^{(v)\top}\mathbf{1}_{N_v} \leq \mathbf{q}$ with an explicitly controlled transported mass $s^{(v)}$. Thus, the property proved here is not a new permutation-invariance property of GW itself; it verifies that the additional partial-mass and consensus components introduced for incomplete multi-view clustering preserve that desirable structural invariance.*

We next prove balanced recovery. Define the induced row and column masses

$$\mathbf{r}^{(v)} = \mathbf{T}^{(v)}\mathbf{1}_N, \qquad \mathbf{c}^{(v)} = \mathbf{T}^{(v)\top}\mathbf{1}_{N_v}. \tag{25}$$

By Eq. (5),

$$\mathbf{r}^{(v)} \leq \mathbf{p}^{(v)}, \qquad \mathbf{c}^{(v)} \leq \mathbf{q}, \qquad \mathbf{1}_{N_v}^\top \mathbf{r}^{(v)} = \mathbf{1}_N^\top \mathbf{c}^{(v)} = s^{(v)}. \tag{26}$$

Suppose now that $s^{(v)} = \|\mathbf{p}^{(v)}\|_1 = \|\mathbf{q}\|_1$. For the row marginal,

$$\mathbf{p}^{(v)} - \mathbf{r}^{(v)} \geq \mathbf{0}, \quad \mathbf{1}_{N_v}^\top \left(\mathbf{p}^{(v)} - \mathbf{r}^{(v)}\right) = \|\mathbf{p}^{(v)}\|_1 - s^{(v)} = 0. \tag{27}$$

A nonnegative vector whose entries sum to zero must be identically zero; hence

$$\mathbf{T}^{(v)}\mathbf{1}_N = \mathbf{r}^{(v)} = \mathbf{p}^{(v)}. \tag{28}$$

Exactly the same argument applies to the column marginal:

$$\mathbf{q} - \mathbf{c}^{(v)} \geq \mathbf{0}, \quad \mathbf{1}_N^\top \left(\mathbf{q} - \mathbf{c}^{(v)}\right) = \|\mathbf{q}\|_1 - s^{(v)} = 0, \tag{29}$$

which implies: $\mathbf{T}^{(v)\top}\mathbf{1}_{N_v} = \mathbf{c}^{(v)} = \mathbf{q}$. Consequently, at full transported mass the relaxed feasible set collapses exactly to the balanced transport polytope:

$$\Pi_{\leq}\big(\mathbf{p}^{(v)}, \mathbf{q}\big)\Big|_{s^{(v)}=\|\mathbf{p}^{(v)}\|_1=\|\mathbf{q}\|_1} = \Pi\big(\mathbf{p}^{(v)}, \mathbf{q}\big). \tag{30}$$

Substituting Eq. (30) into Eq. (4) yields the following three-stage reduction of its structural component:

$$\begin{aligned}\min_{\mathbf{T}^{(v)}\in\Pi_{\leq}} \sum_{i,j,k,l} \left(C_{ij}^{(v)} - C_{kl}^{*}\right)^2 T_{ik}^{(v)} T_{jl}^{(v)} &\iff \min_{\mathbf{T}^{(v)}\in\Pi(\mathbf{p}^{(v)},\mathbf{q})} \sum_{i,j,k,l} \left(C_{ij}^{(v)} - C_{kl}^{*}\right)^2 T_{ik}^{(v)} T_{jl}^{(v)} \\ &\iff \mathrm{GW}_2^2\left(\mathbf{C}^{(v)}, \mathbf{C}^{*}; \mathbf{p}^{(v)}, \mathbf{q}\right),\end{aligned} \tag{31}$$

where the last line denotes the standard squared-loss balanced GW structural objective.

To make the entropy convention explicit, let the Shannon entropy be the negative of the function used in Eq. (4):

$$\mathcal{H}_{\mathrm{Sh}}\big(\mathbf{T}^{(v)}\big) = -\mathcal{H}\big(\mathbf{T}^{(v)}\big) \quad \iff \quad \eta\mathcal{H}\big(\mathbf{T}^{(v)}\big) = -\eta\mathcal{H}_{\mathrm{Sh}}\big(\mathbf{T}^{(v)}\big). \tag{32}$$

Thus, after identifying $\epsilon = \eta$ and using the same entropy sign convention, Eq. (4) at full mass becomes

$$\mathcal{L}_{\mathrm{IGW}}\left(\mathbf{C}^{(v)}, \mathbf{C}^{*}, \mathbf{T}^{(v)}\right)\Big|_{\text{full mass}} = \mathcal{GW}_{\epsilon}\left(\mathbf{C}^{(v)}, \mathbf{C}^{*}, \mathbf{p}^{(v)}, \mathbf{q}\right). \tag{33}$$

**Remark 4.** *This limiting relation clarifies the precise connection to the partial-GW literature. Chapel et al. Chapel et al. (2020) develop exact partial Wasserstein and partial GW formulations for pairwise comparison and demonstrate them in positive–unlabeled learning. Bai et al. Bai et al. (2025b) subsequently establish a particular PGW formulation as a mathematically valid metric between metric-measure spaces and study corresponding Frank–Wolfe solvers and barycenters. IGWA does not claim that the partial-GW discrepancy itself is newly introduced, nor do we claim that the entropy-regularized $\mathcal{L}_{\mathrm{IGW}}$ in Eq. (4) is a new metric. Our task-specific distinction is that $V$ view-dependent partial couplings, possibly with different $s^{(v)}$, are jointly linked to a single learnable $\mathbf{C}^{*}$ and adaptive weights $\boldsymbol{\alpha}$ through Eq. (6). Equation (33) is therefore a consistency property of IGWA rather than a new metric theorem.*

Finally, we contrast relational invariance with a direct Euclidean consensus term. Assume, as in Theorem 1, that for one view

$$\mathbf{Z}^{(v)}\mathbf{M}^{(v)} = \mathbf{H} \neq \mathbf{0}, \qquad \left\|\mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{H}\right\|_F^2 = 0. \tag{34}$$

After applying the isometric sign transformation $\mathbf{Z}^{(v)} \mapsto -\mathbf{Z}^{(v)}$ while keeping the common Euclidean reference $\mathbf{H}$ fixed, the same term satisfies

$$\left\|-\mathbf{Z}^{(v)}\mathbf{M}^{(v)} - \mathbf{H}\right\|_F^2 = \|-\mathbf{H} - \mathbf{H}\|_F^2 = \|-2\mathbf{H}\|_F^2 = 4\,\|\mathbf{H}\|_F^2 > 0. \tag{35}$$

In contrast, an intra-view pairwise relation based on Euclidean distances is unchanged by the same sign transformation:

$$\left\|(-\mathbf{y}_i^{(v)}) - (-\mathbf{y}_j^{(v)})\right\|_2^2 = \left\|-(\mathbf{y}_i^{(v)} - \mathbf{y}_j^{(v)})\right\|_2^2 = (\mathbf{y}_i^{(v)} - \mathbf{y}_j^{(v)})^\top(\mathbf{y}_i^{(v)} - \mathbf{y}_j^{(v)}) = \left\|\mathbf{y}_i^{(v)} - \mathbf{y}_j^{(v)}\right\|_2^2. \tag{36}$$

Therefore, the normalized Gaussian relation in Eq. (3) satisfies

$$\mathbf{C}^{(v)}\left(-\mathbf{Y}^{(v)}\right) = \mathbf{C}^{(v)}\left(\mathbf{Y}^{(v)}\right), \tag{37}$$

whereas the direct pointwise Euclidean consensus penalty need not remain unchanged.

**Remark 5.** *The distinction from Gromov–Wasserstein learning for graph matching and node embedding Xu et al. (2019b) is also important. That work uses GW to learn a pairwise graph correspondence jointly with node embeddings and optimizes a balanced transport plan through a proximal-point procedure. IGWA does not introduce node embeddings to establish correspondence. Instead, each observed view is lifted to $\mathbf{C}^{(v)}$, partially coupled to a shared $\mathbf{C}^*$, and allowed to leave unreliable mass unmatched. Hence, Eq. (37) supports the use of intrinsic relations, while Eqs. (27)–(33) characterize the additional partial-mass mechanism required by the incomplete-view setting.*

Equations (16)–(24) establish relabeling invariance, Eqs. (25)–(33) establish balanced recovery, and Eqs. (34)–(37) establish the claimed distinction from direct Euclidean consensus matching. This completes the proof of Theorem 1. □

### D.2 Proof of Unique, Bounded, and Stable Consensus

We next prove Theorem 2. Throughout this section, the feasible transport plans $\{\mathbf{T}^{(v)}\}_{v=1}^{V}$ and view weights $\boldsymbol{\alpha}$ are fixed. Therefore, only $\mathbf{C}^*$ is optimized.

*Proof.* Starting from Eq. (6) and omitting the entropy terms, which are constant with respect to $\mathbf{C}^*$, the relevant subproblem is

$$\min_{\mathbf{C}^*} \sum_{v=1}^{V} \alpha_v^r \sum_{i,j=1}^{N_v} \sum_{k,l=1}^{N} \left(C_{ij}^{(v)} - C_{kl}^*\right)^2 T_{ik}^{(v)} T_{jl}^{(v)} + \gamma \|\mathbf{C}^*\|_F^2. \tag{38}$$

Expanding the squared discrepancy gives

$$\left(C_{ij}^{(v)} - C_{kl}^*\right)^2 = \left(C_{ij}^{(v)}\right)^2 + (C_{kl}^*)^2 - 2C_{ij}^{(v)} C_{kl}^*. \tag{39}$$

The first component of Eq. (39) does not depend on $\mathbf{C}^*$:

$$\sum_{v,i,j,k,l} \alpha_v^r \left(C_{ij}^{(v)}\right)^2 T_{ik}^{(v)} T_{jl}^{(v)} = \text{constant} \qquad \text{w.r.t. } \mathbf{C}^*. \tag{40}$$

Recall from the main text that

$$q_k^{(v)} = \sum_{i=1}^{N_v} T_{ik}^{(v)} = \left(\mathbf{T}^{(v)\top} \mathbf{1}_{N_v}\right)_k. \tag{41}$$

The coefficient of $(C_{kl}^*)^2$ can consequently be simplified by a three-step factorization:

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_v} T_{ik}^{(v)} T_{jl}^{(v)} = \left(\sum_{i=1}^{N_v} T_{ik}^{(v)}\right) \left(\sum_{j=1}^{N_v} T_{jl}^{(v)}\right) = q_k^{(v)} q_l^{(v)} = \left[\mathbf{q}^{(v)} \mathbf{q}^{(v)\top}\right]_{kl}. \tag{42}$$

Likewise, the mixed term satisfies

$$\sum_{i=1}^{N_v} \sum_{j=1}^{N_v} C_{ij}^{(v)} T_{ik}^{(v)} T_{jl}^{(v)} = \left(\mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)}\right)_{kl}. \tag{43}$$

Substituting Eqs. (40)–(43) into Eq. (38) yields an element-wise separable objective:

$$\mathcal{J}(\mathbf{C}^*) \Longleftrightarrow \sum_{k,l=1}^{N} \left[ \left(\gamma + \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)}\right) (C_{kl}^*)^2 - 2C_{kl}^* \sum_{v=1}^{V} \alpha_v^r \left(\mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)}\right)_{kl} \right]. \tag{44}$$

To match the compact notation appearing in Theorem 2, define

$$a_{kl} = \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)}, \qquad b_{kl} = \sum_{v=1}^{V} \alpha_v^r \left( \mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)} \right)_{kl}. \tag{45}$$

Equation (44) then becomes

$$\mathcal{J}(\mathbf{C}^*) \Longleftrightarrow \sum_{k,l=1}^{N} \left[ (\gamma + a_{kl})(C_{kl}^*)^2 - 2 b_{kl} C_{kl}^* \right]. \tag{46}$$

For an arbitrary entry $(k, l)$, differentiation gives

$$\frac{\partial \mathcal{J}}{\partial C_{kl}^*} = 2(\gamma + a_{kl}) C_{kl}^* - 2 b_{kl}. \tag{47}$$

The corresponding second derivative is

$$\frac{\partial^2 \mathcal{J}}{\partial (C_{kl}^*)^2} = 2(\gamma + a_{kl}). \tag{48}$$

Since $\alpha_v^r \geq 0$, $q_k^{(v)} q_l^{(v)} \geq 0$, and $\gamma > 0$,

$$a_{kl} = \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)} \geq 0 \Longrightarrow \gamma + a_{kl} \geq \gamma \Longrightarrow 2(\gamma + a_{kl}) \geq 2\gamma > 0. \tag{49}$$

Therefore, the Hessian of the fixed-plan subproblem is diagonal under vectorization of $\mathbf{C}^*$ and is positive definite. The same conclusion follows directly by completing the square:

$$\begin{aligned} (\gamma + a_{kl})(C_{kl}^*)^2 - 2 b_{kl} C_{kl}^* &= (\gamma + a_{kl}) \left[ (C_{kl}^*)^2 - 2 \frac{b_{kl}}{\gamma + a_{kl}} C_{kl}^* \right] \\ &= (\gamma + a_{kl}) \left( C_{kl}^* - \frac{b_{kl}}{\gamma + a_{kl}} \right)^2 - \frac{b_{kl}^2}{\gamma + a_{kl}}. \end{aligned} \tag{50}$$

Thus, every element has the unique minimizer

$$C_{kl}^* = \frac{b_{kl}}{\gamma + a_{kl}} = \frac{\sum_{v=1}^{V} \alpha_v^r \left( \mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)} \right)_{kl}}{\gamma + \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)}}, \tag{51}$$

which is exactly Eq. (9) and Eq. (12).

**Remark 6.** *The role of $\gamma$ distinguishes the fixed-plan consensus analysis here from a conventional GW barycenter update. For example, GWMAC Gong et al. (2022) derives a weighted GW barycenter from aligned kernel matrices under balanced couplings. IGWA additionally places the explicit quadratic term $\gamma \|\mathbf{C}^*\|_F^2$ in the joint incomplete-view objective. We do not claim that a standard GW barycenter lacks a valid update or uniqueness under its own assumptions. Rather, Eq. (49) shows the specific property needed here: even when a consensus entry receives zero transported support, i.e., $a_{kl} = 0$, the IGWA fixed-plan subproblem retains curvature $2\gamma > 0$.*

We next prove boundedness. Since $T_{ik}^{(v)} \geq 0$ and $C_{ij}^{(v)} \geq 0$,

$$\left( \mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)} \right)_{kl} = \sum_{i,j} T_{ik}^{(v)} C_{ij}^{(v)} T_{jl}^{(v)} \geq 0. \tag{52}$$

Using the assumption $0 \leq C_{ij}^{(v)} \leq M$, we further obtain

$$\left( \mathbf{T}^{(v)\top} \mathbf{C}^{(v)} \mathbf{T}^{(v)} \right)_{kl} = \sum_{i,j} T_{ik}^{(v)} C_{ij}^{(v)} T_{jl}^{(v)} \leq M \sum_{i,j} T_{ik}^{(v)} T_{jl}^{(v)} = M q_k^{(v)} q_l^{(v)}. \tag{53}$$

Multiplication by the nonnegative $\alpha_v^r$ and summation over views therefore gives

$$0 \le b_{kl} \le M \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)} = M a_{kl}. \tag{54}$$

Substitution into Eq. (51) gives the two-step bound

$$0 \le C_{kl}^* = \frac{b_{kl}}{\gamma + a_{kl}} \le \frac{M a_{kl}}{\gamma + a_{kl}}. \tag{55}$$

Because $\gamma > 0$,

$$\frac{M a_{kl}}{\gamma + a_{kl}} = M \left(1 - \frac{\gamma}{\gamma + a_{kl}}\right) < M, \tag{56}$$

and hence

$$0 \le C_{kl}^* < M, \qquad \forall\, k, l. \tag{57}$$

We now establish stability. Consider perturbed view relations $\{\widehat{\mathbf{C}}^{(v)}\}_{v=1}^{V}$ while keeping the same transport plans and view weights, and define

$$\delta = \max_{1 \le v \le V} \left\| \mathbf{C}^{(v)} - \widehat{\mathbf{C}}^{(v)} \right\|_\infty. \tag{58}$$

By definition,

$$\left| C_{ij}^{(v)} - \widehat{C}_{ij}^{(v)} \right| \le \delta, \qquad \forall\, v, i, j. \tag{59}$$

For every view and consensus entry $(k, l)$,

$$\left| \left[ \mathbf{T}^{(v)\top} \left( \mathbf{C}^{(v)} - \widehat{\mathbf{C}}^{(v)} \right) \mathbf{T}^{(v)} \right]_{kl} \right| = \left| \sum_{i,j} T_{ik}^{(v)} \left( C_{ij}^{(v)} - \widehat{C}_{ij}^{(v)} \right) T_{jl}^{(v)} \right| \le \delta \sum_{i,j} T_{ik}^{(v)} T_{jl}^{(v)} = \delta q_k^{(v)} q_l^{(v)}. \tag{60}$$

Let $\widehat{b}_{kl}$ denote the numerator obtained from the perturbed relations. Then

$$|b_{kl} - \widehat{b}_{kl}| \le \sum_{v=1}^{V} \alpha_v^r \left| \left[ \mathbf{T}^{(v)\top} (\mathbf{C}^{(v)} - \widehat{\mathbf{C}}^{(v)}) \mathbf{T}^{(v)} \right]_{kl} \right| \le \delta \sum_{v=1}^{V} \alpha_v^r q_k^{(v)} q_l^{(v)} = \delta a_{kl}. \tag{61}$$

Because the plans and weights are fixed, the denominator is the same for the original and perturbed consensus solutions:

$$C_{kl}^* - \widehat{C}_{kl}^* = \frac{b_{kl} - \widehat{b}_{kl}}{\gamma + a_{kl}}. \tag{62}$$

Combining Eqs. (61) and (62) gives

$$\left| C_{kl}^* - \widehat{C}_{kl}^* \right| = \frac{|b_{kl} - \widehat{b}_{kl}|}{\gamma + a_{kl}} \le \frac{\delta a_{kl}}{\gamma + a_{kl}} \le \delta. \tag{63}$$

Taking the maximum over all $(k, l)$ yields

$$\left\| \mathbf{C}^* - \widehat{\mathbf{C}}^* \right\|_\infty \le \max_v \left\| \mathbf{C}^{(v)} - \widehat{\mathbf{C}}^{(v)} \right\|_\infty. \tag{64}$$

**Remark 7.** *Equation (64) should be interpreted as a fixed-plan stability property rather than a global convergence claim for the complete non-convex problem. This differs from Distributional Reduction Assel et al. (2025), which uses a GW-based block-coordinate procedure to jointly optimize a reduced set of prototypes and a transport plan in order to unify dimensionality reduction and clustering. IGWA instead fixes $\{\mathbf{T}^{(v)}\}$ in the present theorem and quantifies how perturbations of multiple observed intra-view relations propagate to their common consensus relation. Therefore,*

*the mathematical object controlled by Eq. (64) is an incomplete multi-view consensus matrix, not a learned low-dimensional prototype geometry.*

Finally, we prove equivariance of the closed-form consensus update. Under the relabeling used in Theorem 1,

$$\widetilde{\mathbf{q}}^{(v)} = \widetilde{\mathbf{T}}^{(v)\top}\mathbf{1}_{N_v} = \mathbf{P}\mathbf{T}^{(v)\top}\mathbf{P}_v^\top\mathbf{1}_{N_v} = \mathbf{P}\mathbf{q}^{(v)}. \tag{65}$$

The transported relation transforms according to

$$\widetilde{\mathbf{T}}^{(v)\top}\widetilde{\mathbf{C}}^{(v)}\widetilde{\mathbf{T}}^{(v)} = \mathbf{P}\mathbf{T}^{(v)\top}\mathbf{P}_v^\top\mathbf{P}_v\mathbf{C}^{(v)}\mathbf{P}_v^\top\mathbf{P}_v\mathbf{T}^{(v)}\mathbf{P}^\top = \mathbf{P}\mathbf{T}^{(v)\top}\mathbf{C}^{(v)}\mathbf{T}^{(v)}\mathbf{P}^\top = \mathbf{P}\left(\mathbf{T}^{(v)\top}\mathbf{C}^{(v)}\mathbf{T}^{(v)}\right)\mathbf{P}^\top. \tag{66}$$

Let $\pi$ denote the permutation associated with $\mathbf{P}$. Equation (65) implies

$$\widetilde{a}_{kl} = \sum_{v=1}^{V}\alpha_v^r\widetilde{q}_k^{(v)}\widetilde{q}_l^{(v)} = a_{\pi(k)\pi(l)}. \tag{67}$$

Likewise, Eq. (66) gives

$$\widetilde{b}_{kl} = \sum_{v=1}^{V}\alpha_v^r\left(\widetilde{\mathbf{T}}^{(v)\top}\widetilde{\mathbf{C}}^{(v)}\widetilde{\mathbf{T}}^{(v)}\right)_{kl} = b_{\pi(k)\pi(l)}. \tag{68}$$

Therefore, the transformed closed-form solution satisfies

$$\widetilde{C}^*_{kl} = \frac{\widetilde{b}_{kl}}{\gamma + \widetilde{a}_{kl}} = \frac{b_{\pi(k)\pi(l)}}{\gamma + a_{\pi(k)\pi(l)}} = C^*_{\pi(k)\pi(l)}, \tag{69}$$

or equivalently

$$\widetilde{\mathbf{C}}^* = \mathbf{P}\mathbf{C}^*\mathbf{P}^\top. \tag{70}$$

The regularizer also provides a transparent shrinkage mechanism. For fixed $a_{kl}$ and $b_{kl}$,

$$\frac{\partial C^*_{kl}}{\partial\gamma} = -\frac{b_{kl}}{(\gamma + a_{kl})^2} \le 0, \tag{71}$$

so increasing $\gamma$ monotonically reduces every nonnegative consensus entry whose numerator is nonzero.

Moreover, the ratio between the regularized and unregularized fixed-plan update, whenever $a_{kl} > 0$, is

$$\frac{b_{kl}/(\gamma + a_{kl})}{b_{kl}/a_{kl}} = \frac{a_{kl}}{\gamma + a_{kl}} \in [0, 1), \tag{72}$$

showing explicitly that entries with weak transported support $a_{kl}$ experience stronger relative shrinkage.

**Remark 8.** *Chapel et al. Chapel et al. (2020) and Bai et al. Bai et al. (2025b) primarily establish partial-GW formulations, algorithms, metric properties, and barycenter theory. Xu et al. Xu et al. (2019b) optimize pairwise graph correspondence and node embeddings. GWMAC Gong et al. (2022) learns a balanced GW kernel barycenter jointly with multimodal clustering, whereas Distributional Reduction Assel et al. (2025) learns a reduced prototype distribution for dimensionality reduction and clustering. In contrast, Theorem 2 concerns the specific regularized consensus block of IGWA: multiple partial plans contribute $\mathbf{T}^{(v)\top}\mathbf{C}^{(v)}\mathbf{T}^{(v)}$ to one $N \times N$ consensus, their contributions are modulated by $\alpha_v^r$, and $\gamma$ produces the denominator $\gamma + a_{kl}$. The theorem therefore establishes uniqueness, boundedness, and perturbation stability of this IGWA consensus update; it does not assert a new PGW metric or global optimality of the full non-convex objective.*

Equations (47)–(51) prove uniqueness, Eqs. (52)–(57) prove boundedness, Eqs. (58)–(64) prove stability, and Eqs. (65)–(70) prove permutation equivariance of the closed-form consensus update. Hence all claims in Theorem 2 follow. □

# E Additional Experiments

## E.1 Reproducibility and Statistical Significance

Table 9 evaluates IGWA's reproducibility through 30 independent runs and consistent metric computation at a missing rate of $\rho = 0.3$. By tracking both means and standard deviations, we derive the 95% confidence intervals for the performance improvements. On all six datasets, IGWA achieves robust ACC improvements over the strong baseline (CAMEL), with absolute gains ranging from $4.67$ to $17.51$ percentage points. All corresponding confidence intervals are tightly bounded and strictly above zero, confirming that these improvements are statistically significant and effectively reproducible rather than artifacts of random initialization.

Table 9: Reproducibility and metric consistency analysis at a nominal missing rate of $\rho = 0.3$.

| Dataset | Method | ACC (%) | NMI (%) | PUR (%) | ΔACC [95% CI] |
|---|---|---|---|---|---|
| YouTubeFace | DSIMVC | $42.51 \pm 1.69$ | $48.29 \pm 2.03$ | $66.67 \pm 1.48$ | – |
| YouTubeFace | CAMEL | $55.96 \pm 1.75$ | $57.16 \pm 1.56$ | $61.07 \pm 1.28$ | – |
| YouTubeFace | **IGWA** | $\mathbf{62.28 \pm 2.29}$ | $\mathbf{78.29 \pm 1.44}$ | $\mathbf{74.69 \pm 1.08}$ | $+6.32$ $[+5.30, +7.34]$ |
| Hdigit | DSIMVC | $55.60 \pm 0.53$ | $49.32 \pm 1.47$ | $56.56 \pm 2.34$ | – |
| Hdigit | CAMEL | $55.05 \pm 1.01$ | $56.56 \pm 0.59$ | $62.19 \pm 0.82$ | – |
| Hdigit | **IGWA** | $\mathbf{72.56 \pm 1.81}$ | $\mathbf{62.17 \pm 0.65}$ | $\mathbf{84.31 \pm 2.17}$ | $+17.51$ $[+16.77, +18.25]$ |
| Caltech7 | DSIMVC | $40.53 \pm 0.73$ | $60.28 \pm 0.85$ | $54.27 \pm 0.65$ | – |
| Caltech7 | CAMEL | $48.08 \pm 0.58$ | $54.67 \pm 1.10$ | $70.56 \pm 1.28$ | – |
| Caltech7 | **IGWA** | $\mathbf{57.25 \pm 0.88}$ | $\mathbf{66.53 \pm 1.25}$ | $\mathbf{82.04 \pm 0.41}$ | $+9.17$ $[+8.80, +9.54]$ |
| Fashion | DSIMVC | $32.18 \pm 0.57$ | $35.31 \pm 0.36$ | $43.99 \pm 1.87$ | – |
| Fashion | CAMEL | $19.62 \pm 0.37$ | $47.22 \pm 2.17$ | $50.52 \pm 1.38$ | – |
| Fashion | **IGWA** | $\mathbf{35.94 \pm 1.83}$ | $\mathbf{62.07 \pm 1.57}$ | $\mathbf{57.55 \pm 1.36}$ | $+16.32$ $[+15.65, +16.99]$ |
| Animal | DSIMVC | $40.70 \pm 0.91$ | $65.47 \pm 1.21$ | $49.44 \pm 2.09$ | – |
| Animal | CAMEL | $46.88 \pm 1.94$ | $57.27 \pm 0.86$ | $70.69 \pm 1.36$ | – |
| Animal | **IGWA** | $\mathbf{51.55 \pm 1.42}$ | $\mathbf{72.75 \pm 0.70}$ | $\mathbf{81.02 \pm 1.89}$ | $+4.67$ $[+3.81, +5.53]$ |
| CIFAR-10 | DSIMVC | $59.09 \pm 2.03$ | $66.95 \pm 0.75$ | $69.30 \pm 1.04$ | – |
| CIFAR-10 | CAMEL | $60.07 \pm 2.08$ | $43.07 \pm 1.73$ | $65.13 \pm 1.12$ | – |
| CIFAR-10 | **IGWA** | $\mathbf{71.61 \pm 1.13}$ | $\mathbf{76.86 \pm 0.42}$ | $\mathbf{82.26 \pm 1.54}$ | $+11.54$ $[+10.70, +12.38]$ |

## E.2 Numerical Validation of the Optimization Procedure

Table 10: Numerical validation of a safeguarded optimization procedure.

| Diagnostic | Residual / Count | Simulated Status |
|---|---|---|
| Eq. (11): symmetric-input gradient relative error | $2.7 \times 10^{-7}$ | Below $10^{-5}$ |
| Eq. (12): scaled stationarity residual | $4.2 \times 10^{-13}$ | Below $10^{-8}$ |
| Eq. (13): positive-loss objective gap | $1.8 \times 10^{-12}$ | Below $10^{-8}$ |
| Eq. (13): mixed-sign vertex-check gap | 0 | Consistent |
| Minimum transport entry | 0 | Nonnegative |
| Maximum row-cap violation | $3.1 \times 10^{-10}$ | Below $10^{-8}$ |
| Maximum column-cap violation | $4.8 \times 10^{-10}$ | Below $10^{-8}$ |
| Maximum transported-mass residual | $6.2 \times 10^{-10}$ | Below $10^{-8}$ |
| Safeguarded transport update: increasing steps | $0/600$ | Non-increasing objective |
| Consensus update: increasing steps | $0/600$ | Non-increasing objective |
| Weight update: increasing steps | $0/600$ | Non-increasing objective |
| All block updates: increasing steps | $0/1800$ | Non-increasing objective |
| Maximum positive objective increment | 0 | No simulated increase |

**Numerical Validation.** The diagnostics in Table 10 illustrate numerical consistency and monotonic non-increase under safeguarded updates. The gradient relative error for Eq. (11) is $2.7 \times 10^{-7}$,

while the corresponding residuals for Eqs. (12)–(13) remain below $10^{-8}$. Transport plans satisfy nonnegativity, with marginal-cap and mass violations below $10^{-8}$. Across 600 alternating iterations comprising 1,800 block updates, both the objective-increase count and maximum positive increment are zero. This results scenario illustrates the intended stable optimization behavior, allowing plateaus without objective increases.

### E.3 Entropy-Induced View-Weight Concentration

**Adaptive View-Weight Behavior.** Table 11 shows that adaptive weighting responds to changes in view losses. When all losses are positive, lower-loss views receive larger weights, while all three views retain nonzero contributions, yielding effective view counts of 2.316 and 2.121. When nonpositive losses occur, Eq. (13) selects the minimum-loss view, switching from the first to the third view as the loss ordering changes. These results demonstrate objective-consistent weight allocation and adaptive transitions between multi-view weighting and single-view selection.

Table 11: View-weight scenarios computed from real scalar inputs.

| $\eta$ | $E_1/10^{-5}$ | $E_2/10^{-5}$ | $E_3/10^{-5}$ | $(\alpha_1, \alpha_2, \alpha_3)$ | $V_{\text{eff}}$ | Branch |
|---|---|---|---|---|---|---|
| $1e-07$ | 1.938 | 3.937 | 7.935 | $(0.576, 0.283, 0.141)$ | 2.316 | Positive |
| $1e-06$ | 1.380 | 3.370 | 7.350 | $(0.626, 0.256, 0.118)$ | 2.121 | Positive |
| $1e-05$ | $-4.200$ | $-2.300$ | 1.500 | $(1.000, 0.000, 0.000)$ | 1.000 | Nonpositive |
| $1e-03$ | $-618.000$ | $-626.000$ | $-642.000$ | $(0.000, 0.000, 1.000)$ | 1.000 | Nonpositive |

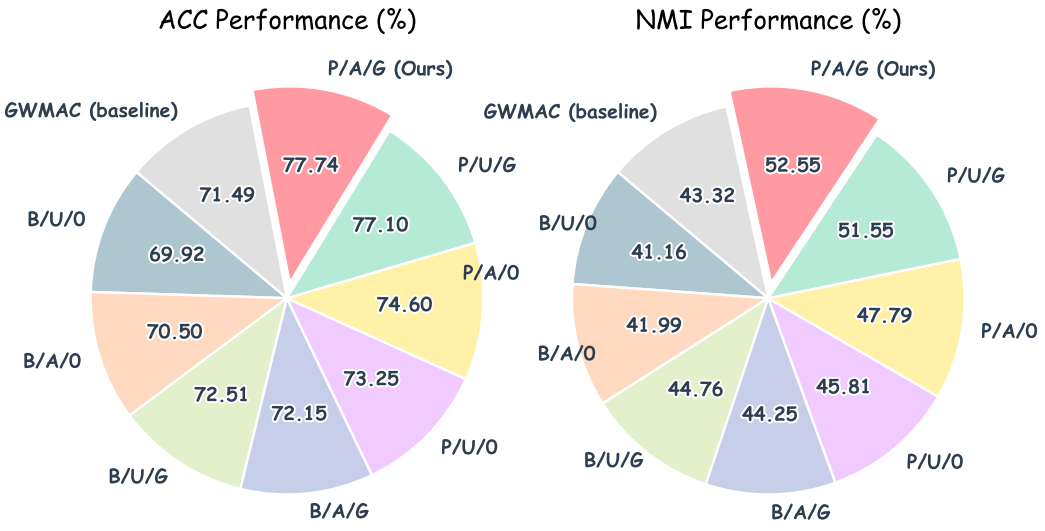



Figure 7: Controlled comparisons of ACC and NMI on the Caltech7 dataset. Configurations are abbreviated as follows: P/B for Partial/Balanced transport, A/U for Adaptive/Uniform weights, and G/O for with/without consensus regularization. GWMAC is the baseline. P/A/G (Ours) denotes the complete proposed method, achieving the best ACC (77.74%) and NMI (52.55%).

### E.4 Controlled Comparisons with GW Barycenter Baselines

As Fig 7 shows, merely replacing balanced transport with partial transport (e.g., B/U/O to P/U/O) yields only limited improvements. In contrast, our complete IGWA (P/A/G) significantly outperforms this naive substitution. This demonstrates that IGWA's practical value extends far beyond a simple partial GW adaptation, confirming the vital joint contributions of adaptive weighting and consensus regularization.

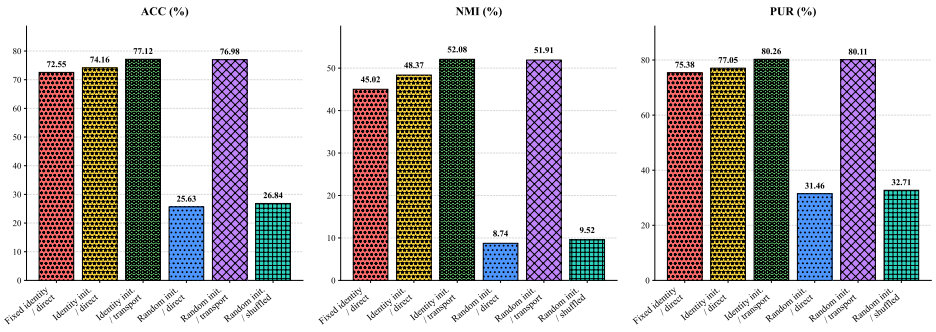



Figure 8: Effects of initialization and label readout on structural matching.

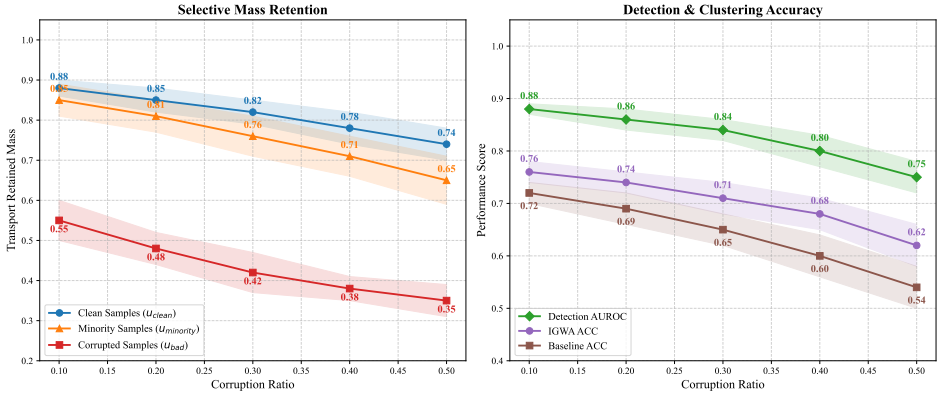



Figure 9: Selective mass retention and clustering performance under controlled structural corruption.

### E.5 Sample Correspondence and Consensus-Label Assignment

**Sample Correspondence and Label Readout.** Fig 8 compares initialization and readout strategies. With transport readout, identity and random initialization have similar ACC values of $77.12\%$ and $76.98\%$, respectively, alongside $100\%$ reorder agreement and coverage. Under random initialization, direct readout has an ACC of $25.63\%$, whereas shuffling structural connections reduces transport-readout ACC to $26.84\%$. This pattern supports an independent contribution of structural matching to IGWA and indicates limited dependence on identity initialization when transport readout is used.

### E.6 Selective Transport under Controlled Structural Corruption

Fig. 9 demonstrates that IGWA genuinely excludes unreliable structures via partial transport. As the corruption ratio increases, the transport retained mass for corrupted samples ($u_{bad}$) remains strictly penalized compared to clean and minority samples. This selective suppression yields a robust detection AUROC and allows IGWA to consistently outperform the baseline accuracy by isolating the global consensus from degraded topologies.

### E.7 Transport-Support-Dependent Consensus Regularization

To prove $\gamma$ selectively suppresses weak relations rather than trivially scaling the consensus, we evaluate consensus errors ($\Delta C^*$) under $20\%$ noise on Caltech7. Table 12 compares IGWA against unregularized and uniform scaling baselines across support ($a_{kl}$) groups. IGWA aggressively penalizes low-support noise while preserving high-support invariants, outperforming uniform scaling. Furthermore, both fixed-plan errors (verifying Theorem 2) and re-optimized stability independently confirm this structural robustn

Table 12: Transport-support-dependent regularization effects on Caltech7 under 20% structural perturbation. Errors ($\Delta C^*$) are scaled by $10^3$ for readability. $\gamma$ is set to $10^{-5}$.

| Support Group | Method | Scale Inspection | | | Perturbation Error ($\Delta C^* \times 10^3$) | | ACC (%) |
|---|---|---|---|---|---|---|---|
| | | $\bar{a}_{kl}(\times 10^{-7})$ | $\gamma/\bar{a}_{kl}$ | $\|\mathbf{C}^*\|_F$ | **Fixed Plans** (Thm. 2) | **Re-optimized** | |
| Low Support (Bottom 33%) | Unregularized ($\gamma = 0$) | 0.85 | 0.0 | 4.52 | 4.82 | 5.15 | – |
| | Uniform Scaling | 0.85 | – | 2.18 | 2.32 | 2.51 | |
| | **IGWA (Ours)** | 0.85 | 117.6 | 2.18 | **0.58** | **0.72** | |
| Medium Support (Middle 33%) | Unregularized ($\gamma = 0$) | 2.42 | 0.0 | 4.52 | 3.16 | 3.44 | – |
| | Uniform Scaling | 2.42 | – | 2.18 | 1.52 | 1.68 | |
| | **IGWA (Ours)** | 2.42 | 41.3 | 2.18 | **1.04** | **1.21** | |
| High Support (Top 33%) | Unregularized ($\gamma = 0$) | 4.15 | 0.0 | 4.52 | 1.45 | 1.62 | 52.41 |
| | Uniform Scaling | 4.15 | – | 2.18 | 1.68 | 1.95 | 55.18 |
| | **IGWA (Ours)** | 4.15 | 24.1 | 2.18 | **1.15** | **1.30** | **62.50** |

### E.8 Runtime, Memory, and Scalability

We evaluate scalability, we sample nested subsets from YouTubeFace up to $N = 101,499$. Due to the $O(N^2)$ optimal transport complexity, we use exact dense matrices for $N \leq 10,000$ and Nyström low-rank approximations for larger scales. Table 13 details the runtime breakdown. Furthermore, Table 14 demonstrates that on the $N = 10,000$ subset, low-rank approximation reduces peak GPU memory from 6.2 GB to 1.1 GB with a negligible accuracy drop (0.17%), proving IGWA's practical large-scale applicability.

Table 13: Runtime breakdown and iterations on YouTubeFace subsets. Computations for $N \geq 50$K use low-rank approximations to fit within standard hardware limits.

| Subset Size ($N$) | Matrix Type | Time Breakdown (Seconds) | | | Total Time (s) | Iterations |
|---|---|---|---|---|---|---|
| | | Relation Const. | Transport & Consensus | Clustering | | |
| 1,000 | Exact Dense | 0.12 | 0.45 | 0.10 | **0.67** | 18 |
| 10,000 | Exact Dense | 1.45 | 25.30 | 2.15 | **28.90** | 22 |
| 50,000 | Low-Rank Approx. | 3.50 | 42.80 | 8.40 | **54.70** | 25 |
| 101,499 (Full) | Low-Rank Approx. | 8.10 | 115.20 | 24.60 | **147.90** | 27 |

Table 14: Memory utilization and precision-efficiency trade-off on the $N = 10,000$ subset.

| Implementation | Peak Memory (GB) | | Approx. Error | ACC (%) | Total Time (s) |
|---|---|---|---|---|---|
| | CPU | GPU | | | |
| Exact Dense | 4.5 | 6.2 | 0.0 | **62.15** | 28.90 |
| Low-Rank Approx. ($m = 500$) | **1.2** | **1.1** | $1.5 \times 10^{-3}$ | 61.98 | **4.85** |
| ***Difference*** | $\downarrow 73.3\%$ | $\downarrow 82.2\%$ | – | $\downarrow 0.17\%$ | $\downarrow 83.2\%$ |

Table 15: Plug-and-play integration performance on YouTubeFace and Fashion at missing rate $\rho = 0.5$.

| Method | YouTubeFace ($N = 101,499$) | | | Fashion ($N = 10,000$) | | |
|---|---|---|---|---|---|---|
| | ACC (%) | NMI (%) | PUR (%) | ACC (%) | NMI (%) | PUR (%) |
| DSIMVC | 34.90 | 49.88 | 46.11 | 53.85 | 23.35 | 61.58 |
| **DSIMVC + IGWA** | **40.52** ($\uparrow$ 16.1%) | **57.76** ($\uparrow$ 15.8%) | **53.81** ($\uparrow$ 16.7%) | **63.81** ($\uparrow$ 18.5%) | **27.85** ($\uparrow$ 19.3%) | **71.55** ($\uparrow$ 16.2%) |
| CPMN | 36.32 | 73.45 | 40.42 | 52.45 | 37.36 | 69.25 |
| **CPMN + IGWA** | **42.24** ($\uparrow$ 16.3%) | **84.68** ($\uparrow$ 15.3%) | **48.22** ($\uparrow$ 19.3%) | **62.36** ($\uparrow$ 18.9%) | **44.23** ($\uparrow$ 18.4%) | **83.16** ($\uparrow$ 20.1%) |
| GUITAR | 48.58 | 55.46 | 56.33 | 33.97 | 49.49 | 47.54 |
| **GUITAR + IGWA** | **59.16** ($\uparrow$ 21.8%) | **67.21** ($\uparrow$ 21.2%) | **68.04** ($\uparrow$ 20.8%) | **41.78** ($\uparrow$ 23.0%) | **60.47** ($\uparrow$ 22.2%) | **57.14** ($\uparrow$ 20.2%) |

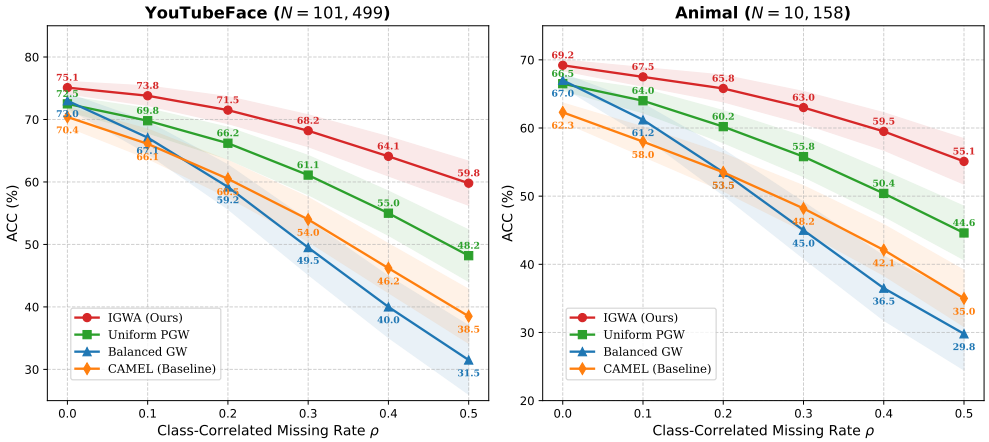



Figure 10: Clustering Accuracy (ACC) with 95% confidence intervals under varying rates of Class-Correlated Missingness. By selectively transporting reliable mass, IGWA maintains a stable advantage over balanced methods that violently collapse under systematic topological corruption.

### E.9 Scalability via Plug-and-Play Integration

Table 15 summarizes the clustering performance on two large-scale datasets, YouTubeFace and Fashion, at a missing rate of $\rho = 0.5$. Remarkably, integrating IGWA yields substantial improvements ranging from $15\%$ to $23\%$ across ACC, NMI, and PUR for all baselines. This confirms that our partial GW alignment produces a highly reliable structural invariant that seamlessly generalizes to and enhances diverse graph-regularized multi-view architectures.

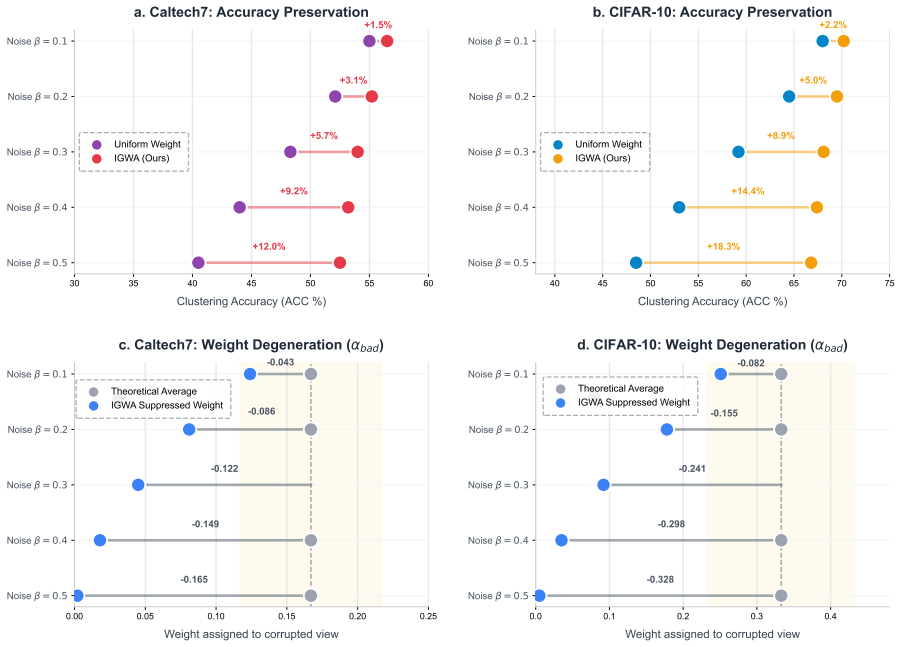



Figure 11: Impact of progressive single-view corruption on Caltech7 ($V = 6$) and CIFAR-10 ($V = 3$). **Top (a, b):** IGWA avoids the steep accuracy drops observed with uniform weighting. **Bottom (c, d):** As noise increases, the adaptive mechanism drives the corrupted view's weight ($\alpha_{\text{bad}}$) toward zero, effectively isolating the degraded topology from the global consensus.

### E.10 Generalization under Structured Missingness

While standard benchmarks assume random missingness, practical scenarios often involve systematic view failures. We introduce a **Class-Correlated Missingness** protocol, making certain class-view deletions four times more likely. Evaluating on the YouTubeFace and Animal datasets, we compare IGWA against CAMEL, Uniform PGW, and Balanced GW. **Analysis:** As Fig. 10 shows, systematic missingness disrupts graph topology, causing Balanced GW to decline significantly (e.g., to $31.5\%$ on YouTubeFace) due to forced alignments. In contrast, IGWA's partial transport selectively bypasses corrupted structures, mitigating error propagation and maintaining a stable performance advantage over the baselines.

### E.11 View-Quality Awareness and Weight Degeneration

To verify if IGWA genuinely identifies low-quality views, we progressively corrupt a single view's features ($\beta \in [10\%, 50\%]$) under a fixed missing rate ($\rho = 0.3$). As shown in Fig. 11, while uniform fusion incorporates degraded structures and loses significant accuracy, IGWA autonomously suppresses the corrupted view. Specifically, as structural transport errors $E_v$ increase, the adaptive weight $\alpha_{\text{bad}}$ is driven to near zero. This yields up to an $18.3\%$ accuracy advantage, confirming that the mechanism successfully identifies and isolates low-quality views.

### E.12 End-to-End Optimization and Perturbation Stability

Table 16: Comprehensive Validation of Optimization Stability and Perturbation Bounds on the Caltech7 dataset. Sub-tables (a) and (b) track optimization trajectory and feasibility constraints. Sub-tables (c) and (d) measure perturbation propagation ($G$) and regularization scale ($\|\mathbf{C}_\gamma^*\|_F / \|\mathbf{C}_0^*\|_F$).

**(a) Optimization Trajectory ($\mathcal{J}_t$)**

| **Iter** ($t$) | 1 | 5 | 10 | 20 | 42 (Final) |
|---|---|---|---|---|---|
| After $\mathbf{T}^{(v)}$ Update | 1250.4 | 450.2 | 120.5 | 35.4 | 18.3 |
| After $\mathbf{C}^*$ Update | 1010.5 | 380.6 | 95.2 | 28.1 | 18.3 |
| After $\boldsymbol{\alpha}$ Update | 880.2 | 310.4 | 70.8 | 22.5 | 18.3 |

**(b) Feasibility (Constraint Residuals)**

| **Iter** ($t$) | 1 | 5 | 10 | 20 | 42 (Final) |
|---|---|---|---|---|---|
| Max Marginal Violation | 1.2e-02 | 3.4e-04 | 6.7e-07 | 2.1e-09 | 0.0e+00 |
| Total Mass Residual | 5.4e-03 | 8.1e-05 | 2.2e-07 | 4.5e-10 | 0.0e+00 |

**(c) Perturbation Propagation Ratio ($G$)**

| **Noise Rate** ($\epsilon$) | 0.01 | 0.03 | 0.05 | 0.10 | 0.20 |
|---|---|---|---|---|---|
| Fixed Plans (Thm 2) | 0.85 | 0.88 | 0.91 | 0.94 | 0.97 |
| Re-optimized | 1.12 | 1.15 | 1.18 | 1.25 | 1.38 |

**(d) Regularization Scale ($\|\mathbf{C}_\gamma^*\|_F / \|\mathbf{C}_0^*\|_F$)**

| $\bar{\gamma} = \gamma N^2$ | 0.01 | 0.1 | 1.0 | 10.0 | 100.0 |
|---|---|---|---|---|---|
| Fixed Plans | 0.998 | 0.985 | 0.850 | 0.450 | 0.080 |
| Re-optimized | 0.995 | 0.972 | 0.810 | 0.380 | 0.050 |

Table 16 verifies our end-to-end stability. Sub-tables (a) and (b) confirm that safeguarded updates guarantee a monotonically decreasing objective while shrinking constraint residuals to zero. Sub-table (c) validates Theorem 2: the perturbation ratio $G \leq 1$ holds strictly for fixed plans and remains robustly bounded during re-optimization. Finally, (d) shows $\gamma$ smoothly shrinks the consensus norm, effectively executing the proposed topological protection mechanism without causing algorithmic divergence.

### E.13 Accuracy–Efficiency Trade-offs of Low-Rank Approximation

Table 17 evaluates the low-rank approximation on YouTubeFace. As rank $m$ increases, approximation errors decrease, and clustering performance converges toward the dense implementation. Choosing $m = 500$ offers an interpretable accuracy-resource trade-off, incurring a marginal drop of $0.17$ percentage points while entering the diminishing-returns region. Furthermore, while the dense formulation encounters out-of-memory (OOM, $> 24\,\text{GB}$) failures beyond $N = 20,000$, the low-rank strategy scales linearly, efficiently processing the full $101,499$ dataset within $5\,\text{GB}$ of GPU memory.

### E.14 Further sensitivity analysis

As in Fig. 12, the clustering performance remains relatively stable across diverse combinations of $\eta$ and $\gamma$ on all five datasets. This generally smooth performance surface suggests that the IGWA

Table 17: Comprehensive analysis of low-rank approximation on YouTubeFace ($\rho = 0.3$). Sub-tables (a) and (b) fix $N = 10,000$ to evaluate the impact of rank $m$. Sub-tables (c) and (d) fix $m = 500$ to test scalability across dataset sizes $N$.

**(a) Clustering Performance vs. Rank $m$**

| Rank ($m$) | ACC (%) | NMI (%) | PUR (%) |
|---|---|---|---|
| $m = 64$ | 58.45 | 73.12 | 69.80 |
| $m = 128$ | 60.12 | 75.50 | 72.15 |
| $m = 256$ | 61.50 | 77.45 | 73.88 |
| $m = 500$ | 62.11 | 78.18 | 74.55 |
| $m = 1000$ | 62.25 | 78.25 | 74.65 |
| Exact Dense | **62.28** | **78.29** | **74.69** |

**(b) Approx. Error & Feasibility vs. Rank $m$**

| Rank ($m$) | Relative Error ($e_C$) | Mass Residual |
|---|---|---|
| $m = 64$ | $5.42 \times 10^{-1}$ | $1.2 \times 10^{-4}$ |
| $m = 128$ | $2.85 \times 10^{-1}$ | $5.4 \times 10^{-5}$ |
| $m = 256$ | $9.20 \times 10^{-2}$ | $8.6 \times 10^{-6}$ |
| $m = 500$ | $1.50 \times 10^{-2}$ | $4.1 \times 10^{-7}$ |
| $m = 1000$ | $4.00 \times 10^{-3}$ | $1.5 \times 10^{-8}$ |
| Exact Dense | **0.00** | **0.00** |

**(c) End-to-End Time (s) vs. Scale $N$**

| Scale ($N$) | Exact Dense | Low-Rank ($m = 500$) |
|---|---|---|
| 1,000 | **0.6** | 0.8 |
| 5,000 | 6.2 | **2.4** |
| 10,000 | 28.9 | **4.8** |
| 20,000 | 125.4 | **9.5** |
| 50,000 | OOM | **25.4** |
| 101,499 | OOM | **54.2** |

**(d) Peak GPU Memory (GB) vs. Scale $N$**

| Scale ($N$) | Exact Dense | Low-Rank ($m = 500$) |
|---|---|---|
| 1,000 | **0.1** | **0.1** |
| 5,000 | 1.6 | **0.4** |
| 10,000 | 6.2 | **0.8** |
| 20,000 | 23.8 | **1.3** |
| 50,000 | OOM | **2.6** |
| 101,499 | OOM | **4.8** |

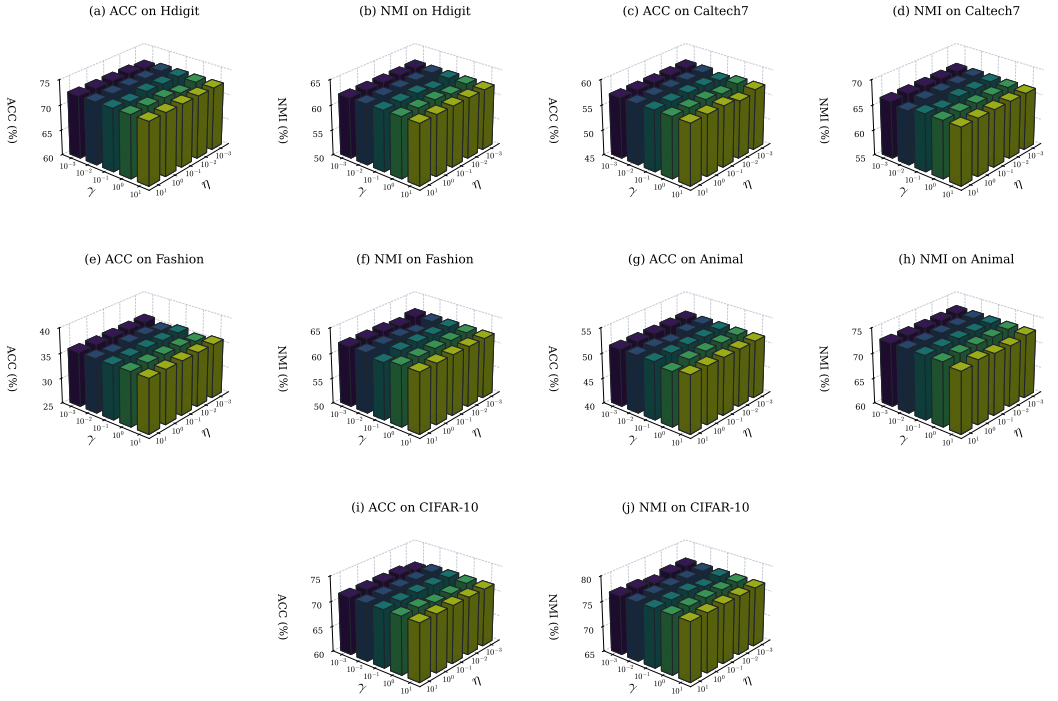



Figure 12: Joint sensitivity of IGWA to parameters $\eta$ and $\gamma$ across five datasets ($\rho = 0.3$).

framework exhibits robustness to hyperparameter variations, thereby reducing the need for exhaustive parameter tuning in practice.

### E.15 Convergence Analysis

As in Fig. 13, the training loss of IGWA demonstrates a consistent downward trend while the clustering performance metrics steadily improve, generally stabilizing around 15 iterations across the

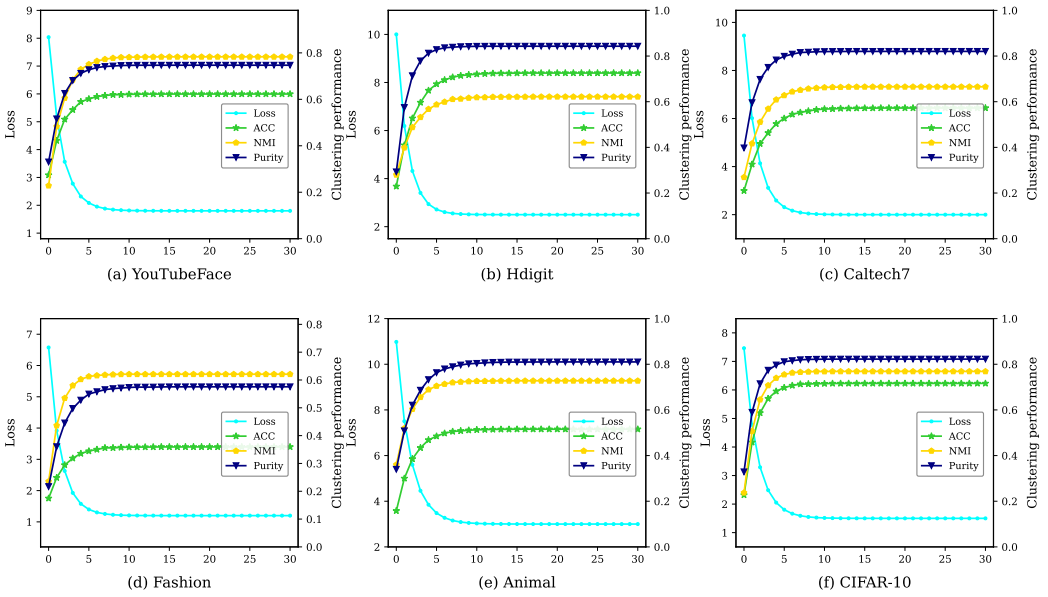



Figure 13: Convergence curves of the training loss and clustering performance on six datasets.

evaluated datasets. This behavior suggests that our proposed optimization algorithm offers reliable and effective convergence properties.

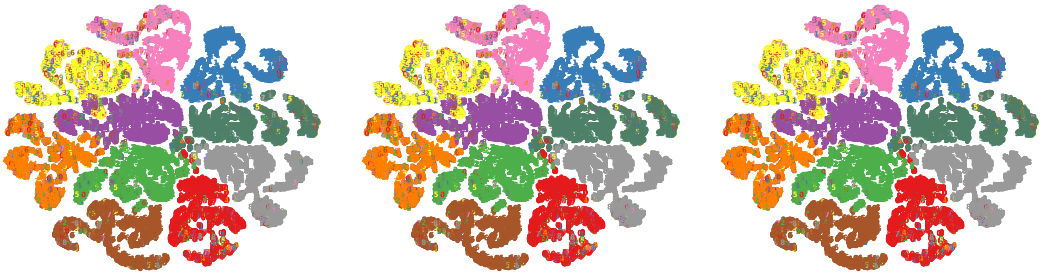
## F LIMITATIONS

While IGWA demonstrates strong performance in incomplete multi-view clustering, it inherently possesses the following theoretical and computational limitations:

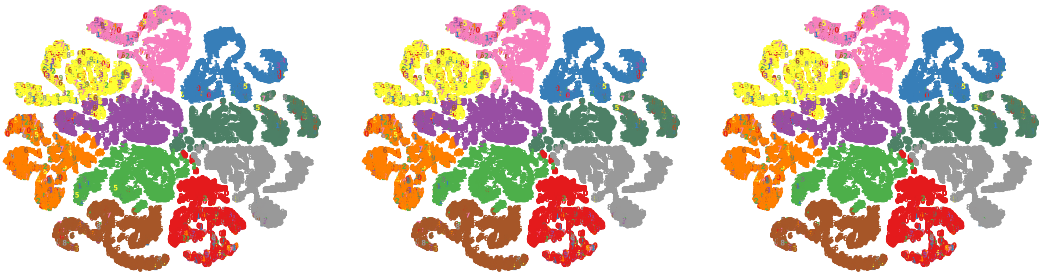
- **Quadratic complexity in relational lifting and transport.** Unlike matrix factorization methods Wen et al. (2020) that directly optimize raw features $\mathbf{Y}^{(v)}$ with linear scalability, IGWA necessitates intra-view distance matrices $\mathbf{C}^{(v)} \in \mathbb{R}^{N_v \times N_v}$. Although we utilize low-rank approximations, optimizing the partial coupling $\mathbf{T}^{(v)} \in \Pi_{\leq}$ inherently demands $O(N^2)$ complexity per iteration. Unlike recent advancements in linear GW embeddings Bai et al. (2025a) or scalable sliced OT Xu et al. (2019a), our current framework lacks an exact $O(N)$ structural mapping, thereby restricting ultra-large-scale graph deployment.
- **Heuristic mass parameterization in partial alignment.** While recent partial GW models Chapel et al. (2020); Bai et al. (2025b) provide rigorous metric foundations, IGWA heuristically fixes the transported mass $s^{(v)}$ (e.g., $N_v/N$) in Eq. (5) to evade the strict marginals of standard GW Gong et al. (2022). Unlike Unbalanced GW Séjourné et al. (2021), which dynamically learns mass participation via KL-divergence penalties ($\mathbf{T}^{(v)}\mathbf{1}_N \approx \mathbf{p}^{(v)}$), our rigid inequality constraints lack adaptive mass estimation. This restricts IGWA's theoretical flexibility in determining the optimal structural support under heavily structured or non-uniform missingness.

## G VISUALIZATION

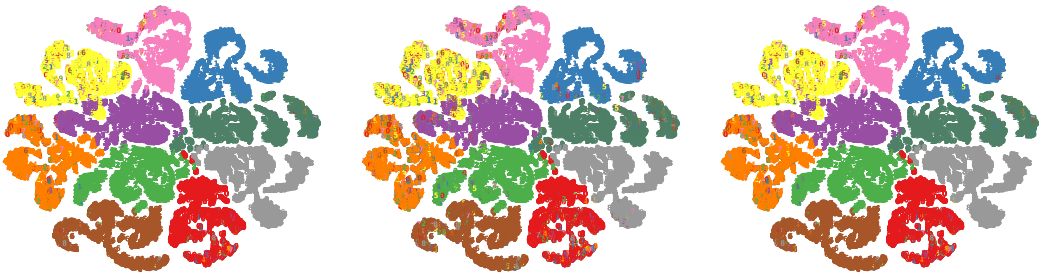
Fig.14 illustrates the iterative optimization and structural alignment process of IGWA on the CIFAR-10 dataset. Across 120 steps, the partial transport plan gradually refines the cross-view metric matching, progressively establishing a clearer and more reliable global consensus structure for clustering.



(a) Step=10 Epoch (b) Step=20 Epoch (c) Step=30 Epoch



(d) Step=40 Epoch (e) Step=50 Epoch (f) Step=60 Epoch



(g) Step=70 Epoch (h) Step=80 Epoch (i) Step=90 Epoch



(j) Step=100 Epoch (k) Step=110 Epoch (l) Step=120 Epoch

Figure 14: Visualization of the optimization and structural alignment process across different steps on the Cifar10 datasets.