

Gromov–Wasserstein Methods for Multi-View Relational Embedding and Clustering

Rafael Pereira Eufrazio, Eduardo Fernandes Montesuma, and Charles Casimiro Cavalcante

Abstract—Learning low-dimensional representations from multi-view relational data is challenging when underlying geometries differ across views. We propose Bary-GWMDS, a Gromov-Wasserstein-based method that operates directly on distance matrices to learn a consensus embedding preserving shared relational structure. By leveraging intrinsic distances, the approach naturally handles nonlinear distortions across views. We also introduce Mean-GWMDS-C, a clustering-oriented formulation that averages distance matrices and learns reduced-support representations via a consensus Gromov-Wasserstein transport. Experiments on synthetic and real-world datasets show that the proposed framework yields stable and geometrically meaningful embeddings.

Keywords—Gromov-Wasserstein distance, manifold learning, multi-view learning, relational embedding, prototype-based clustering

I. INTRODUCTION

Learning low-dimensional representations from relational data is a fundamental problem in signal processing and machine learning, with applications ranging from manifold learning to multi-modal data analysis [1]. In signal processing applications, multi-view data naturally arise from heterogeneous sensing modalities, such as sensor networks, image descriptors, and time-series measurements, where each view captures complementary aspects of the underlying signal structure. In many practical scenarios, the same set of samples is observed through multiple views, each inducing a distinct relational structure [1], [2]. Effectively aggregating such heterogeneous view-dependent information into a single embedding remains challenging, especially when the underlying geometries differ across views.

Classical multi-view embedding methods typically rely on Euclidean averaging or linear alignment strategies, which may fail under nonlinear distortions or heterogeneous geometric structures [3]–[6]. In contrast, the Gromov-Wasserstein (GW) distance provides an intrinsic notion of similarity between metric spaces, enabling the comparison and aggregation of relational data independently of the ambient representation [7], [8]. Unlike methods that prioritize local neighborhoods, GW aligns global pairwise relations across heterogeneous spaces

without explicit pointwise correspondences, at the cost of repeatedly solving computationally expensive transport problems.

In this work, we propose Bary-GWMDS, which aggregates view-specific distance matrices through a GW barycenter and then applies Gromov-Wasserstein MDS (GW-MDS) to the consensus structure. Unlike the single-view formulations in [9], [10], it addresses multiple relational views; relative to the multi-view methods in [11], it uses a GW-barycentric consensus rather than arithmetic distance averaging or joint view-wise optimization. We also introduce Mean-GWMDS-C, which extends Mean-GWMDS to reduced-support clustering, with assignments induced by a single consensus transport plan.

Experimental results on synthetic and real-world datasets demonstrate the effectiveness of the proposed methods.

The remainder of this paper is organized as follows. Section II introduces the background, including the problem setup and the GW formulation. Section III presents the proposed multi-view framework, detailing both the Bary-GWMDS and Mean-GWMDS-C approaches. Section IV describes the experimental setup and discusses the results on synthetic and real-world multi-view datasets. Finally, Section V concludes the paper.

II. BACKGROUND

In this section, we present the necessary background on GW geometry and multi-view relational modeling, starting with the problem formulation.

A. Problem Setup

We consider a multi-view dataset consisting of S views defined over a common set of n samples. Each view is represented by a relational metric space $(\mathcal{X}, D^{(s)})$, where $D^{(s)} \in \mathbb{R}^{n \times n}$ denotes the pairwise distance matrix associated with view s .

Our goal is to learn a low-dimensional embedding that captures a consensus relational structure across views, based solely on their pairwise distance representations. Such representations can also be leveraged for downstream tasks such as clustering.

B. Gromov-Wasserstein Framework

The GW distance compares probability distributions defined on different metric spaces by aligning their intrinsic relational structures [9].

Given two finite metric measure spaces $(\mathcal{X}, D_X, \hat{\mu})$ and $(\mathcal{Y}, D_Y, \hat{\nu})$, with $\hat{\mu} = \sum_{i=1}^n a_i \delta_{x_i}$ and $\hat{\nu} = \sum_{j=1}^m b_j \delta_{y_j}$, where $\mathbf{a} = (a_i)_{i=1}^n$ and $\mathbf{b} = (b_j)_{j=1}^m$ are their probability weight

R. P. Eufrazio is with the Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Canindé, Ceará, Brazil, and with the Federal University of Ceara, Fortaleza, Ceará, Brazil, e-mail: rafael.eufrazio@ifce.edu.br.

E. F. Montesuma is with Sigma Nova, Paris, France, e-mail: eduardo.montesuma@sigmanova.ai.

C. C. Cavalcante is with the Federal University of Ceara, Fortaleza, Ceará, Brazil, e-mail: charles@ufc.br.

This work was partially supported by CNPq Procs. 308512/2023-5 and 420341/2025-0, CNPq/INCT STREAM (Signal processing and TRansmission for Environmental Analysis and Monitoring) 409179/2024-8 and by CAPES - Finance Code 001.

vectors, let $D_X^{ik} = D_X(x_i, x_k)$ and $D_Y^{j\ell} = D_Y(y_j, y_\ell)$. The squared GW distance is defined as

$$\text{GW}^2(\hat{\mu}, \hat{\nu}) = \min_{\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b})} \sum_{i,j,k,\ell} \left(D_X^{ik} - D_Y^{j\ell} \right)^2 T_{ij} T_{k\ell}. \quad (1)$$

where $\Pi(\mathbf{a}, \mathbf{b}) = \{ \mathbf{T} \in \mathbb{R}_+^{n \times m} : \mathbf{T} \mathbf{1}_m = \mathbf{a}, \mathbf{T}^\top \mathbf{1}_n = \mathbf{b} \}$ is the set of admissible transport plans. Each entry T_{ij} represents the amount of probability mass transported from x_i to y_j .

The semi-relaxed GW (srGW) variant fixes only the source marginal, $\Pi^{\text{sr}}(\mathbf{a}) = \{ \mathbf{T} \in \mathbb{R}_+^{n \times K} : \mathbf{T} \mathbf{1}_K = \mathbf{a} \}$, enabling a reduced-support representation with $K \ll n$ prototypes [12], [13].

Relational structures can be aggregated via the GW barycenter, defined as

$$\bar{D} = \arg \min_{D \in \mathbb{R}^{n \times n}} \sum_{s=1}^S \lambda_s \text{GW}^2(D^{(s)}, D), \quad (2)$$

where $\lambda_s \geq 0$ and $\sum_{s=1}^S \lambda_s = 1$. In this work, we use uniform weights $\lambda_s = 1/S$ [14].

C. GW-MDS Embedding

Specifically, we seek an embedding $Y \in \mathbb{R}^{n \times d}$ whose induced distance matrix D_Y best matches the relational structure encoded in D [9], [10]:

$$\min_{Y \in \mathbb{R}^{n \times d}} \text{GW}^2(D, D_Y), \quad D_Y(i, j) = \|y_i - y_j\|_2. \quad (3)$$

Here, D and D_Y denote pairwise distance matrices associated with empirical measures supported on $\mathcal{X}$ and $\mathcal{Y}$, respectively. This formulation can be interpreted as a generalized multidimensional scaling problem, where relational alignment is mediated by an optimal transport plan $\mathbf{T}$.

More explicitly, the objective depends jointly on the embedding Y and the transport plan $\mathbf{T}$, leading to a non-convex optimization problem due to their coupling, which is addressed via an alternating optimization strategy.

The optimization is performed directly over the embedding coordinates Y using gradient-based updates to minimize $\text{GW}^2(D, D_Y)$. In particular, the embedding is iteratively updated as

$$Y^{(t+1)} = Y^{(t)} - \eta \nabla_Y \text{GW}^2(D, D_Y^{(t)}), \quad (4)$$

while the transport plan $\mathbf{T}$ is recomputed at each iteration. This alternating procedure jointly estimates the embedding and the relational correspondence between structures.

Importantly, no explicit normalization is applied to D_Y , as the scale of the embedding is implicitly determined by the GW objective.

III. PROPOSED MULTI-VIEW FRAMEWORK

In this section, we present a GW-based framework for multi-view data analysis. We first introduce Bary-GWMDS for embedding, followed by Mean-GWMDS-C for clustering¹.

¹Code is available on GitHub:

<https://github.com/casape/bary-gwmds-mean-gwmds-c>

Algorithm 1 Bary-GWMDS

Input: $\{D^{(s)}\}_{s=1}^S$, embedding dimension d

Output: $Y \in \mathbb{R}^{n \times d}$

- 1: Compute the GW barycenter $\bar{D}$ of $\{D^{(s)}\}_{s=1}^S$
 - 2: Initialize $Y \in \mathbb{R}^{n \times d}$
 - 3: Optimize Y by minimizing $\text{GW}(\bar{D}, D_Y)$
 - 4: **return** Y
-

Algorithm 2 Mean-GWMDS-C

Input: $\{D^{(s)}\}_{s=1}^S$, prototypes K , dim d

Output: $Y \in \mathbb{R}^{K \times d}$

- 1: $\bar{D} \leftarrow \frac{1}{S} \sum_{s=1}^S D^{(s)}$
 - 2: Initialize $Y \in \mathbb{R}^{K \times d}$
 - 3: Optimize Y by minimizing $\text{GW}(\bar{D}, D_Y)$
 - 4: **return** Y (labels: $\hat{c}(i) = \arg \max_k \mathbf{T}_{ik}$)
-

A. Bary-GWMDS

We introduce Bary-GWMDS (Barycentric Gromov-Wasserstein Multidimensional Scaling) (Algorithm 1), a two-stage method for multi-view embedding. The proposed approach first aggregates view-dependent relational structures using a GW barycenter [14], and then computes a low-dimensional embedding by applying GW-MDS [9] to the resulting consensus metric.

More specifically, given a set of distance matrices $\{D^{(s)}\}_{s=1}^S$, the GW barycenter is used to compute a consensus distance matrix $\bar{D}$ that captures the shared relational structure across views. Subsequently, a low-dimensional embedding Y is obtained by solving a GW-MDS problem based on $\bar{D}$.

B. Mean-GWMDS-C

Beyond full-support embedding, we consider a clustering-oriented multi-view formulation, termed Mean-GWMDS-C (Mean Gromov-Wasserstein Multidimensional Scaling for Clustering) (Algorithm 2), which aims at learning compact representations defined over a reduced set of representative points. In this setting, the latent embedding $Y \in \mathbb{R}^{K \times d}$, with $K \ll n$, encodes the geometry of learned prototypes rather than individual samples.

Mean-GWMDS-C operates by first aggregating the view-dependent relational structures through direct averaging of the corresponding distance matrices. Given the averaged multi-view distance matrix, a reduced-support relational representation is learned using a (semi-relaxed) GW-based [7], [12] objective, and the prototype embedding Y is subsequently optimized so as to induce a distance matrix D_Y consistent with the aggregated relational geometry. Solving GW problems independently across views would induce multiple, view-dependent transport plans, hindering coherent clustering. Mean-GWMDS-C resolves this issue by aggregating relational structures prior to optimization, yielding a single consensus transport plan and favoring structural compactness over point-wise fidelity.

After optimization, Mean-GWMDS-C yields a transport plan $\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b}) \subset \mathbb{R}_+^{n \times K}$, where T_{ik} represents the mass

transported from sample i to prototype k . Hard and soft assignments are respectively given by $\hat{c}(i) = \arg \max_k T_{ik}$ and $P_{ik} = T_{ik}/a_i$ [13]. Thus, clustering is induced by transport to relational prototypes without requiring explicit Euclidean centroids or decision boundaries.

As such, Mean-GWMDs-C provides a complementary perspective to Bary-GWMDs, targeting scenarios in which reduced representations and clustering-oriented analysis are preferred over full-support embeddings.

IV. EXPERIMENTS AND DISCUSSION

In this section, we evaluate the proposed framework through experiments, considering both embedding quality and clustering performance.

A. Bary-GWMDs with Geodesic Distances

We begin by evaluating the proposed Bary-GWMDs framework. In all experiments, each view is represented by a pairwise distance matrix computed using geodesic distances, which better capture the intrinsic geometry of nonlinear manifolds. Geodesic distances are obtained by constructing a neighborhood graph and computing shortest-path distances.

Preliminary experiments indicate that using Euclidean distances in the barycentric aggregation leads to unstable or distorted consensus structures. In particular, Euclidean metrics tend to overemphasize local linearity and fail to preserve global manifold relationships, which adversely affects the GW barycenter computation.

By contrast, geodesic distances yield well-structured barycenters that preserve the intrinsic relational organization common to all views. This choice results in stable consensus distance matrices and enables the subsequent GW-MDS optimization to produce meaningful low-dimensional embeddings.

To assess the effectiveness of the Bary-GWMDs approach, we compare it with Multi-ISOMAP [15], a multi-view method based on geodesic distances. Multi-ISOMAP extends the classical ISOMAP framework to the multi-view setting. This baseline is particularly relevant to our study, as both methods rely on geodesic distances to capture the intrinsic geometry of nonlinear manifolds.

B. Multi-View Construction

To evaluate the proposed methods under controlled geometric variability, we consider synthetic multi-view datasets generated from a common underlying manifold.

The two views are generated from the same manifold through a 40° rotation about the z -axis and the scaling-shearing map $A = \begin{bmatrix} 1.8 & 0.3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix}$, respectively; the former preserves distances, whereas the latter distorts them while retaining their topology.

Table I reports view-wise and mean Pearson correlation [16] between the learned embeddings and the original geodesic distance matrices on synthetic manifolds. Fig. 2 illustrates the Swiss roll embeddings obtained from each view. For Bary-GWMDs, correlations differ substantially across views (0.96 for View 1 and 0.26 for View 2), reflecting its focus

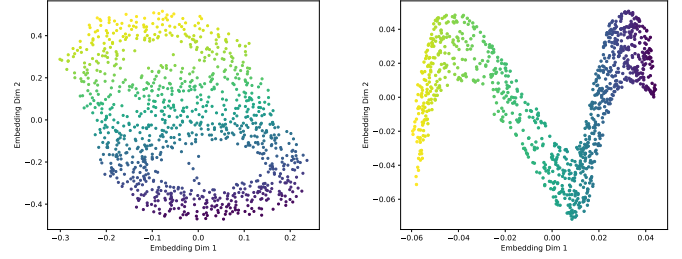


FIG. 1: Swiss roll embeddings: **Left:** Bary-GWMDs embedding. **Right:** Multi-ISOMAP embedding.

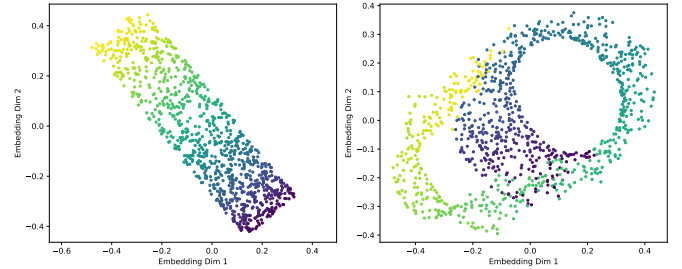


FIG. 2: Embeddings obtained separately from each view of the Swiss roll dataset. **Left:** embedding View 1. **Right:** embedding View 2.

on preserving view-specific relational structure rather than enforcing uniform alignment.

Multi-ISOMAP yields more balanced correlations (0.74 and 0.67), but fails to preserve the intrinsic Swiss roll geometry due to direct distance averaging, which blurs view-dependent relational information. Overall, these results show that correlation alone does not guarantee geometric fidelity, and that barycentric optimal transport enables more faithful manifold preservation.

C. Mean-GWMDs-C: Clustered Multi-View Representations

We now consider a clustering-oriented formulation of the proposed framework, termed Mean-GWMDs-C, which learns compact multi-view representations using a reduced set of relational prototypes. Inspired by the reduced-support interpretation of semi-relaxed GW, the method directly averages the view-wise geodesic distance matrices and optimizes a balanced GW objective with fixed uniform marginals.

In this setting, cluster assignments are obtained from the transport plan mapping samples to prototypes. We evaluate Mean-GWMDs-C on several real-world multi-view datasets, including MultiFeature (MulFe) [17], MSRC [18], Handwritten (Han) [19], and Caltech101-7 (Calt) [20]. The proposed method is compared with Multi-View K-Means [21] and Multi-View Spectral Clustering [22], using the same number of clusters and external evaluation criteria.

1) *Evaluation Metrics:* Clustering performance is assessed using Normalized Mutual Information (NMI) [23] and Adjusted Rand Index (ARI) [24], with higher values indicating better agreement with ground-truth labels. Table II reports NMI and ARI clustering results on four multi-view datasets

		S-curve	Swiss Roll	Mobius
Bary-GWMDS	view 1	0.98	0.96	0.93
	view 2	0.91	0.26	0.96
	mean	0.945	0.610	0.945
Multi-Isomap	view 1	0.80	0.74	0.93
	view 2	0.85	0.67	0.95
	mean	0.825	0.705	0.940

TABLE I: View-wise and mean correlation between learned embeddings.

		MulFe	Calt	Han	MSRC
Mean-GWMDS-C	NMI	0.942	0.260	0.773	0.463
	ARI	0.958	0.165	0.730	0.341
MV KMeans	NMI	0.759	0.240	0.549	0.264
	ARI	0.773	0.185	0.382	0.129
MV Spectral Clust.	NMI	0.933	0.249	0.851	0.577
	ARI	0.946	0.174	0.793	0.495

TABLE II: Clustering performance comparison in terms of NMI and ARI.

with distinct geometric and semantic characteristics. On the Multiple Features dataset, which exhibits a strong shared geometry across views, Mean-GWMDS-C achieves the best performance, attaining the highest NMI (0.942) and ARI (0.958). A similar behavior is observed on Caltech101-7, where Mean-GWMDS-C obtains the highest NMI, indicating the effectiveness of geometry-aware relational aggregation under heterogeneous visual descriptors.

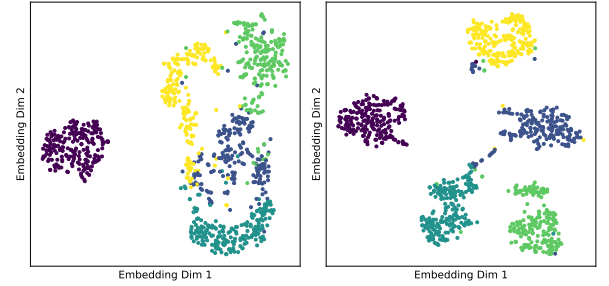
For the Handwritten and MSRC-v5 datasets, Multi-view Spectral Clustering yields the highest scores, benefiting from compact clusters and strong local neighborhood consistency. Nevertheless, Mean-GWMDS-C remains competitive and consistently outperforms multi-view K-Means.

Overall, these results highlight that GW-based relational aggregation is particularly advantageous when views share compatible geometric structures, positioning Mean-GWMDS-C as a robust alternative for multi-view clustering.

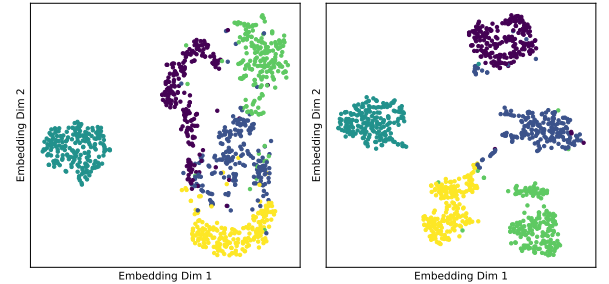
Fig. 3 provides a qualitative illustration of the clustering behavior of Mean-GWMDS-C on the Multiple Features dataset. This visualization highlights the ability of Mean-GWMDS-C to integrate multi-view relational information into a compact representation while preserving meaningful geometric separation between clusters.

Table III and Fig. 4 show the impact of the embedding dimensionality on Mean-GWMDS-C. The method achieves stable performance for low-dimensional embeddings on Multi-Feature and Caltech101-7, while moderate increases in dimension benefit the Handwritten and MSRC datasets. Overall, the results indicate that Mean-GWMDS-C is robust to the choice of embedding dimension and performs well with compact latent representations.

Fig. 5 illustrates the sensitivity of Mean-GWMDS-C to the number of neighbors used in the construction of relational graphs. Across all datasets, clustering performance remains relatively stable within a moderate range of neighborhood sizes, indicating robustness to this hyperparameter. On datasets



(a) Two-dimensional t-SNE embeddings of two feature views from the Multiple Features dataset.



(b) Visualizations of two views from the Multiple Features dataset, colored by Mean-GWMDS-C clusters.

FIG. 3: Visualization of the Multiple Features dataset, showing two feature views and the cluster structure recovered by Mean-GWMDS-C.

dim Y	NMI / ARI							
	MULFE		CALT		HAN		MSRC	
	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
2	0.9118	0.9256	0.2642	0.1673	0.7654	0.7221	0.4638	0.3411
4	0.9550	0.9679	0.2644	0.1675	0.8610	0.8519	0.5182	0.4090
8	0.9550	0.9679	0.2890	0.1831	0.8764	0.8720	0.5377	0.4330
16	0.9550	0.9679	0.2908	0.1841	0.8798	0.8740	0.5377	0.4330
32	0.9550	0.9679	0.2902	0.1838	0.8875	0.8841	0.4596	0.3117

TABLE III: Clustering performance of Mean-GWMDS-C as a function of the embedding dimensionality.

such as MulFe and Handwritten, NMI and ARI exhibit a clear peak for intermediate neighborhood values, while overly small or large neighborhoods tend to degrade performance. Overall, these results suggest that Mean-GWMDS-C benefits from a balanced neighborhood selection, avoiding both overly local and excessively dense relational structures.

Implementation details. All experiments use uniform masses, the squared loss, and 12-nearest-neighbor Isomap, without entropic regularization. GW subproblems are solved with POT [25] using tolerance 10^{-5} . For Bary-GWMDS, each view-wise distance matrix is normalized by its maximum. The barycenter is initialized from the first view and computed for 50 iterations with tolerance 10^{-12} , followed by 100 Adam iterations with learning rate 0.1. For Mean-GWMDS-C, distances are normalized by their mean off-diagonal value and prototype coordinates are optimized with Adam for 3000 iterations with learning rate 0.02.

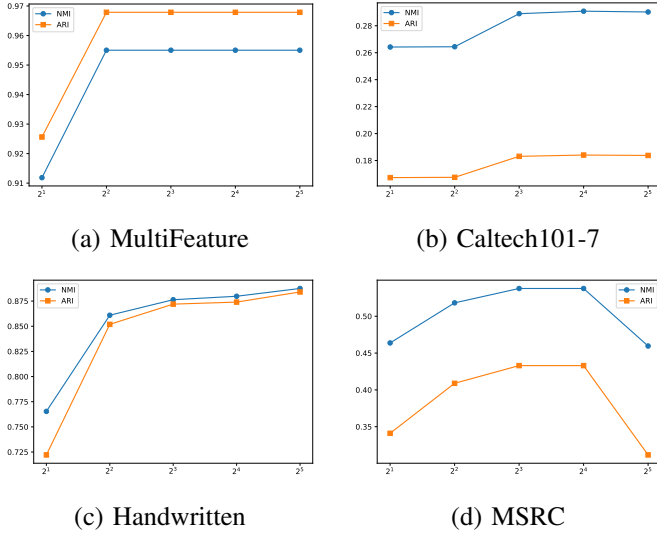


FIG. 4: Clustering performance of Mean-GWMDs-C as a function of the embedding dimensionality on different multi-view datasets, evaluated using NMI and ARI.

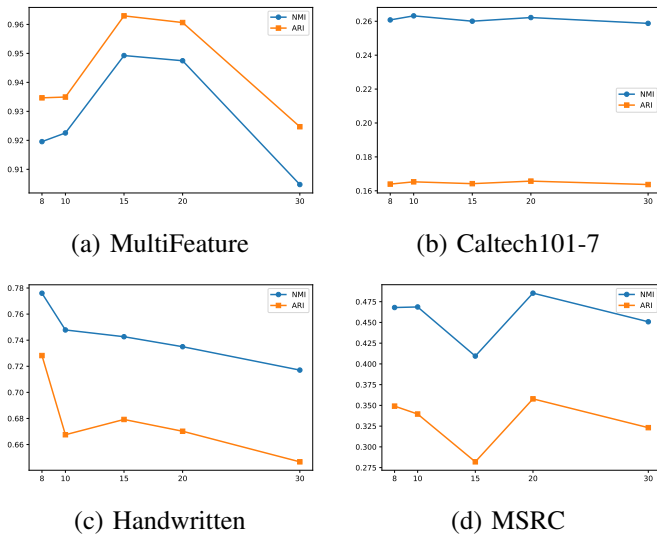


FIG. 5: Comparison of NMI and ARI as a function of the number of neighbors on different multi-view datasets.

V. CONCLUSION

The results show that GW-based relational modeling allows Bary-GWMDs to preserve intrinsic geometry often degraded by direct averaging, while Mean-GWMDs-C learns compact, relationally consistent representations. Overall, geometry-aware optimal transport provides an effective framework for multi-view embedding and clustering.

REFERENCES

- [1] Z. Yu, Z. Dong, C. Yu, K. Yang, Z. Fan, and C. P. Chen, "A review on multi-view learning," *Frontiers of Computer Science*, vol. 19, no. 7, p. 197334, 2025.
- [2] Y. Qin, X. Zhang, S. Yu, and G. Feng, "A survey on representation learning for multi-view data," *Neural Networks*, vol. 181, p. 106842, 2025.

- [3] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
- [4] A. R. Chowdhury, A. Gupta, and S. Das, "Deep multi-view clustering: A comprehensive survey of the contemporary techniques," *Information Fusion*, p. 103012, 2025.
- [5] M. Salter-Townshend and T. H. McCormick, "Latent space models for multiview network data," *The annals of applied statistics*, vol. 11, no. 3, p. 1217, 2017.
- [6] R. Lin, S. Du, S. Wang, and W. Guo, "Multi-view clustering via optimal transport algorithm," *Knowledge-Based Systems*, vol. 279, p. 110954, 2023.
- [7] T. Kerdoncuff, R. Emonet, and M. Sebban, "Sampled Gromov-Wasserstein," *Machine Learning*, vol. 110, no. 8, pp. 2151–2186, 2021.
- [8] E. F. Montesuma, F. M. N. Mboula, and A. Souloumiac, "Recent advances in optimal transport for machine learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [9] R. P. Eufrazio, E. F. Montesuma, and C. C. Cavalcante, "A dimensionality reduction technique based on the Gromov-Wasserstein distance," in *International Conference on Geometric Science of Information*. Springer, 2025, pp. 111–120.
- [10] R. P. Eufrazio, E. F. Montesuma, and C. C. Cavalcante, "Nonlinear dimensionality reduction through optimal transport between incomparable spaces: Rp eufrazio et al." *Information Geometry*, pp. 1–36, 2026.
- [11] R. P. Eufrazio, E. F. Montesuma, and C. C. Cavalcante, "Structure-preserving multi-view embedding using gromov-wasserstein optimal transport," 2026. [Online]. Available: <https://arxiv.org/abs/2604.02610>
- [12] R. A. Clark, T. Needham, and T. Weighill, "Generalized dimension reduction using semi-relaxed Gromov-Wasserstein distance," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 15, 2025, pp. 16 082–16 090.
- [13] H. Van Assel, Cédric Vincent-Cuaz, Nicolas Courty, Rémi Flamary, Pascal Frossard, Titouan Vayer, "Distributional Reduction: Unifying Dimensionality Reduction and Clustering with Gromov-Wasserstein," *Transactions on Machine Learning Research*, 2025. [Online]. Available: <https://openreview.net/forum?id=cllm6SS354>
- [14] G. Peyré, M. Cuturi, and J. Solomon, "Gromov-wasserstein averaging of kernel and distance matrices," in *International conference on machine learning*. PMLR, 2016, pp. 2664–2672.
- [15] T. Rodosthenous, V. Shahrezaei, and M. Evangelou, "Multi-view data visualisation via manifold learning," *PeerJ Computer Science*, vol. 10, p. e1993, 2024.
- [16] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.
- [17] M. van Breukelen, R. P. Duin, D. M. Tax, and J. Den Hartog, "Handwritten digit recognition by combined classifiers," *Kybernetika*, vol. 34, no. 4, pp. 381–386, 1998.
- [18] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International journal of computer vision*, vol. 81, no. 1, pp. 2–23, 2009.
- [19] F. Y. Chin, K. H. Lem, and K. M. Wong, "Improving handwritten digit recognition using hybrid feature selection algorithm," *Applied Computing and Informatics*, 2022.
- [20] M.-S. Chen, C.-D. Wang, and J.-H. Lai, "Low-rank tensor based proximity learning for multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 5076–5090, 2022.
- [21] S. Bickel and T. Scheffer, "Multi-view clustering," in *Icdm*, vol. 4, no. 2004, 2004, pp. 19–26.
- [22] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 393–400.
- [23] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of machine learning research*, vol. 3, no. Dec, pp. 583–617, 2002.
- [24] C. X. Gao, D. Dwyer, Y. Zhu, C. L. Smith, L. Du, K. M. Filia, J. Bayer, J. M. Mensink, T. Wang, C. Bergmeir et al., "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Research*, vol. 327, p. 115265, 2023.
- [25] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier et al., "POT: Python Optimal Transport," *Journal of Machine Learning Research*, vol. 22, no. 78, pp. 1–8, 2021.