RECOVERING CONSENSUS IN MULTI-MODAL GRAPHS: PROVABLE MULTI-MODAL SPECTRAL CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper studies clustering algorithms for multi-modal graphs, in which each modality might have a different structure of clusters caused by the unique information of each modality, and the modality-consensus clusters are defined such that vertices belong to the same cluster if and only if they appear in the optimal cluster of every single modality. The paper presents an algorithm to approximate these modality-consensus clusters, and provides a theoretical analysis of the algorithm's output with respect to the optimal clusters. The paper compares the performance of the proposed algorithm with the previous state-of-the-art on both synthetic and real-world datasets, with experimental results demonstrating its superiority.

1 INTRODUCTION

Graph clustering is a fundamental topic in unsupervised learning with wide-ranging applications across machine learning and data science. Traditionally, graph clustering is mainly studied from a unimodal perspective, and this approach is inherently limited. Specifically, they rely solely on modality-specific information, often resulting in a coarse or incomplete partition of the data. By contrast, integrating data from multiple modalities can reveal complementary insights and expose more precise structural relationships. This potential has made multi-modal clustering a very active research area over the past years. On the other side, existing multi-modal clustering methods implicitly assume that different modalities share a common optimal clustering structure. This assumption can be overly restrictive for real-world multi-modal data, where each modality may reveal a distinct yet complementary clustering structure due to its modality-specific information. For example, in multi-modal medical data, medical images may group patients according to tumor morphology, whereas genomic profiles may partition them according to molecular subtypes. Rather than treating such structural discrepancies as conflicts, they can provide complementary cues for uncovering more fine-grained joint structures. Specifically, as illustrated in Fig. 1 (b), the non-empty intersections of modality-specific optimal clusters naturally induce a finer partition that simultaneously preserves the clustering information revealed by different modalities. This observation raises a very interesting question: how can we recover such a joint clustering structure from multi-modal graphs when their modality-specific optimal partitions may differ?

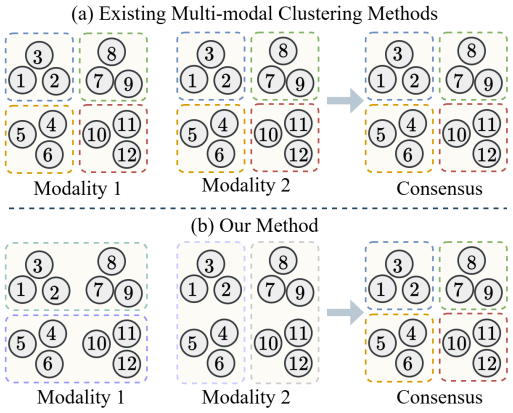


Figure 1: (a) Existing multi-modal clustering methods assume different modalities share the same optimal clusters, failing to deal with modalities with different optimal cluster structures. (b) Our method defines the modality consensus optimal clusters as the non-empty intersection of modality-specific optimal clusters.

Motivated by this question, in this work we introduce a novel algorithm for multi-modal graph clustering and provide a rigorous theoretical analysis of its performance; the overall framework of our algorithm is shown in Figure 2. We define the optimal clustering as the consensus across all modalities: two vertices belong to the same cluster if and only if they are co-clustered in the

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.

optimal partition of every individual modality. Our method approximates this consensus by jointly leveraging modality-specific embeddings, and we establish its performance guarantee under a natural assumption about the structure of the input multi-modal graphs. To the best of our knowledge, this is the first multi-modal graph clustering algorithm with provable theoretical guarantees of this type. We conduct experiments on both synthetic and real-world datasets, and the experimental results confirm the effectiveness of our designed algorithm.

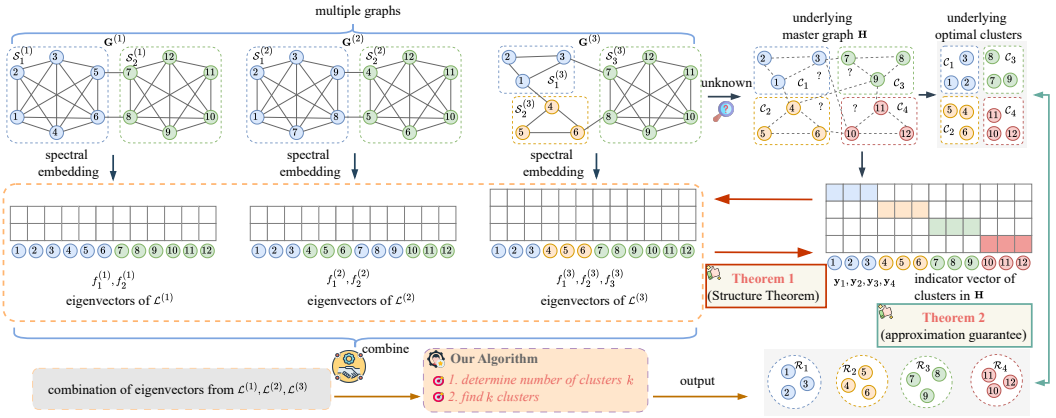


Figure 2: Overview of our work

2 Related Work

Our work relates to a large and growing number of multi-modal clustering methods. These methods are mainly based on different learning strategies, e.g., diffusion maps (Lindenbaum et al., 2020), modality disentanglement (Xu et al., 2021), collaborative training (Trosten et al., 2023), and adaptive multi-modal fusion (Liu et al., 2025b). While several studies jointly utilize multi-modal spectral embeddings to capture the underlying cluster structures Kumar & Daumé (2011), they primarily rely on heuristic designs that lack theoretical analysis. In addition, previous algorithms are mainly built upon the assumption of cluster structure consistency across different modalities, whereas we address clustering for multi-modal graphs in which each modality might have a different structure of clusters. While recent work (Tsitsikas & Papalexakis, 2025) allows different graphs to exhibit distinct clustering structures, its goal is to jointly cluster graphs and samples into multiple structural groups instead of recovering a unified fine-grained partition induced by the modality-specific optimal clusters. Moreover, it remains a heuristic approach without a theoretical foundation. On the other side, theoretical studies of multi-modal clustering mainly focuses on the block models (Cohen-Addad et al., 2024) or analyzing the generalization errors on multi-modal clustering (Liang et al., 2025; 2023), from which our work is fundamentally different.

Our work is also related to a sequence of research on the theoretical analysis of graph clustering on well-clustered graphs (Cohen-Addad et al., 2017; Macgregor & Sun, 2022; Mizutani, 2021; Peng et al., 2017; Sun & Zanetti, 2019). The primary focus of these works is unimodal graphs and, as one of the conceptual contribution, we show that the tools developed in this direction can be generalized for multi-modal graph clustering.

3 Preliminaries

Let $\mathbf{G} = (\mathcal{V}, \mathbf{E}, w)$ be an undirected graph with n vertices and weight function $w : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_{\geq 0}$. For any edge $e = \{u, v\} \in \mathbf{E}$, we use $w_{\mathbf{G}}(u, v)$ or $w_{\mathbf{G}}(e)$ to express the weight of e. For a vertex $u \in V$, we denote its degree by $\deg_{\mathbf{G}}(u) \triangleq \sum_{v \in \mathcal{V}} w_{\mathbf{G}}(u, v)$, and the volume for any $\mathcal{S} \subseteq \mathcal{V}$ is defined as $\mathrm{vol}_{\mathbf{G}}(\mathcal{S}) \triangleq \sum_{u \in \mathcal{S}} \deg_{\mathbf{G}}(u)$. For any $\mathcal{S}, \mathcal{T} \subset \mathcal{V}$, we define the cut value between $\mathcal{S}$ and $\mathcal{T}$ by $w_{\mathbf{G}}(\mathcal{S}, \mathcal{T}) \triangleq \sum_{e \in \mathbf{E}(\mathcal{S}, \mathcal{T})} w_{\mathbf{G}}(e)$, where $\mathbf{E}(\mathcal{S}, \mathcal{T})$ is the set of edges between $\mathcal{S}$ and $\mathcal{T}$. Moreover,

for any $\mathcal{S} \subset \mathcal{V}$, the conductance of $\mathcal{V}$ is defined as

$$\Phi_{\mathbf{G}}(\mathcal{S}) \triangleq \frac{w_{\mathbf{G}}(\mathcal{S}, \mathcal{V} \setminus \mathcal{S})}{\min\{\mathrm{vol}_{\mathbf{G}}(\mathcal{S}), \mathrm{vol}_{\mathbf{G}}(\mathcal{V} \setminus \mathcal{S})\}}$$

if $\mathcal{S} \neq \emptyset$, and $\Phi_{\mathbf{G}}(\mathcal{S}) = 1$ if $\mathcal{S} = \emptyset$. For any $k \geq 2$, we call subsets of vertices $\mathcal{A}_1, \ldots, \mathcal{A}_k$ a k-way partition of $\mathbf{G}$ if $\bigcup_{i=1}^{k} \mathcal{A}_i = \mathcal{V}$ and $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for different i and j. We define the k-way expansion of $\mathbf{G}$ by $\rho_{\mathbf{G}}(k) \triangleq \min_{\text{partition } \mathcal{A}_1, \ldots, \mathcal{A}_k} \max_{i \in [k]} \Phi_{\mathbf{G}}(\mathcal{A}_i)$, where for any $k \in \mathbb{Z}^+$ we define $[k] \triangleq \{1, \ldots, k\}$.

We use ℓ to express the total number of modalities considered in the problem, and use $\mathbf{G}^{(i)} = \left(\mathcal{V}, \mathbf{E}^{(i)}, w^{(i)}\right)$ to define the graph corresponding to the i-th modality for $i \in [\ell]$. We assume that graph $\mathbf{G}^{(i)}$ has k_i clusters, and the optimal clusters of $\mathbf{G}^{(i)}$ are $\mathcal{S}_1^{(i)}, \ldots, \mathcal{S}_{k_i}^{(i)}$. We define the normalized indicator vector of $\mathcal{S}_j^{(i)}$ for $j \in [k_i]$ as $\mathbf{g}_j^{(i)} \in \mathbb{R}^n$, where $\mathbf{g}_j^{(i)}(u) = \sqrt{1/|\mathcal{S}_j^{(i)}|}$ if $u \in \mathcal{S}^{(i)}$ and $\mathbf{g}_j^{(i)}(u) = 0$ otherwise. Note that $\left\|\mathbf{g}_j^{(i)}\right\|^2 = 1$ for any $i \in [\ell]$ and $j \in [k_i]$.

We assume there exists an underlying *master graph* $\mathbf{H}$ with the same vertex set $\mathcal{V}$, and $\mathbf{H}$ encodes the cluster structure of the graphs corresponding to different modalities. Notice that any two vertices u, v belong to the same cluster in $\mathbf{H}$ if and only if u, v belong to the same optimal cluster $\mathcal{S}_j^{(i)}$ for every $1 \leq i \leq \ell$. This property naturally results in a partition of $\mathcal{V}$, denoted by $\mathcal{C}_1, \ldots, \mathcal{C}_k$ for some $k \leq \prod_{i=1}^{\ell} k_i$. We call this partition the *optimal partition* of $\mathcal{V}$, and define the normalized indicator vector of $\mathcal{C}_j$ for $j \in [k]$ as $\mathbf{y}_j \in \mathbb{R}^n$, where $\mathbf{y}_j(u) = \sqrt{1/|\mathcal{C}_j|}$ if $u \in \mathcal{C}_j$ and $\mathbf{y}_j(u) = 0$ otherwise. Note that $\|\mathbf{y}_j\|^2 = 1$ for any $j \in [k]$.

Our analysis is based on the spectral properties of graphs, and we list the basics of spectral graph theory. For any $\mathbf{G}^{(i)} = \left(\mathcal{V}, \mathbf{E}^{(i)}, w^{(i)}\right)$, let $\mathbf{D}_{\mathbf{G}^{(i)}} \in \mathbb{R}^{n \times n}$ be the diagonal matrix defined by $\mathbf{D}_{\mathbf{G}^{(i)}}(u, u) = \deg_{\mathbf{G}^{(i)}}(u)$ for any $u \in V$. We denote by $\mathbf{A}_{\mathbf{G}^{(i)}} \in \mathbb{R}^{n \times n}$ the adjacency matrix of $\mathbf{G}^{(i)}$, where $\mathbf{A}_{\mathbf{G}^{(i)}}(u, v) = w^{(i)}_{\mathbf{G}^{(i)}}(u, v)$. The normalized Laplacian matrix of $\mathbf{G}^{(i)}$ is defined as $\mathcal{L}^{(i)} \triangleq \mathbf{I} - \mathbf{D}_{G^{(i)}}^{-1/2} \mathbf{A}_{G^{(i)}} \mathbf{D}_{G^{(i)}}^{-1/2}$, where $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix. The normalized Laplacian matrix $\mathcal{L}^{(i)}$ is symmetric and real-valued, and has n real eigenvalues which we write as $0 = \lambda_1^{(i)} \leq \ldots \leq \lambda_n^{(i)} \leq 2$; we use $\mathbf{f}_j^{(i)} \in \mathbb{R}^n$ for $i \in [\ell]$ and $j \in [n]$ to express the eigenvector of $\mathcal{L}^{(i)}$ corresponding to $\lambda_j^{(i)}$. The following lemma will be used in our analysis.

Lemma 3.1 (Higher-order Cheeger inequality, (Lee et al., 2014)). *There is an absolute constant C' such that it holds for any graph $\mathbf{G}$ and $k \geq 2$ that $\lambda_k(\mathcal{L}_{\mathbf{G}})/2 \leq \rho_{\mathbf{G}}(k) \leq C' \cdot k^3 \sqrt{\lambda_k(\mathcal{L}_{\mathbf{G}})}$, where $\lambda_k(\mathcal{L}_{\mathbf{G}})$ is the k-th smallest eigenvalue of $\mathcal{L}_{\mathbf{G}}$.*

4 Relating the Optimal Partition to the Eigenvectors of $\mathcal{L}^{(i)}$

In this section we show that the indicator vectors of the optimal clusters $\mathcal{C}_1, \ldots, \mathcal{C}_k$ of $\mathbf{H}$ can be approximately by the bottom eigenvectors of $\mathcal{L}^{(i)}$ for $i \in [\ell]$. Notice that, while computing an optimal partition of every $\mathbf{G}^{(i)}$ is NP-hard and it's intractable to explicitly construct $\mathbf{H}$ with $\mathcal{C}_1, \ldots, \mathcal{C}_k$ as the optimal clusters, we prove that the optimal cluster information of this *non-explicitly* constructed $\mathbf{H}$ can be approximately obtained from the spectral properties of $\mathcal{L}^{(i)}$. From this perspective, we find the result very interesting and could have more applications in studying multi-modal clustering.

4.1 The Structure Theorem for Classical Clustering

We first examine the case of any unimodal graph $\mathbf{G}^{(i)}$, and review the condition needed to ensure that there is a well-defined structure of clusters in $\mathbf{G}^{(i)}$. By definition, we say that $\mathbf{G}^{(i)}$ has k_i well-defined clusters if one can partition $\mathcal{V}$ into k_i subsets, and the induced subgraph from each subset is an expander graph; on the other side, every $(k_i + 1)$-partition of $\mathcal{V}$ would introduce a dense cut between the partitioned subsets. Formalizing the intuition, Peng et al. Peng et al. (2017) introduces

the parameter

$$\Upsilon_{\mathbf{G}^{(i)}}(k_i) = \frac{\lambda_{k+1}^{(i)}}{\rho_{\mathbf{G}^{(i)}}(k_i)}, \tag{1}$$

and shows any graph $\mathbf{G}^{(i)}$ with a large value of $\Upsilon_{\mathbf{G}^{(i)}}(k_i)$ for some k_i ensures that $\mathbf{G}^{(i)}$ has k_i well-defined clusters. [1]

There's a sequence of work quantifying the relationship between the spectral properties of $\mathcal{L}^{(i)}$ and the cluster structure of $\mathbf{G}^{(i)}$ with respect to $\Upsilon_{\mathbf{G}^{(i)}}(k_i)$ (Mizutani, 2021; Sun & Zanetti, 2019; Chen et al., 2016; Czumaj et al., 2015; Kloumann et al., 2017; Kolev & Mehlhorn, 2016; Louis & Venkat, 2019; Peng, 2020; Peng & Yoshida, 2020), and the following result will be used in our analysis.

Lemma 4.1 (The Structure Theorem, (Macgregor & Sun, 2022)). *The following statements hold: (1) For any $i \in [\ell]$ and $j \in [k_i]$, there is $\hat{\mathbf{f}}_j^{(i)} \in \mathbb{R}^n$, which is a linear combination of $\mathbf{f}_1^{(i)}, \ldots, \mathbf{f}_{k_i}^{(i)}$, such that $\left\|\mathbf{g}_j^{(i)} - \hat{\mathbf{f}}_j^{(i)}\right\|^2 \leq 1/\Upsilon_{\mathbf{G}^{(i)}}(k_i)$. (2) For any $i \in [\ell]$ and $j \in [k_i]$, there are vectors $\hat{\mathbf{g}}_1^{(i)}, \ldots, \hat{\mathbf{g}}_{k_i}^{(i)}$, each of which is a linear combination of $\mathbf{g}_1^{(i)}, \ldots, \mathbf{g}_{k_i}^{(i)}$, such that $\sum_{j=1}^{k_i} \left\|\mathbf{f}_j^{(i)} - \hat{\mathbf{g}}_j^{(i)}\right\|^2 \leq k_i/\Upsilon_{\mathbf{G}^{(i)}}(k_i)$.*

4.2 The Structure Theorem for Multi-Modal Clustering

Building on Lemma 4.1, we develop a quantitative connection between the spectral properties of $\mathcal{L}^{(i)}$ and the structure of optimal clusters in $\mathbf{H}$, and present the structure theorem for multi-modal graph clustering. Our result is as follows:

Theorem 1 (The Structure Theorem for Multi-modal Clustering). *The following statements hold:*

1. *Given any $\mathbf{y}_i$ for $i \in [k]$, there is a linear combination of $\mathbf{f}_1^{(1)}, \ldots, \mathbf{f}_{k_1}^{(1)}, \ldots, \mathbf{f}_1^{(\ell)}, \ldots, \mathbf{f}_{k_\ell}^{(\ell)}$, denoted by $\mathbf{h}_i \in \mathbb{R}^n$, such that $\|\mathbf{y}_i - \mathbf{h}_i\|^2 \leq \sum_{j=1}^{\ell} k_j/\Upsilon_{\mathbf{G}^{(j)}}(k_j)$.*

2. *For any $i \in [\ell]$, there are vectors $\hat{\mathbf{y}}_1^{(i)}, \ldots, \hat{\mathbf{y}}_{k_i}^{(i)}$, each of which is a linear combination of $\mathbf{y}_1, \ldots, \mathbf{y}_k$, such that $\sum_{j=1}^{k_i} \left\|\mathbf{f}_j^{(i)} - \hat{\mathbf{y}}_j^{(i)}\right\|^2 \leq k_i/\Upsilon_{\mathbf{G}^{(i)}}(k_i)$.*

Theorem 1 shows that every $\mathbf{y}_i$ can be approximated by a linear combination of the bottom k_j eigenvectors *from ℓ different modalities*. Thus, we can approximately recover the optimal modality-consensus clusters from the space jointly spanned by the modality-specific embeddings. Moreover, the upper bound $\ell \sum_{j=1}^{\ell} k_j/\Upsilon_{\mathbf{G}^{(j)}}$ of the first part implies that the clearer structure of clusters in any unimodal $\mathbf{G}^{(i)}$, the clearer one in $\mathbf{H}$. To the best of our knowledge, this presents the *first* theoretical result that quantifies the relationship between $\mathbf{G}^{(i)}$ and $\mathbf{H}$.

Proof Sketch of Theorem 1. We first employ a sequence of orthogonal projection matrices $\{\mathbf{P}^{(i)}\}_{i=1}^{\ell}$ to decompose the indicator vector $\mathbf{y}_i$ onto the subspace spanned by the indicators $\mathbf{g}_j^{(i)}$ for $i \in [\ell]$ and $j \in [k_i]$. This allows us to represent every $\mathbf{y}_i$ as a linear combination of the indicator vectors. Next, we construct $\mathbf{h}_i$ by examining the subspace defined by the bottom eigenvectors of the respective graph Laplacian matrices. We further decompose the approximation error $\|\mathbf{y}_i - \mathbf{h}_i\|^2$ into modality-specific tail errors. By the Cauchy–Schwarz inequality, Parseval's identity, and Lemma 4.1, we bound this tail and combine the modality-wise bounds to prove Theorem 1

[1] To explain the meaning of $\Upsilon_{\mathbf{G}^{(i)}}(k_i)$, we assume that $\mathbf{G}^{(i)}$ has k_i well-defined clusters $\left\{\mathcal{S}_j^{(i)}\right\}_{j=1}^{k_i}$. By definition, the values of $\Phi_{\mathbf{G}^{(i)}}(\mathcal{S}_j^{(i)})$ for every $\mathcal{S}_j^{(i)}$, as well as $\rho_{\mathbf{G}_i}(k_i)$, are low; on the other hand, any (k_i+1)-way partition of $\mathcal{V}$ would separate the vertices of some $\mathcal{S}_j^{(i)}$, and this leads to a high value of $\rho_{\mathbf{G}^{(i)}}(k_i+1)$. Combining this with Lemma 3.1, a lower bound on $\Upsilon_{\mathbf{G}^{(i)}}(k_i)$ would be sufficient to ensure that $\mathbf{G}^{(i)}$ has exactly k_i clusters.

5 MULTI-MODAL GRAPH CLUSTERING ALGORITHM

In this section we present and analyze a multi-modal graph clustering algorithm. We assume without loss of generality that there is a universal constant $C \in \mathbb{R}^+$ such that every modality $\mathbf{G}^{(i)}$ has a clear structure of clusters formalized by $\Upsilon_{\mathbf{G}^{(i)}}(k_i) \geq C$ for some $k_i \in [n]$, which results in well-defined optimal clusters $\mathcal{C}_1, \ldots, \mathcal{C}_k$. This assumption is well supported by the classical eigen-gap heuristic Ng et al. (2001), and is known the weakest assumption for spectral clustering to work Macgregor & Sun (2022). The performance of our algorithm is as follows:

Theorem 2 (Main Theorem). *Let $\{\mathbf{G}^{(i)}\}_{i=1}^{\ell}$ be a set of graphs, and each graph has k_i clusters of almost balanced size and $\Upsilon_{\mathbf{G}^{(i)}}(k_i) \geq C$ for a universal constant[2]. Let $\{\mathcal{R}_s\}_{s=1}^{k}$ be the output of our algorithm and $\{\mathcal{C}_s\}_{s=1}^{k}$ the optimal clusters of $\mathbf{H}$, and satisfying $|\mathcal{C}_s| = \Omega(n/k)$. Without loss of generality, let the optimal correspondent of $\mathcal{R}_s$ be $\mathcal{C}_s$. Then, it holds that*

$$\sum_{s=1}^{k} |\mathcal{R}_s \triangle \mathcal{C}_s| = O\left(\frac{k}{\max\{k_j\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \cdot n.$$

The clustering error of the multi-modal algorithm is governed by a weighted average of the modality-wise cluster structure parameters, with weights determined by the number of clusters in each modality. In the special case where all modalities admit the same optimal cluster structure, the bound reduces to $O\left(1/\ell \cdot \sum_{i=1}^{\ell} 1/\Upsilon_{\mathbf{G}^{(i)}}(k_i)\right) \cdot n$, recovering an average dependence on the clusterability of the individual modalities[3].

To examine the significance of this result, recall that clusters of n vertices occurring in practice usually have high conductance, and different clusters are connected by sparse cuts. This results in $\Upsilon_{\mathbf{G}^{(i)}}(k_i) \geq \log^c n$ for some constant c if $\mathbf{G}^{(i)}$ has k_i clusters connected by sparse cuts. In this scenario, by Theorem 2 most vertices are correctly classified by the algorithm, and the total misclassified vertices by our algorithm is $o(n)$.

5.1 MAIN ALGORITHM

Throughout the subsection we assume that the optimal number of clusters k in $\mathbf{H}$ is known in advance, and our main algorithm consists of three steps: the algorithm first computes the bottom k_i eigenvectors of every $\mathcal{L}^{(i)}$ for $i \in [\ell]$, and uses these eigenvectors to construct an embedding. Secondly, the algorithm applies k-means on the embedded points and uses its output as the partition of $\mathcal{V}$. See Algorithm 1 for formal description.

Algorithm 1 Multi-modal Graph Clustering

Require: Multi-modal Laplacian matrices $\{\mathcal{L}^{(i)}\}_{i=1}^{\ell}$, number of modality-specific clusters $\{k_i\}_{i=1}^{\ell}$, and total number of cluster k
Ensure: Clusters $\{\mathcal{A}_1, \ldots, \mathcal{A}_k\}$
1: **for** $i = 1$ **to** ℓ **do**
2: Compute the eigenvectors $\mathbf{f}_1^{(i)}, \ldots, \mathbf{f}_{k_i}^{(i)}$ of $\mathcal{L}^{(i)}$
3: Construct the embedding: $F^{(i)}(u) \triangleq \left(\mathbf{f}_1^{(i)}(u), \ldots, \mathbf{f}_{k_i}^{(i)}(u)\right)$
4: **end for**
5: Embed each $u \in \mathcal{V}$ to $F(u) \triangleq (F^{(1)}(u), \ldots, F^{(\ell)}(u))^\top$
6: Apply k-means on the embedded points $\{F(u)\}_{u \in \mathcal{V}}$
7: Partition $\mathcal{V}$ into k clusters
8: **return** $\{\mathcal{A}_1, \ldots, \mathcal{A}_k\}$

[2] It's known from Macgregor & Sun (2022) that one can remove the condition of almost-balanced clusters by assuming a stronger condition on $\Upsilon_{\mathbf{G}^{(i)}}$, i.e., $\Upsilon_{\mathbf{G}^{(i)}} = \Omega(k_i)$. The same tradeoff applies to our analysis, implying that our result holds for graphs with unbalanced clusters when we slightly increase the lower bound on $\Upsilon_{\mathbf{G}^{(i)}}$.

[3] The modality weights can be further refined by incorporating the eigen gap of each modality to reduce the contribution of modalities with worse quality.

Remark 5.1. In Algorithm 1 the number of clusters k_i in $\mathbf{G}^{(i)}$ is given as input. Notice that, by the eigen-gap heuristic Ng et al. (2001), $\mathbf{G}^{(i)}$ has k_i clusters if and only if there is a large gap between the k-th and $(k+1)$-th eigenvalue of $\mathcal{L}^{(i)}$. Hence, we can determine the value of k_i by computing the eigenvalue of $\mathcal{L}^{(i)}$.

Now we analyze Algorithm 1 and sketch the proof of Theorem 2. By definition, we can write every optimal cluster $\mathcal{C}_s$ for $s \in [k]$ as $\mathcal{C}_s = \bigcap_{i=1}^{\ell} \mathcal{S}^{(i)}_{\theta^{(i)}(s)}$, where $\theta^{(i)}(s)$ is the only index satisfying $\mathcal{C}_s \subseteq \mathcal{S}^{(i)}_{\theta^{(i)}(s)}$. For every cluster $\mathcal{C}_s$, we define $\mathbf{p}_s = \left(\mathbf{p}_s^{(1)}, \mathbf{p}_s^{(2)}, \ldots, \mathbf{p}_s^{(\ell)}\right)$, where $\mathbf{p}_s^{(i)}(j) = 1/\sqrt{|\mathcal{C}_s|}\left\langle \mathbf{f}_j^{(i)}, \mathbf{g}^{(i)}_{\theta^{(i)}(s)} \right\rangle \left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s \right\rangle$. We refer $\{\mathbf{p}_s\}_{s=1}^k$ as the approximate center of the point embedding from the optimal clusters $\{\mathcal{C}_s\}_{s=1}^k$ and prove that the resultant total k-means cost of the embedding points can be upper-bounded.

Lemma 5.2. *It holds that* $\sum_{s=1}^{k} \sum_{u \in \mathcal{C}_s} \|F(u) - \mathbf{p}_s\|^2 \leq \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}$.

Lemma 5.2 shows that the cost of an optimal k-means on $\{F(u)\}_{u \in \mathcal{V}}$ is at most $\sum_{i=1}^{\ell} k_i / \Upsilon_{\mathbf{G}^{(i)}}(k_i)$. We further prove that any different $\mathbf{p}^{(s)}$ and $\mathbf{p}^{(z)}$ are far from each other, and their distance can be lower bounded as well.

Lemma 5.3. *It holds for any different* $s, z \in [k]$ *that*

$$\|\mathbf{p}_s - \mathbf{p}_z\|^2 \geq \sum_{i=1}^{\ell} \frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|, \left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

Lemmas 5.2 and 5.3 allow us to construct a bijection σ between our output clusters and the optimal ones, such that under σ the total number of misclassified vertices can be upper bounded. Our analysis on constructing σ and bounding the total number of misclassified vertices is similar to the one in (Macgregor & Sun, 2022), and we refer the reader to the appendix for complete analysis.

5.2 Computing the value of k

This subsection addresses the assumption that k is known in advance, and presents an efficient algorithm to compute the optimal number of clusters. Our analysis is based on the case distinction:

First of all, we prove that partitioning $\mathcal{V}$ into $k' \geq k$ clusters has low k-means cost. By the definition of the optimal clusters in $\mathbf{H}$, we know that k is at most $k^* \triangleq \min\left\{\prod_{i=1}^{\ell} k_i, n\right\}$. As the total k-means cost for any fixed input set decreases with respect to k, by Lemma 5.2 we have the following result:

Lemma 5.4. *Let* $\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'})$ *be the cost returned by a* k'*-means algorithm with approximation ratio* APT*, where* $k' \in [k, k^*]$*. Then, it holds that*

$$\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'}) \leq \mathsf{APT} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Secondly, we prove that partitioning $\mathcal{V}$ into $k-1$ clusters has high cost. Our analysis is based on the following intuition. Notice that it holds by Lemma 5.2 that most embedded points $F(u)$ with $u \in \mathcal{C}_s$ is close to their approximate center $\mathbf{p}_s$, and it holds by Lemma 5.3 that any different points $\mathbf{p}_s$ and $\mathbf{p}_z$ are far from each other. Hence, when partitioning these n points into $k-1$ clusters, a constant fraction of the points from the "disappeared" cluster will be forced to merge to the clusters with the approximate center far from its original one, which results in high $(k-1)$-means cost.

Lemma 5.5. *Let* $\mathsf{COST}^*(\mathcal{Q}_1, \ldots, \mathcal{Q}_{k-1})$ *be the optimal* $(k-1)$*-means cost on the instance* $\{F(u)\}_{u \in \mathcal{V}}$ *returned by a* $(k-1)$*-means algorithm. Then it holds that*

$$\mathsf{COST}^*(\mathcal{Q}_1, \ldots, \mathcal{Q}_{k-1}) \geq \sum_{i=1}^{\ell} \left(\frac{k_i^2}{16} - \frac{2k_i^2}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

Combining Lemmas 5.4 and 5.5 as well as the assumption on $\Upsilon_{\mathbf{G}^{(i)}}$ for $i \in [\ell]$, we have for any $k' \geq k$ that $\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'}) \leq \frac{1}{20} \cdot \sum_{i=1}^{\ell} k_i$, and $\mathsf{COST}^*(\mathcal{Q}_1, \ldots, \mathcal{Q}_{k-1}) \geq \frac{1}{20} \cdot \sum_{i=1}^{\ell} k_i^2$. Inspired by this, we can repeatedly run Algorithm 1 with all possible values of k, starting with the highest possible one k^*, until we reach $k-1$ for which the total cost is at least $\left(\sum_{i=1}^{\ell} k_i^2\right)/20$. Then, we can use the output from the second last iteration corresponding to k as the algorithm's final output.

5.3 Practical Consideration

Section 5.2 shows that the exact value of k can be computed directly by repeatedly executing Algorithm 1 and keeping track of the k-means cost function. This algorithm guarantees to find the exact value, but has high time complexity. To address this, we present a heuristic algorithm to approximate the value of k, and our heuristic algorithm is as follows:

Initialization: firstly, we run Algorithm 1 by setting the number of cluster $k' = \min\left\{\prod_{i=1}^{\ell} k_i, n\right\}$, and obtain clusters $\mathcal{R}_1, \ldots, \mathcal{R}_{k'}$. Since $k' \geq k$, the same optimal cluster could be wrongly decomposed in this initialization step. Secondly, for every $\mathbf{G}^{(i)}$ we apply spectral clustering to compute its approximate clusters denoted by $\mathcal{A}_1^{(i)}, \ldots, \mathcal{A}_{k_i}^{(i)}$. We introduce the *signature vector* $\mathsf{sig} : \mathcal{V} \to [k_1] \times \ldots \times [k_\ell]$, such that the i-th entry of $\mathsf{sig}(u)$ is j if $u \in \mathcal{A}_j^{(i)}$, i.e., when running spectral clustering on $\mathbf{G}^{(i)}$, vertex u is assigned to the j-th cluster. Since every $\mathcal{A}_j^{(i)}$ is close to its corresponding optimal cluster by the performance guarantee of spectral clustering on every individual $\mathbf{G}^{(i)}$, we expect most vertices in $\mathcal{R}_j$ for $j \in [k']$ have the same signature.

Merging Clusters: the merging step consists of iterations. In every iteration, we sample a fixed number of vertices $u_1, \ldots, u_t$ from every $\mathcal{R}_j$, and use the signature most frequently appeared in $\mathsf{sig}(u_1), \ldots, \mathsf{sig}(u_t)$ as the signature of $\mathcal{R}_j$. We merge the clusters with the same signature and proceed with the next iteration, until the algorithm doesn't further merge clusters in an iteration.

By the description, our designed heuristic applies Algorithm 1 with a larger value of k to generate a coarse partition, and generates signatures for vertices and clusters. By the performance guarantee of spectral clustering, two clusters share the same signature if they actually belong to the same optimal one, and will be merged by the algorithm. Due to the fixed number of vertices sampled in each iteration, our designed heuristic runs fast in practice.

6 Experiments

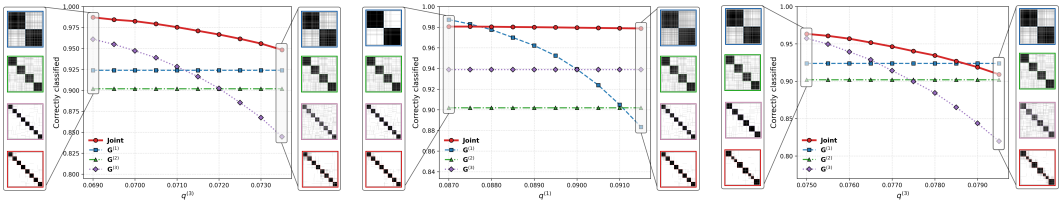
We evaluate the clustering performance of the proposed algorithm on both synthetic and real-world datasets. Our algorithm is implemented in Python 3.8, and all experiments are conducted on a computer server with an AMD 9700x CPU@3.8 GHz processor, 64G RAM, and RTX 5090 GPU.

6.1 Results on Synthetic Data

We first study the clustering performance of our algorithm on graphs $\{\mathbf{G}^{(i)}\}_{i=1}^{\ell}$ from the stochastic block model (SBM) (Abbe, 2018) parameterised by $N, \ell, k_1, \ldots, k_\ell, p$, and $q_1, \ldots, q_\ell$. Here, every graph $\mathbf{G}^{(i)}$ has N vertices labeled 1 to N, and k_i clusters. The vertices of $\mathbf{G}^{(i)}$ are partitioned sequentially into k_i blocks of size $n^{(i)} = N/k_i$, such that the j-th cluster contains vertices labeled from $(j-1) \cdot n^{(i)} + 1$ to $j \cdot n^{(i)}$. To generate graph $\mathbf{G}^{(i)}$ with clusters $\left\{\mathcal{S}_j^{(i)}\right\}_{j=1}^{k_i}$, for every pair of vertices $u \in \mathcal{S}_x^{(i)}$ and $v \in \mathcal{S}_y^{(i)}$ we add an edge $\{u, v\}$ with probability p if $x = y$, and add $\{u, v\}$ with probability $q^{(i)}$ if $x \neq y$.

The metric used for evaluating the performance of spectral clustering on every single graph $\mathbf{G}^{(i)}$ is defined by $(\sum_{j=1}^{k_i} |\mathcal{A}_j^{(i)} \cap \mathcal{S}_j^{(i)}|/|\mathcal{S}_j^{(i)}|)/k_i$. We evaluate the clustering performance of our algorithm by $(\sum_{j=1}^{k} |\mathcal{R}_j \cap \mathcal{C}_j|/|\mathcal{C}_j|)/k$. Notice that the cluster indexes for optimal clusters are transferred by permutation, such that the output clusters and the optimal clusters with the same cluster indexes are "optimal" correspondent with each other. To this end, these two metrics are consistent with the

approximation guarantee for classical graph clustering and multi-modal graph clustering, facilitating the measurement of the average corrected classified vertices ratio with respect to each cluster.



(a) $p = 0.1$, $q^{(1)} = 0.0905$, $q^{(2)} = 0.0835$ (b) $p = 0.1$, $q^{(2)} = 0.0835$, $q^{(3)} = 0.0705$ (c) $p = 0.1$, $q^{(1)} = 0.0905$, $q^{(2)} = 0.0835$

Figure 3: Comparison on the performance of our algorithm and classical spectral clustering. The legends called $\mathbf{G}^{(1)}$, $\mathbf{G}^{(2)}$, and $\mathbf{G}^{(3)}$ correspond to the output of spectral clustering on each graph, and "Joint" refers to the performance of our algorithm. Clearer block diagonal structures correspond to better clustering performance. In addition, we plot the sample correlation matrices, where an entry is non-zero if two vertices are in the same cluster depicted by the output of the clustering algorithm. This visualization clearly indicates that our algorithm effectively captures the underlying cluster structure of $\mathbf{H}$.

Case 1. We look at the case where the underlying total number of clusters is equal to the number of clusters in one of the $\mathbf{G}^{(i)}$s. We construct three graphs by setting the parameters as follows:

- Graph $\mathbf{G}^{(1)}$ has $k_1 = 2$ clusters, each of which has $n^{(1)} = 6,000$ vertices. Vertices $1 \sim 6,000$ belong to cluster 1, and vertices $6,001 \sim 12,000$ belong to cluster 2.
- Graph $\mathbf{G}^{(2)}$ has $k_2 = 4$ clusters, each of which has $n^{(2)} = 3,000$ vertices. The clusters are formed by sequential blocks of vertices: $1 \sim 3,000$ (cluster 1), $3,001 \sim 6,000$ (cluster 2), $6,001 \sim 9,000$ (cluster 3), and $9,001 \sim 12,000$ (cluster 4).
- Graph $\mathbf{G}^{(3)}$ has $k_3 = 8$ clusters, each of which has $n^{(3)} = 1,500$ vertices. The vertices are equally divided.

By construction, there are 8 clusters in $\mathbf{H}$, and these clusters are consistent to the ones in $G^{(3)}$. To evaluate the performance of our algorithm, we set $p = 0.1$, $q^{(1)} = 0.0905$, $q^{(2)} = 0.0835$, and change the value of $q^{(3)}$ from 0.069 to 0.0735. As reported in Figure 3a, both of our algorithm and spectral clustering' performances become worse when $q^{(3)}$ increases. However, there is a regime of $q^{(3)}$ for which our algorithm's performance is much better than applying spectral clustering on single graphs alone.

Next, we fix $p = 0.1$, $q^{(2)} = 0.0835$, $q^{(3)} = 0.0705$, and change the value of $q^{(1)}$ from 0.087 to 0.0915. As shown in Figure 3b, in this setup the performance of our algorithm is better than that of applying $\mathbf{G}^{(3)}$ alone, even with a poor quality graph $\mathbf{G}^{(1)}$.

Case 2. We further examine a more complicated and interesting setting, in which the total number of clusters in $\mathbf{H}$ is more than the number of clusters observed from each graph. To this end, we construct three graphs with the following parameters:

- Graph $\mathbf{G}^{(1)}$ has $k_1 = 2$ clusters, each of which has $n^{(1)} = 6,000$ vertices. Vertices $1 \sim 6,000$ belong to cluster 1, and vertices $6,001 \sim 12,000$ belong to cluster 2.
- Graph $\mathbf{G}^{(2)}$ has $k_2 = 4$ clusters, each of which has $n^{(2)} = 3,000$ vertices. The clusters are formed by sequential blocks of vertices: $1 \sim 3,000$ (cluster 1), $3,001 \sim 6,000$ (cluster 2), $6,001 \sim 9,000$ (cluster 3), and $9,001 \sim 12,000$ (cluster 4).
- Graph $\mathbf{G}^{(3)}$ has $k_3 = 6$ clusters, each of which has $n^{(3)} = 2,000$ vertices. The vertices are equally divided.

This construction results in the master graph $\mathbf{H}$ with 8 clusters. We fix $q^{(1)} = 0.0905$, $q^{(2)} = 0.0835$, and change the value of $q^{(3)}$ from 0.075 to 0.0795. As shown in Figure 3c, the ratio of the corrected

classified vertices from our algorithm remains high even for a large value of $q^{(3)}$, under which spectral clustering is unable to correctly classify most vertices in $\mathbf{G}^{(3)}$.

6.2 Results on Real-World Data

We further report the performance of our algorithm on three widely used multi-modal datasets: MNIST-USPS (Peng et al., 2019), nuswide (Zhen et al., 2019), and Reuters (Amini et al., 2009); the details of these datasets are reported in the appendix. The Reuters dataset contains high-dimensional features, so we reduce the dimensions by retaining the 300 columns with the largest variances, resulting in a 300-dimensional input for this dataset. Since different modalities in each dataset contain the same number of clusters, we randomly merge 1/3 clusters in each modality by setting the corresponding features to all zeros matrix to formulate the datasets. For the Reuters dataset, we apply the same procedure in the first three modalities. For all datasets, feature values are normalized to the range $[0, 1]$ before training.

We compare the performance of our algorithm with five shallow multi-modal clustering methods: AWP (Nie et al., 2018), ASR-ETR (Ji & Feng, 2023), UDBGL (Fang et al., 2023), AEVC (Liu et al., 2024), and LMTC (Liu et al., 2025a). As reported in Table 1, our algorithm consistently outperforms all competing methods with respect to ACC, NMI, ARI, and Fscore. These results collectively validate the effectiveness of our designed algorithm.

Table 1: Performance of the algorithms on the three datasets

Algorithms	MNIST-USPS				nuswide				Reuters			
	ACC	NMI	ARI	Fscore	ACC	NMI	ARI	Fscore	ACC	NMI	ARI	Fscore
AWP	75.70	77.46	64.07	67.88	38.27	44.86	24.17	34.73	30.54	12.50	0.24	34.14
ASR-ETR	65.01	59.89	50.42	55.48	51.66	49.26	34.66	41.32	50.88	42.30	34.35	47.31
UDBGL	67.11	60.48	51.73	56.72	52.50	51.49	35.29	42.09	39.44	21.19	10.53	33.18
AEVC	72.64	65.96	58.14	62.53	39.16	52.78	33.79	41.81	38.28	26.01	9.28	36.26
LMTC	72.18	71.38	62.48	66.38	**61.79**	52.05	42.61	48.54	49.48	31.65	26.63	42.62
Ours	**83.76**	**88.31**	**80.8**	**86.41**	58.22	**63.83**	**45.28**	**53.62**	**62.92**	**68.87**	**54.98**	**69.06**
CVCL	60.82	78.45	56.03	71.24	42.67	52.46	31.12	47.00	54.92	45.30	36.58	50.02
MVCAN	88.44	84.57	79.85	82.44	60.42	61.25	44.79	55.07	66.88	68.95	55.17	67.42
SCM	90.58	84.10	81.31	83.53	48.17	57.74	41.62	52.38	63.70	49.40	42.40	55.92
SparseMVC	90.40	89.39	83.36	85.65	60.86	62.80	50.48	57.67	70.30	70.32	56.98	70.54
Ours+CRL	**95.40**	**90.75**	**90.30**	**91.37**	**65.48**	63.04	**50.64**	**57.98**	**74.04**	**75.53**	**63.60**	**75.94**

To further demonstrate the effectiveness of the proposed method, we compare our work with four representative deep multi-modal clustering methods: CVCL (Chen et al., 2023), MVCAN (Xu et al., 2024), SCM (Luo et al., 2024), and SparseMVC (Liu et al., 2025b). As shown in Table **??**, when directly comparing our algorithm with the deep learning based approaches, our algorithm is not the best performer; this is mainly because the other approaches apply neural networks to learn the features. To make a fair comparison with deep models, we utilize a weighted contrastive learning loss from the work in (Li et al., 2025) to learn higher quality representation from each modality. We term this variation of our method as Our+CRL. Specifically, for each modality an autoencoder is used to extract representation under the guidance of a weighted contrastive learning loss and a feature reconstruction loss. Then, we perform our method on the newly learned multi-modal representation for clustering. As reported in Table 1, our method equipped with learned multi-modal representation works better than the other deep methods on three datasets. The results verify the effectiveness of the proposed method. We highlight that, among the tested algorithms, our algorithm is the only one with a theoretical guarantee.

7 Conclusions

This paper studies clustering algorithms for capturing the modality-consensus clusters from multi-modal graphs and defines the modality-consensus clusters as data points in the same cluster if and only if they appear in the same optimal cluster for every single modality. The paper designs a multi-modal graph clustering algorithm via performing k-means on concatenated modality-specific spectral embeddings, which is shown to approximately capture the modality-consensus clusters. Experimental results on both synthetic and real-world datasets further demonstrate the effectiveness of our designed algorithm.

REFERENCES

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

Massih R Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views-an application to multilingual text categorization. *Advances in Neural Information Processing Systems*, 22, 2009.

Jie Chen, Hua Mao, Wai Lok Woo, and Xi Peng. Deep multiview clustering by contrasting cluster assignments. In *International Conference on Computer Vision*, pp. 16752–16761, 2023.

Jiecao Chen, He Sun, David Woodruff, and Qin Zhang. Communication-optimal distributed clustering. *Advances in Neural Information Processing Systems*, 29, 2016.

Vincent Cohen-Addad, Varun Kanade, and Frederik Mallmann-Trenn. Hierarchical clustering beyond the worst-case. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pp. 6201–6209, 2017.

Vincent Cohen-Addad, Tommaso D'Orsi, Silvio Lattanzi, and Rajai Nasser. Multi-view stochastic block models. In *International Conference on Machine Learning*, pp. 9180–9207, 2024.

Artur Czumaj, Pan Peng, and Christian Sohler. Testing cluster structure of graphs. In *ACM Symposium on Theory of Computing*, pp. 723–732, 2015.

Si-Guo Fang, Dong Huang, Xiao-Sha Cai, Chang-Dong Wang, Chaobo He, and Yong Tang. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):11436–11447, 2023.

Jintian Ji and Songhe Feng. Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization. In *International Conference on Computer Vision*, pp. 19343–19352, 2023.

Isabel M Kloumann, Johan Ugander, and Jon Kleinberg. Block models and personalized pagerank. *Proceedings of the National Academy of Sciences*, 114(1):33–38, 2017.

Pavel Kolev and Kurt Mehlhorn. A note on spectral clustering. In *European Symposium on Algorithms*, pp. 57–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016.

Abhishek Kumar and Hal Daumé. A co-training approach for multi-view spectral clustering. In *International Conference on Machine Learning*, pp. 393–400, 2011.

James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37:1–37:30, 2014.

Zhenglai Li, Yuqi Shi, Xiao He, and Chang Tang. Mask-informed deep contrastive incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12): 11822–11833, 2025.

Weixuan Liang, Xinwang Liu, Yong Liu, Chuan Ma, Yunping Zhao, Zhe Liu, and En Zhu. Consistency of multiple kernel clustering. In *International Conference on Machine Learning*, pp. 20650–20676, 2023.

Weixuan Liang, Xinwang Liu, Ke Liang, Jiyuan Liu, and En Zhu. Coke: Core kernel for more efficient approximation of kernel weights in multiple kernel clustering. In *Forty-second International Conference on Machine Learning*, 2025.

Ofir Lindenbaum, Arie Yeredor, Moshe Salhov, and Amir Averbuch. Multi-view diffusion maps. *Information Fusion*, 55:127–149, 2020.

Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian Qu, Yu Feng, Renxiang Guan, et al. Large-scale multi-view tensor clustering with implicit linear kernels. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 20727–20736, 2025a.

Ruimeng Liu, Xin Zou, Chang Tang, Xiao Zheng, Xingchen Hu, Kun Sun, and Xinwang Liu. Sparsemvc: Probing cross-view sparsity variations for multi-view clustering. In *Neural Information Processing Systems*, 2025b.

Suyuan Liu, Ke Liang, Zhibin Dong, Siwei Wang, Xihong Yang, Sihang Zhou, En Zhu, and Xinwang Liu. Learn from view correlation: An anchor enhancement strategy for multi-view clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 26151–26161, 2024.

Anand Louis and Rakesh Venkat. Planted models for k-way edge and vertex expansion. In *Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 23–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019.

Caixuan Luo, Jie Xu, Yazhou Ren, Junbo Ma, and Xiaofeng Zhu. Simple contrastive multi-view clustering with data-level fusion. In *International Joint Conference on Artificial Intelligence*, pp. 4697–4705, 2024.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Peter Macgregor and He Sun. A tighter analysis of spectral clustering, and beyond. In *International conference on machine learning*, pp. 14717–14742. PMLR, 2022.

Tomohiko Mizutani. Improved analysis of spectral algorithm for clustering. *Optimization Letters*, 15 (4):1303–1325, 2021.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pp. 849–856, 2001.

Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *International Conference on Knowledge Discovery and Data Mining*, pp. 2022–2030, 2018.

Pan Peng. Robust clustering oracle and local reconstructor of cluster structure of graphs. In *ACM-SIAM Symposium on Discrete Algorithms*, pp. 2953–2972. SIAM, 2020.

Pan Peng and Yuichi Yoshida. Average sensitivity of spectral clustering. In *International Conference on Knowledge Discovery and Data Mining*, pp. 1132–1140, 2020.

Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! *SIAM Journal on Computing*, 46(2):710–743, 2017.

Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *International conference on machine learning*, pp. 5092–5101, 2019.

He Sun and Luca Zanetti. Distributed graph clustering and sparsification. *ACM Transactions on Parallel Computing*, 6(3):1–23, 2019.

Daniel J Trosten, Sigurd Løkse, Robert Jenssen, and Michael C Kampffmeyer. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23976–23985, 2023.

Yorgos Tsitsikas and Evangelos E Papalexakis. Multi-view spectral clustering for graphs with multiple view structures. In *SIAM International Conference on Data Mining*, pp. 270–278. SIAM, 2025.

Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *IEEE/CVF International Conference on Computer Vision*, pp. 9234–9243, 2021.

Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, and Xiaofeng Zhu. Investigating and mitigating the side effects of noisy views for self-supervised clustering algorithms in practical multi-view scenarios. In *Computer Vision and Pattern Recognition*, pp. 22957–22966, 2024.

Liangli Zhen, Peng Hu, Xu Wang, and Dezhong Peng. Deep supervised cross-modal retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10394–10403, 2019.

APPENDIX

Appendix Contents

A Necessary statement **13**

A.1 AI use statement 13

A.2 Ethics statement 13

A.3 Reproducibility statement 13

B Omitted Details for Section 4 **13**

C Omitted Details for Section 5 **16**

C.1 Proof of Lemma 5.2 and 5.3 16

C.2 Proof of Theorem 2 20

C.3 Proof of Lemma 5.4 24

C.4 Proof of Lemma 5.5 25

D Omitted Details from Section 6 **29**

D.1 Omitted Details for t-SNE Visualization Results 29

D.2 Omitted Details for Multi-modal Datasets 29

A NECESSARY STATEMENT

A.1 AI USE STATEMENT

In this work, we used generative AI tools to assist with presentation refinement, code debugging and refactoring. Generative AI tools were also used to improve the clarity and organization of some technical explanations. All AI-assisted content was carefully reviewed by the authors. In particular, AI-assisted code modifications were inspected and tested before being used in the experiments. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

A.2 ETHICS STATEMENT

This work focuses on the theoretical and algorithmic study of multi-modal clustering and does not involve human-subject experiments or the collection of personally identifiable information.

A.3 REPRODUCIBILITY STATEMENT

We have taken several steps to facilitate the reproducibility of our results. The assumptions and formal definitions underlying the theoretical analysis are stated explicitly in the main text, while complete proofs of the theoretical results are provided in the appendix. Detailed descriptions of the proposed algorithm are also provided in the main paper and appendix.

B OMITTED DETAILS FOR SECTION 4

In this section, we present the detailed proof of Theorem 1.

Proof. We first prove the first part of Theorem 1.

For any $s \in [\ell]$, let $\mathbf{P}^{(s)}$ denote the orthogonal projection matrix onto $\operatorname{span}\left\{\mathbf{g}_1^{(s)}, \ldots, \mathbf{g}_{k_s}^{(s)}\right\}$. Recall that any two vertices u, v belong to the same cluster in $\mathbf{H}$ if and only if they belong to the same optimal cluster $\mathcal{S}_j^{(s)}$ for every $s \in [\ell]$. According to the construction of $\mathbf{H}$, every $\mathbf{y}_i$, $i \in [k]$, is generated from $\left\{\mathbf{g}_j^{(s)} : s \in [\ell],\ j \in [k_s]\right\}$, while every $\mathbf{g}_j^{(s)}$ can be reconstructed from $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$. Hence,

$$\operatorname{span}\left\{\mathbf{g}_1^{(1)}, \ldots, \mathbf{g}_{k_1}^{(1)}, \ldots, \mathbf{g}_1^{(\ell)}, \ldots, \mathbf{g}_{k_\ell}^{(\ell)}\right\} = \operatorname{span}\left\{\mathbf{y}_1, \ldots, \mathbf{y}_k\right\}.$$

Therefore, $\mathbf{y}_i$ admits the successive projection decomposition

$$\begin{aligned}
\mathbf{y}_i = & \sum_{p=1}^{k_1} \left\langle \mathbf{P}^{(1)}\mathbf{y}_i, \mathbf{g}_p^{(1)} \right\rangle \mathbf{g}_p^{(1)} \\
& + \sum_{p=1}^{k_2} \left\langle \mathbf{P}^{(2)} \left(\mathbf{I} - \mathbf{P}^{(1)}\right) \mathbf{y}_i, \mathbf{g}_p^{(2)} \right\rangle \mathbf{g}_p^{(2)} \\
& + \cdots + \sum_{p=1}^{k_\ell} \left\langle \mathbf{P}^{(\ell)} \prod_{r=1}^{\ell-1} \left(\mathbf{I} - \mathbf{P}^{(\ell-r)}\right) \mathbf{y}_i, \mathbf{g}_p^{(\ell)} \right\rangle \mathbf{g}_p^{(\ell)}.
\end{aligned} \tag{2}$$

For convenience, for every $s \in [\ell]$, define

$$\mathbf{a}_i^{(s)} \triangleq \mathbf{P}^{(s)} \prod_{r=1}^{s-1} \left(\mathbf{I} - \mathbf{P}^{(s-r)}\right) \mathbf{y}_i, \tag{3}$$

where the empty product for $s = 1$ is understood as the identity matrix. Since $\mathbf{a}_i^{(s)} \in \operatorname{span}\left\{\mathbf{g}_1^{(s)}, \ldots, \mathbf{g}_{k_s}^{(s)}\right\}$, we have $\mathbf{a}_i^{(s)} = \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \mathbf{g}_p^{(s)}$. Thus, equation 2 can be written as $\mathbf{y}_i = \sum_{s=1}^{\ell} \mathbf{a}_i^{(s)}$.

We first establish an energy bound for the vectors $\{\mathbf{a}_i^{(s)}\}_{s=1}^{\ell}$. Define $\mathbf{r}_i^{(0)} = \mathbf{y}_i$, and, for every $s \in [\ell]$, we have $\mathbf{a}_i^{(s)} = \mathbf{P}^{(s)}\mathbf{r}_i^{(s-1)}$ and $\mathbf{r}_i^{(s)} = \left(\mathbf{I} - \mathbf{P}^{(s)}\right)\mathbf{r}_i^{(s-1)}$. Since $\mathbf{P}^{(s)}$ is an orthogonal projection matrix, we have $\mathbf{a}_i^{(s)} \perp \mathbf{r}_i^{(s)}$. Moreover, we have $\mathbf{r}_i^{(s-1)} = \mathbf{a}_i^{(s)} + \mathbf{r}_i^{(s)}$. This gives that

$$\left\|\mathbf{r}_i^{(s-1)}\right\|^2 = \left\|\mathbf{a}_i^{(s)}\right\|^2 + \left\|\mathbf{r}_i^{(s)}\right\|^2.$$

Summing the above equality over $s = 1, \ldots, \ell$ yields

$$\sum_{s=1}^{\ell} \left\|\mathbf{a}_i^{(s)}\right\|^2 = \|\mathbf{y}_i\|^2 - \left\|\mathbf{r}_i^{(\ell)}\right\|^2 \leq \|\mathbf{y}_i\|^2 = 1, \tag{4}$$

where the last equality follows from the normalization of $\mathbf{y}_i$.

We next expand every $\mathbf{g}_p^{(s)}$ in the eigenvectors $\{\mathbf{f}_t^{(s)}\}_{t=1}^{n}$ as $\mathbf{g}_p^{(s)} = \sum_{t=1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)}$. Substituting the above expansion into equation 2, we obtain

$$\mathbf{y}_i = \sum_{s=1}^{\ell} \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \left(\sum_{t=1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)} \right). \tag{5}$$

Define

$$\mathbf{h}_i \triangleq \sum_{s=1}^{\ell} \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \left(\sum_{t=1}^{k_s} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)} \right). \tag{6}$$

Subtracting equation 6 from equation 5 gives

$$\mathbf{y}_i - \mathbf{h}_i = \sum_{s=1}^{\ell} \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \left(\sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)} \right). \tag{7}$$

For every $s \in [\ell]$ and $p \in [k_s]$, let

$$\mathbf{b}_p^{(s)} \triangleq \sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)}. \tag{8}$$

Further define

$$\mathbf{e}_i^{(s)} \triangleq \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \mathbf{b}_p^{(s)}. \tag{9}$$

Then

$$\mathbf{y}_i - \mathbf{h}_i = \sum_{s=1}^{\ell} \mathbf{e}_i^{(s)}. \tag{10}$$

We now bound each $\mathbf{e}_i^{(s)}$. For arbitrary $s \in [\ell]$, by the Cauchy–Schwarz inequality, we have

$$\left\|\mathbf{e}_i^{(s)}\right\|^2 = \left\| \sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle \mathbf{b}_p^{(s)} \right\|^2 \leq \left(\sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle^2 \right) \left(\sum_{p=1}^{k_s} \left\|\mathbf{b}_p^{(s)}\right\|^2 \right). \tag{11}$$

Since $\mathbf{a}_i^{(s)}$ belongs to $\operatorname{span}\{\mathbf{g}_1^{(s)}, \ldots, \mathbf{g}_{k_s}^{(s)}\}$ and the normalized cluster indicator vectors $\mathbf{g}_1^{(s)}, \ldots, \mathbf{g}_{k_s}^{(s)}$ are orthonormal, Parseval's identity gives

$$\sum_{p=1}^{k_s} \left\langle \mathbf{a}_i^{(s)}, \mathbf{g}_p^{(s)} \right\rangle^2 = \left\|\mathbf{a}_i^{(s)}\right\|^2. \tag{12}$$

On the other hand, since $\{\mathbf{f}_t^{(s)}\}_{t=1}^n$ is an orthonormal basis, we have

$$\left\|\mathbf{b}_p^{(s)}\right\|^2 = \left\|\sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle \mathbf{f}_t^{(s)}\right\|^2 = \sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle^2. \tag{13}$$

By the first part of Lemma 4.1,

$$\sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle^2 \le \frac{1}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)} \qquad \text{for every } p \in [k_s]. \tag{14}$$

Therefore,

$$\sum_{p=1}^{k_s} \left\|\mathbf{b}_p^{(s)}\right\|^2 = \sum_{p=1}^{k_s} \sum_{t=k_s+1}^{n} \left\langle \mathbf{g}_p^{(s)}, \mathbf{f}_t^{(s)} \right\rangle^2 \le \frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}. \tag{15}$$

Combining equation 11, equation 12, and equation 15, we obtain

$$\left\|\mathbf{e}_i^{(s)}\right\|^2 \le \frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)} \left\|\mathbf{a}_i^{(s)}\right\|^2. \tag{16}$$

Finally, combining equation 10 with the triangle inequality and equation 16, we have

$$\|\mathbf{y}_i - \mathbf{h}_i\| \le \sum_{s=1}^{\ell} \left\|\mathbf{e}_i^{(s)}\right\| \le \sum_{s=1}^{\ell} \sqrt{\frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}} \left\|\mathbf{a}_i^{(s)}\right\|. \tag{17}$$

Applying the Cauchy–Schwarz inequality to the last expression yields

$$\begin{aligned} \|\mathbf{y}_i - \mathbf{h}_i\|^2 &\le \left(\sum_{s=1}^{\ell} \sqrt{\frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}} \left\|\mathbf{a}_i^{(s)}\right\|\right)^2 \\ &\le \left(\sum_{s=1}^{\ell} \frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}\right) \left(\sum_{s=1}^{\ell} \left\|\mathbf{a}_i^{(s)}\right\|^2\right) \\ &\le \sum_{s=1}^{\ell} \frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}, \end{aligned} \tag{18}$$

where the last inequality follows from equation 4. Therefore,

$$\|\mathbf{y}_i - \mathbf{h}_i\|^2 \le \sum_{s=1}^{\ell} \frac{k_s}{\Upsilon_{\mathbf{G}^{(s)}}(k_s)}.$$

This proves the first statement of the theorem.

Now we prove the second statement. For every $i \in [\ell]$ and $j \in [k_i]$, we set

$$\begin{aligned} \hat{\mathbf{y}}_j^{(i)} &\triangleq \sum_{t=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_t^{(i)} \right\rangle \mathbf{g}_t^{(i)} = \sum_{t=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_t^{(i)} \right\rangle \left(\sum_{p=1}^{k} \left\langle \mathbf{g}_t^{(i)}, \mathbf{y}_p \right\rangle \mathbf{y}_p\right) \\ &= \sum_{t=1}^{k_i} \sum_{p=1}^{k} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_t^{(i)} \right\rangle \left\langle \mathbf{g}_t^{(i)}, \mathbf{y}_p \right\rangle \mathbf{y}_p, \end{aligned}$$

where $\hat{\mathbf{y}}_j^{(i)}$ is a linear combination of $\mathbf{g}_1^{(i)}, \ldots, \mathbf{g}_{k_i}^{(i)}$ and is also a linear combination of $\mathbf{y}_1, \ldots, \mathbf{y}_k$. By applying the second part of Theorem 4.1, we have that $\sum_{j=1}^{k_i} \left\|\mathbf{f}_j^{(i)} - \hat{\mathbf{y}}_j^{(i)}\right\|^2 \le \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}$, which proves the second part of the theorem. □

C Omitted Details for Section 5

In this section, we present the details omitted from Section 5, and this section is organized as follows: we first give the omitted proofs of Lemma 5.2 and 5.3 in Section C.1, then prove Theorem 2 in Section C.2, and finally give the omitted proofs of Lemma 5.5 in Section C.4.

C.1 Proof of Lemma 5.2 and 5.3

In this subsection, we analyze the properties of point embedding and show why simply concatenating modality-specific spectral embedding works for finding the underlying clusters. The proofs of Lemma 5.2 and 5.3 is similar to the one in (Macgregor & Sun, 2022). Here we first proof Lemma 5.2 as follows:

For every cluster $\mathcal{C}_s$, we define the vector $\mathbf{p}_s \in \mathbb{R}^t$ by $\mathbf{p}_s = \left(\mathbf{p}_s^{(1)}, \mathbf{p}_s^{(2)}, \ldots, \mathbf{p}_s^{(\ell)}\right)$, where $\mathbf{p}_s^{(i)}(j) = \frac{1}{\sqrt{|\mathcal{C}_s|}} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(s)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle$. We call $\{\mathbf{p}_s\}_{s=1}^k$ as the approximate centers of the point embedding from the optimal clusters $\{\mathcal{C}_s\}_{s=1}^k$. In following content, we prove that the resultant total k-means cost of the embedding points can be upper-bounded as follows:

Lemma C.1. *It holds that* $\sum_{s=1}^k \sum_{u\in\mathcal{C}_s} \|F(u) - \mathbf{p}_s\|^2 \le \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}$.

Proof. It holds by definition that

$$
\begin{aligned}
\sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \|F(u) - \mathbf{p}_s\|^2 &= \sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \sum_{i=1}^{\ell} \left\| F^{(i)}(u) - \mathbf{p}_s^{(i)} \right\|^2 \\
&= \sum_{i=1}^{\ell} \sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \left[\sum_{j=1}^{k_i} \left(\mathbf{f}_j^{(i)}(u) - \frac{\sum_{r=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_r^{(i)} \right\rangle \left\langle \mathbf{g}_r^{(i)}, \mathbf{y}_s \right\rangle}{\sqrt{|\mathcal{C}_s|}} \right)^2 \right] \\
&= \sum_{i=1}^{\ell} \sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \left[\sum_{j=1}^{k_i} \left(\mathbf{f}_j^{(i)}(u) - \sum_{r=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_r^{(i)} \right\rangle \left\langle \mathbf{g}_r^{(i)}, \mathbf{y}_s \right\rangle \mathbf{y}_s(u) \right)^2 \right] \\
&= \sum_{i=1}^{\ell} \sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \left[\sum_{j=1}^{k_i} \left(\mathbf{f}_j^{(i)}(u) - \sum_{r=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_r^{(i)} \right\rangle \left(\sum_{q=1}^{k} \left\langle \mathbf{g}_r^{(i)}, \mathbf{y}_q \right\rangle \mathbf{y}_q(u) \right) \right)^2 \right] \\
&= \sum_{i=1}^{\ell} \sum_{s=1}^{k} \sum_{u\in\mathcal{C}_s} \left[\sum_{j=1}^{k_i} \left(\mathbf{f}_j^{(i)}(u) - \hat{\mathbf{y}}_j^{(i)}(u) \right)^2 \right] \\
&= \sum_{i=1}^{\ell} \sum_{j=1}^{k_i} \left\| \mathbf{f}_j^{(i)} - \hat{\mathbf{y}}_j^{(i)} \right\|^2 \\
&\le \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},
\end{aligned}
$$

which proves the statement. □

Next, we analyze the lower and upper bound of $\left\|\mathbf{p}_s^{(i)}\right\|^2$ for $i \in [\ell]$ and $s \in [k]$.

Lemma C.2. *It holds for any* $s \in [k]$ *and* $i \in [\ell]$ *that*

$$
\frac{1}{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|} \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \le \left\|\mathbf{p}_s^{(i)}\right\|^2 \le \frac{1}{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}.
$$

Proof. By definition of $\mathbf{p}_s^{(i)}$, we have

$$\begin{aligned}
\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right| \left\|\mathbf{p}_s^{(i)}\right\|^2 &= \frac{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}{|\mathcal{C}_s|} \sum_{j=1}^{k_i} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(s)}^{(i)} \right\rangle^2 \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle^2 \\
&= \frac{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}{|\mathcal{C}_s|} \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle^2 \left\|\hat{\mathbf{f}}_{\theta^{(i)}(s)}^{(i)}\right\|^2 \\
&= \frac{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}{|\mathcal{C}_s|} \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle^2 \left(1 - \left\|\hat{\mathbf{f}}_{\theta^{(i)}(s)}^{(i)} - \mathbf{g}_{\theta^{(i)}(s)}^{(i)}\right\|^2\right) \\
&= \frac{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}{|\mathcal{C}_s|} \left(\sum_{u \in \mathcal{C}_s} \frac{1}{\sqrt{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}} \frac{1}{\sqrt{|\mathcal{C}_s|}}\right)^2 \left(1 - \left\|\hat{\mathbf{f}}_{\theta^{(i)}(s)}^{(i)} - \mathbf{g}_{\theta^{(i)}(s)}^{(i)}\right\|^2\right) \\
&= 1 - \left\|\hat{\mathbf{f}}_{\theta^{(i)}(s)}^{(i)} - \mathbf{g}_{\theta^{(i)}(s)}^{(i)}\right\|^2,
\end{aligned}$$

where the second equality holds by the definition of $\hat{\mathbf{f}}_{\theta^{(i)}(s)}^{(i)}$. By the first part of Theorem 1 we have that

$$\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right| \left\|\mathbf{p}_s^{(i)}\right\|^2 = 1 - \left\|\hat{\mathbf{f}}_s^{(i)} - \mathbf{g}_s^{(i)}\right\|^2 \geq 1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},$$

which implies

$$\frac{1}{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|} \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \leq \left\|\mathbf{p}_s^{(i)}\right\|^2 \leq \frac{1}{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}. \qquad \square$$

Lemma C.3. *It holds for any different $s, z \in [k]$ and $i \in [\ell]$ that*

$$\left\|\sqrt{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|} \cdot \mathbf{p}_s^{(i)} - \sqrt{\left|\mathcal{S}_{\theta^{(i)}(z)}^{(i)}\right|} \cdot \mathbf{p}_z^{(i)}\right\|^2 \geq 2 - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof. We have

$$\begin{aligned}
&\left\|\sqrt{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|} \cdot \mathbf{p}_s^{(i)} - \sqrt{\left|\mathcal{S}_{\theta^{(i)}(z)}^{(i)}\right|} \cdot \mathbf{p}_z^{(i)}\right\|^2 \\
&= \sum_{j=1}^{k_i} \left(\frac{\sqrt{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}}{\sqrt{|\mathcal{C}_s|}} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(s)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle - \frac{\sqrt{\left|\mathcal{S}_{\theta^{(i)}(z)}^{(i)}\right|}}{\sqrt{|\mathcal{C}_z|}} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(z)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(z)}^{(i)}, \mathbf{y}_z \right\rangle\right)^2 \\
&= \sum_{j=1}^{k_i} \frac{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}{|\mathcal{C}_s|} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(s)}^{(i)} \right\rangle^2 \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle^2 + \sum_{j=1}^{k_i} \frac{\left|\mathcal{S}_{\theta^{(i)}(z)}^{(i)}\right|}{|\mathcal{C}_z|} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(z)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(z)}^{(i)}, \mathbf{y}_z \right\rangle^2 \\
&\quad - 2 \sum_{j=1}^{k_i} \left(\frac{\sqrt{\left|\mathcal{S}_{\theta^{(i)}(s)}^{(i)}\right|}}{\sqrt{|\mathcal{C}_s|}} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(s)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(s)}^{(i)}, \mathbf{y}_s \right\rangle\right) \left(\frac{\sqrt{\left|\mathcal{S}_{\theta^{(i)}(z)}^{(i)}\right|}}{\sqrt{|\mathcal{C}_z|}} \left\langle \mathbf{f}_j^{(i)}, \mathbf{g}_{\theta^{(i)}(z)}^{(i)} \right\rangle \left\langle \mathbf{g}_{\theta^{(i)}(z)}^{(i)}, \mathbf{y}_z \right\rangle\right).
\end{aligned} \tag{19}$$

Based on the proof of Lemma C.2, the first two terms of equation 19 are totally upper bounded by $2\left(1-1/\Upsilon_{\mathbf{G}^{(i)}}(k_i)\right)$. For the third term of equation 19, we have that

$$\begin{aligned}
&\sum_{j=1}^{k_i}\left(\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(s)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s\right\rangle\right)\left(\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \mathbf{y}_z\right\rangle\right)\\
&=\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\sum_{j=1}^{k_i}\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(s)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s\right\rangle\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \mathbf{y}_z\right\rangle\\
&\leq\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \mathbf{y}_z\right\rangle\left|\left(\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}\right)^\top\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}\right|.
\end{aligned}$$

Notice that we can upper bound $\left|\left(\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}\right)^\top\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}\right|$ by

$$\begin{aligned}
&\left|\left(\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}\right)^\top\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}\right|\\
&\leq\left|\left(\mathbf{g}^{(i)}_{\theta^{(i)}(s)}+\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}-\mathbf{g}^{(i)}_{\theta^{(i)}(s)}\right)^\top\left(\mathbf{g}^{(i)}_{\theta^{(i)}(z)}+\hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}-\mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right)\right|\\
&\leq\left|\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}-\mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right\rangle+\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}-\mathbf{g}^{(i)}_{\theta^{(i)}(s)}\right\rangle+\left\langle \hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(s)}-\mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \hat{\mathbf{f}}^{(i)}_{\theta^{(i)}(z)}-\mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right\rangle\right|\\
&\leq\frac{3}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
&\sum_{j=1}^{k_i}\left(\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(s)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s\right\rangle\right)\left(\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\left\langle \mathbf{f}^{(i)}_j, \mathbf{g}^{(i)}_{\theta^{(i)}(z)}\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \mathbf{y}_z\right\rangle\right)\\
&\leq\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(s)}, \mathbf{y}_s\right\rangle\left\langle \mathbf{g}^{(i)}_{\theta^{(i)}(z)}, \mathbf{y}_z\right\rangle\frac{3}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\\
&=\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}{\sqrt{|\mathcal{C}_s|}}\frac{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}{\sqrt{|\mathcal{C}_z|}}\left(\sum_{u\in\mathcal{C}_s}\sum_{v\in\mathcal{C}_z}\frac{1}{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}}\frac{1}{\sqrt{|\mathcal{C}_s|}}\frac{1}{\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}}\frac{1}{\sqrt{|\mathcal{C}_z|}}\right)\frac{3}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\\
&\leq\frac{3}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.
\end{aligned}$$

This gives

$$\left\|\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\cdot\mathbf{p}^{(i)}_s-\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}\cdot\mathbf{p}^{(i)}_z\right\|^2\geq 2\left(1-\frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)-\frac{6}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\geq 2-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},$$

which proves the statement. □

Lemma C.4. *It holds for any different $s, z\in[k]$ and $i\in[\ell]$ that*

$$\left\|\frac{\mathbf{p}^{(i)}_s}{\left\|\mathbf{p}^{(i)}_s\right\|}-\frac{\mathbf{p}^{(i)}_z}{\left\|\mathbf{p}^{(i)}_z\right\|}\right\|^2\geq 2-\frac{20}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof. Without loss of generality, we assume that

$$\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\left\|\mathbf{p}^{(i)}_s\right\| \le \sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}\left\|\mathbf{p}^{(i)}_z\right\|$$

Let $\mathbf{a}_s = \sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\cdot\mathbf{p}^{(i)}_s$, and $\mathbf{a}_z = \sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}\cdot\mathbf{p}^{(i)}_z$, we have $\|\mathbf{a}_s\| \le \|\mathbf{a}_z\| \le 1$. By Lemma C.2, we obtain that

$$\begin{aligned}
\left\|\frac{\mathbf{p}^{(i)}_s}{\left\|\mathbf{p}^{(i)}_s\right\|} - \frac{\mathbf{p}^{(i)}_z}{\left\|\mathbf{p}^{(i)}_z\right\|}\right\| &\ge \left\|\mathbf{a}_s - \frac{\|\mathbf{a}_s\|}{\|\mathbf{a}_z\|}\mathbf{a}_z\right\| \\
&\ge \|\mathbf{a}_s - \mathbf{a}_z\| - (\|\mathbf{a}_z\| - \|\mathbf{a}_s\|) \\
&\ge \sqrt{2 - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} - \left(\sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|}\left\|\mathbf{p}^{(i)}_z\right\| - \sqrt{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\left\|\mathbf{p}^{(i)}_s\right\|\right) \\
&\ge \sqrt{2 - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} + \sqrt{1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} - 1 \\
&\ge \sqrt{2}\left(1 - \frac{4}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) + \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) - 1 \\
&\ge \sqrt{2} - \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},
\end{aligned}$$

where the second inequality holds by the triangle inequality, the third and fourth use Lemma C.3, and the fifth utilizes the fact that $\sqrt{1-x} \ge 1 - x$ holds when $0 \le x \le 1$. This gives

$$\left\|\frac{\mathbf{p}^{(i)}_s}{\left\|\mathbf{p}^{(i)}_s\right\|} - \frac{\mathbf{p}^{(i)}_z}{\left\|\mathbf{p}^{(i)}_z\right\|}\right\|^2 \ge \left(\sqrt{2} - \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)^2 \ge 2 - \frac{14\sqrt{2}}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \ge 2 - \frac{20}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

□

Now we show that any different pair of $\mathbf{p}_s$ and p_z are far from each other as follows:

Lemma C.5. *It holds for any different $s, z \in [k]$ that*

$$\|\mathbf{p}_s - \mathbf{p}_z\|^2 \ge \sum_{i=1}^{\ell} \frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|, \left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}}\left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

Proof. Without loss of generality, we assume $\left\|\mathbf{p}^{(i)}_s\right\| \ge \left\|\mathbf{p}^{(i)}_z\right\|$. Then, let $\left\|\mathbf{p}^{(i)}_z\right\| = \alpha\left\|\mathbf{p}^{(i)}_s\right\|$ for $\alpha \in [0, 1]$. By Lemma C.2, it holds that

$$\left\|\mathbf{p}^{(i)}_s\right\| \ge \frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|, \left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}}\left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

In addition, by the proof of Lemma C.4 we have

$$\left\langle \frac{\mathbf{p}^{(i)}_s}{\left\|\mathbf{p}^{(i)}_s\right\|}, \frac{\mathbf{p}^{(i)}_z}{\left\|\mathbf{p}^{(i)}_z\right\|} \right\rangle \le \sqrt{2} - \frac{1}{2}\left\|\frac{\mathbf{p}^{(i)}_s}{\left\|\mathbf{p}^{(i)}_s\right\|} - \frac{\mathbf{p}^{(i)}_z}{\left\|\mathbf{p}^{(i)}_z\right\|}\right\| \le \frac{\sqrt{2}}{2} + \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},$$

where the first inequality follows by that if $x^2 + y^2 = 1$, then $x + y \le \sqrt{2}$. One can understand the equation above by considering the right-angled triangle with one edge given by $\mathbf{p}^{(i)}_s / \left\|\mathbf{p}^{(i)}_s\right\|$ and

another edge given by $\left(\mathbf{p}_s^{(i)}/\left\|\mathbf{p}_s^{(i)}\right\|\right)\cdot\left(\mathbf{p}_z^{(i)}/\left\|\mathbf{p}_z^{(i)}\right\|\right)$. Then,

$$\begin{aligned}\left\|\mathbf{p}_s^{(i)}-\mathbf{p}_z^{(i)}\right\|^2 &= \left\|\mathbf{p}_s^{(i)}\right\|^2+\left\|\mathbf{p}_z^{(i)}\right\|^2-2\left\langle\frac{\mathbf{p}_s^{(i)}}{\left\|\mathbf{p}_s^{(i)}\right\|},\frac{\mathbf{p}_z^{(i)}}{\left\|\mathbf{p}_z^{(i)}\right\|}\right\rangle\left\|\mathbf{p}_s^{(i)}\right\|\left\|\mathbf{p}_z^{(i)}\right\| \\ &\geq (1+\alpha)\left\|\mathbf{p}_s^{(i)}\right\|^2-\left(\sqrt{2}+\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\alpha\left\|\mathbf{p}_s^{(i)}\right\|^2 \\ &\geq \left(1-(\sqrt{2}-1)\alpha-\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\left\|\mathbf{p}_s^{(i)}\right\|^2 \\ &\geq \frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|,\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}}\left(\frac{1}{2}-\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\left(1-\frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \\ &\geq \frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|,\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).\end{aligned}$$

This gives that

$$\|\mathbf{p}_s-\mathbf{p}_z\|^2\geq\sum_{i=1}^{\ell}\frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|,\left|\mathcal{S}^{(i)}_{\theta^{(i)}(z)}\right|\right\}}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right),$$

which proves the statement. □

C.2 Proof of Theorem 2

Now we present the analysis of misclassified error for our algorithm in a manner similar to the way used in (Macgregor & Sun, 2022). Let $\{\mathcal{R}_q\}_{q=1}^k$ be the output of our algorithm, and for every output cluster $\mathcal{R}_q$ we define the corresponding center of the embedded points $\{F(u)\}$ by $\mathbf{c}_q$. We define the function $\sigma:[k]\to[k]$ by

$$\sigma(q)=\arg\min_{s\in[k]}\|\mathbf{p}_s-\mathbf{c}_q\|^2, \tag{20}$$

where $\sigma(q)$ is the order of cluster in $\{\mathcal{C}_s\}_{q=1}^k$ that most close to $\mathbf{c}_q$. With this function, we assign every $\mathbf{c}_q$ with its nearest cluster center $\mathbf{p}_{\sigma(q)}$. Noticed that there might exist different $q,q'\in[k]$ such that $\mathcal{R}_q$ and $\mathcal{R}_{q'}$ map to the same $\mathcal{C}_{\sigma(q)}$. To this end, for any fixed $\sigma:[k']\to[k]$ and $s\in[k]$, we define $\mathcal{M}_{\sigma,s}$ by

$$\mathcal{M}_{\sigma,s}\triangleq\bigcup_{q:\,\sigma(q)=s}\mathcal{R}_q. \tag{21}$$

In the following lemma, we show that the total misclassified size with respect to each cluster under well mapped $\mathcal{R}_s$ and $\mathcal{C}_{\sigma(s)}$ can be upper bounded:

Lemma C.6. *Let $\{\mathbf{G}^{(i)}\}_{i=1}^{\ell}$ be multiple graphs, each graph be with k_i clusters of almost balanced size, and σ and $\mathcal{M}_{\sigma,s}$ be defined as defined in equation 21. If $\Upsilon_{\mathbf{G}^{(i)}}(k_i)\geq 32$, it holds that*

$$\sum_{s=1}^{k}|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|\leq\frac{128\cdot(1+\mathsf{APT})}{\sum_{i=1}^{\ell}k_i}\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\cdot n.$$

Proof. Let $\mathcal{B}_{qs}=\mathcal{R}_q\cap\mathcal{C}_s$ be the vertices in $\mathcal{R}_q$ which belong to the true cluster $\mathcal{C}_s$. Then, we have that

$$\begin{aligned}\sum_{i=1}^{\ell}\sum_{s=1}^{k}\frac{|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|} &= \sum_{i=1}^{\ell}\sum_{q=1}^{k}\sum_{\substack{s=1\\ s\neq\sigma(q)}}^{k}|B_{qs}|\left(\frac{1}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(\sigma(q))}\right|}+\frac{1}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\right) \\ &\leq 2\sum_{i=1}^{\ell}\sum_{q=1}^{k}\sum_{\substack{s=1\\ s\neq\sigma(q)}}^{k}\frac{|B_{qs}|}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(\sigma(q))}\right|,\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|\right\}}.\end{aligned} \tag{22}$$

Noticed that the total k-means cost of the partition returned by our algorithm is

$$\begin{aligned}
\mathsf{COST}(\mathcal{R}_1,\ldots,\mathcal{R}_k) &= \sum_{q=1}^{k}\sum_{u\in\mathcal{R}_q}\|F(u)-\mathbf{c}_q\|^2 \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\|F(u)-\mathbf{c}_q\|^2 \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\left(\frac{\|\mathbf{p}_s-\mathbf{c}_q\|^2}{2}-\|F(u)-\mathbf{p}_s\|^2\right) \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\frac{\left\|\mathbf{p}_s-\mathbf{p}_{\sigma(q)}\right\|^2}{8}-\sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\|F(u)-\mathbf{p}_s\|^2 \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}|\mathcal{B}_{qs}|\frac{\left\|\mathbf{p}_s-\mathbf{p}_{\sigma(i)}\right\|^2}{8}-\sum_{s=1}^{k}\sum_{u\in\mathcal{C}_s}\left\|F(u)-\mathbf{p}^{(s)}\right\|^2,
\end{aligned}$$

where the second inequality follows by the inequality of

$$\|\mathbf{a}-\mathbf{b}\|^2\geq\frac{\|\mathbf{b}-\mathbf{c}\|^2}{2}-\|\mathbf{a}-\mathbf{c}\|^2.$$

Notice that $\mathbf{c}_i$ is closer to $\mathbf{p}_{\sigma(i)}$ than $\mathbf{p}_s$, by applying the triangle inequality and AM-GM inequality we have that

$$\|\mathbf{p}_s-\mathbf{c}_q\|^2\geq\frac{1}{2}\cdot\left(\|\mathbf{p}_s-\mathbf{c}_q\|^2+\left\|\mathbf{p}_{\sigma(i)}-\mathbf{c}_q\right\|^2\right)\geq\frac{1}{4}\cdot\left\|\mathbf{p}_s-\mathbf{p}_{\sigma(i)}\right\|^2.$$

This gives

$$\begin{aligned}
&\mathsf{COST}(\mathcal{R}_1,\ldots,\mathcal{R}_k) \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\frac{\left\|\mathbf{p}_s-\mathbf{p}_{\sigma(q)}\right\|^2}{8}-\sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\sum_{u\in\mathcal{B}_{qs}}\|F(u)-\mathbf{p}_s\|^2 \\
&\geq \sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}|\mathcal{B}_{qs}|\frac{\left\|\mathbf{p}_s-\mathbf{p}_{\sigma(i)}\right\|^2}{8}-\sum_{s=1}^{k}\sum_{u\in\mathcal{C}_s}\left\|F(u)-\mathbf{p}^{(s)}\right\|^2 \\
&\geq \sum_{i=1}^{\ell}\sum_{q=1}^{k}\sum_{\substack{1\leq s\leq k\\ s\neq\sigma(q)}}\frac{|\mathcal{B}_{qs}|}{8}\frac{1}{\min\left\{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|,\left|\mathcal{S}^{(i)}_{\theta^{(i)}(\sigma(i))}\right|\right\}}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)-\sum_{s=1}^{k}\sum_{u\in\mathcal{C}_s}\|F(u)-\mathbf{p}_s\|^2.
\end{aligned}$$

where the last inequality holds by Lemma 5.3. By using equation 22, we have

$$\begin{aligned}
&\mathsf{COST}(\mathcal{R}_1,\ldots,\mathcal{R}_k) \\
&\geq \sum_{i=1}^{\ell}\left(\frac{1}{16}\cdot\left(\sum_{s=1}^{k}\frac{|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\right)\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)-\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).
\end{aligned}$$

Since $\mathsf{COST}\left(\mathcal{R}_1,\ldots,\mathcal{R}_k\right)\leq\mathsf{APT}\cdot\sum_{i=1}^{\ell}k_i/\Upsilon_{\mathbf{G}^{(i)}}(k_i)$, we have

$$\sum_{i=1}^{\ell}\left(\frac{1}{16}\cdot\left(\sum_{s=1}^{k}\frac{|M_{\sigma,s}\triangle\mathcal{C}_s|}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\right)\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)-\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\leq\mathsf{APT}\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

This gives that

$$\sum_{i=1}^{\ell}\left(\sum_{s=1}^{k}\frac{|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\right)\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\leq 16\cdot(1+\mathsf{APT})\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

By the assumption that the clusters in every $\mathbf{g}^{(i)}$ is almost balanced in size, i.e., $\left|\mathcal{S}^{(i)}_q\right| \in (n/(2k_i),(2n)/k_i)$ holds for every $i\in[\ell]$ and $q\in[k_i]$, we have

$$\sum_{s=1}^{k}\frac{|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|}{\left|\mathcal{S}^{(i)}_{\theta^{(i)}(s)}\right|}\geq\frac{k_i}{2n}\sum_{s=1}^{k}|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|\,.$$

Thus, we have

$$\left(\sum_{s=1}^{k}|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|\right)\sum_{i=1}^{\ell}\frac{k_i}{2n}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\leq 16\cdot(1+\mathsf{APT})\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

By the assumption that $\Upsilon_{\mathbf{G}^{(i)}}(k_i)\geq 32$ for any $i\in[\ell]$, we have

$$\begin{aligned}\sum_{s=1}^{k}|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s| &\leq 16\cdot(1+\mathsf{APT})\cdot\left(\sum_{i=1}^{\ell}\frac{k_i}{2\cdot n}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\right)^{-1}\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\\ &\leq\frac{128\cdot(1+\mathsf{APT})}{\sum_{i=1}^{\ell}k_i}\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\cdot n.\end{aligned}$$

This completes the proof. □

Now we present the detailed proof of Theorem 2 in the following part.

Proof of Theorem 2. By Lemma 5.2, we have

$$\mathsf{COST}(\mathcal{C}_1,\ldots,\mathcal{C}_k)\leq\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Combining this with the fact that one can apply an approximate k-means clustering algorithm with approximation ratio APT for spectral clustering, we have that

$$\mathsf{COST}(\mathcal{R}_1,\ldots,\mathcal{R}_k)\leq\mathsf{APT}\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Then, let $\sigma:[k]\to[k]$ be the function which assigns the clusters $\mathcal{R}_1,\ldots,\mathcal{R}_k$ to the ground truth clusters such that

$$\sigma(q)=\arg\min_{j\in[k]}\|\mathbf{p}_s-\mathbf{c}_q\|^2.$$

Then, it holds by Lemma C.6 that

$$\epsilon(\sigma)\triangleq\sum_{s=1}^{k}|\mathcal{M}_{\sigma,s}\triangle\mathcal{C}_s|\leq\frac{128\cdot(1+\mathsf{APT})}{\sum_{i=1}^{\ell}k_i}\cdot\sum_{i=1}^{\ell}\frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\cdot n. \tag{23}$$

Suppose σ is not a permutation. Consequently, there exists $s\in[k]$ where $\mathcal{M}_{\sigma,s}=\emptyset$, and distinct $x,y\in[k]$ such that $\sigma(x)=\sigma(y)=s'$ for some $s'\neq s$. We construct $\sigma':[k]\to[k]$ from σ as follows:

- Set $\sigma'(z)=s$ if $z=x$;
- Set $\sigma'(z)=\sigma(z)$ for any other $z\in[k]\setminus\{x\}$.

This procedure can be applied inductively to transform σ into a permutation. From the definitions of $\epsilon(\cdot)$ and σ', the difference $\epsilon(\sigma') - \epsilon(\sigma)$ can be expressed as:

$$\epsilon(\sigma') - \epsilon(\sigma) = \underbrace{(|\mathcal{M}_{\sigma',s} \triangle \mathcal{C}_s| - |\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s|)}_{=:\alpha} + \underbrace{(|\mathcal{M}_{\sigma',s'} \triangle \mathcal{C}_{s'}| - |\mathcal{M}_{\sigma,s'} \triangle \mathcal{C}_{s'}|)}_{=:\beta}. \tag{24}$$

We analyze four cases based on the signs of α and β. For each, we bound the cost incurred by the transition $\sigma \to \sigma'$, then aggregate these bounds to determine the total cost of the complete construction of a permutation.

Case 1: $\alpha < 0, \beta < 0$**.** In this case, it is clear that $\epsilon(\sigma') - \epsilon(\sigma) \leq 0$, and hence the total introduced cost is at most 0.

Case 2: $\alpha > 0, \beta < 0$**.** In this case, we have

$$\begin{aligned}\epsilon(\sigma') - \epsilon(\sigma) &= |\mathcal{M}_{\sigma',s} \triangle \mathcal{C}_s| - |\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s| + |\mathcal{M}_{\sigma,s'} \triangle \mathcal{C}_{s'}| - |\mathcal{M}_{\sigma',s'} \triangle \mathcal{C}_{s'}| \\ &= |\mathcal{R}_x \setminus \mathcal{C}_s| - |\mathcal{R}_x \cap \mathcal{C}_s| + |\mathcal{R}_x \setminus \mathcal{C}_{s'}| - |\mathcal{R}_x \triangle \mathcal{C}_{s'}| \\ &\leq 2 \cdot |\mathcal{R}_x \setminus \mathcal{C}_{s'}|,\end{aligned}$$

As each set $\mathcal{R}_x$ is moved at most once during the transformation to a permutation, the total cost incurred in this case is bounded by

$$\sum_{s'=1}^{k} \sum_{\mathcal{R}_x \in \mathcal{M}_{\sigma,s'}} 2 \cdot |\mathcal{R}_x \setminus \mathcal{C}_{s'}| \leq 2 \cdot \sum_{s'=1}^{k} |\mathcal{M}_{\sigma,s'} \triangle \mathcal{C}_{s'}| \leq \frac{256 \cdot (1 + \mathsf{APT})}{\sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n.$$

Case 3: $\alpha > 0, \beta > 0$**.** In this case, we have

$$\begin{aligned}\epsilon(\sigma') - \epsilon(\sigma) &= |\mathcal{M}_{\sigma',s} \triangle \mathcal{C}_s| - |\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s| + |\mathcal{M}_{\sigma',s'} \triangle \mathcal{C}_{s'}| - |\mathcal{M}_{\sigma,s'} \triangle \mathcal{C}_{s'}| \\ &= |\mathcal{R}_x \setminus \mathcal{C}_s| - |\mathcal{R}_x \cap \mathcal{C}_s| + |\mathcal{R}_x \cap \mathcal{C}_{s'}| - |\mathcal{R}_x \setminus \mathcal{C}_{s'}| \\ &\leq 2 \cdot |\mathcal{R}_x \cap \mathcal{C}_{s'}| \\ &\leq 2 \cdot |\mathcal{C}_{s'}|,\end{aligned}$$

We will consider the total introduced cost due to this case and Case 4 together, and so let's first examine Case 4.

Case 4: $\alpha < 0, \beta > 0$**.** In this case, we have

$$\begin{aligned}\epsilon(\sigma') - \epsilon(\sigma) &= |\mathcal{M}_{\sigma',s'} \triangle \mathcal{C}_{s'}| - |\mathcal{M}_{\sigma,s'} \triangle \mathcal{C}_{s'}| \\ &\leq |\mathcal{R}_x \triangle \mathcal{C}_{s'}| - |\mathcal{R}_x \setminus \mathcal{C}_{s'}| \\ &\leq |\mathcal{C}_{s'}|.\end{aligned}$$

Now, let us bound the total number of times we need to construct σ' to obtain a permutation. For any s with $\mathcal{M}_{\sigma,s} = \emptyset$, we have

$$\frac{|\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s|}{|\mathcal{C}_s|} = \frac{|\mathcal{C}_s|}{|\mathcal{C}_s|} = 1,$$

so the total number of required iterations is upper bounded by

$$|\{s : \mathcal{M}_{\sigma,s} = \emptyset\}| \leq \sum_{s=1}^{k} \frac{|\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s|}{|\mathcal{C}_s|}$$

Recall the assumption of almost-balanced clusters on $\mathbf{G}^{(i)}$ for $i \in [\ell]$, and the condition of $|\mathcal{C}_s|$ for $s \in [k]$, we have

$$\frac{n}{2k} \leq |\mathcal{C}_s| \leq \frac{2n}{\max\{k_i\}}.$$

This gives that

$$\begin{aligned}|\{s : \mathcal{M}_{\sigma,s} = \emptyset\}| &\leq \frac{2k}{n} \sum_{s=1}^{k} |\mathcal{M}_{\sigma,s} \triangle \mathcal{C}_s| \\ &\leq \frac{256k \cdot (1 + \mathsf{APT})}{\sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\end{aligned}$$

As such, the total introduced cost due to Cases 3 and 4 is at most

$$2 \cdot |\mathcal{C}_s| \frac{256k \cdot (1+\mathsf{APT})}{\sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \leq \frac{1024k \cdot (1+\mathsf{APT})}{\max\{k_i\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n.$$

Putting everything together, we have that

$$\begin{aligned}\epsilon(\sigma^\star) &\leq \epsilon(\sigma) + \frac{1024k \cdot (1+\mathsf{APT})}{\max\{k_i\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n + \frac{256 \cdot (1+\mathsf{APT})}{\sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n \\ &\leq \frac{1408k \cdot (1+\mathsf{APT})}{\max\{k_i\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n.\end{aligned}$$

This implies that

$$\sum_{s=1}^{k} |\mathcal{M}_{\sigma^\star,s} \triangle \mathcal{C}_s)| \leq \frac{1408k \cdot (1+\mathsf{APT})}{\max\{k_i\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \cdot n.$$

Without loss of generality We assume that there is a universal constant $C \in \mathbb{R}^+$ such that every modality $\mathbf{G}^{(i)}$ has a clear structure of clusters formalized by $\Upsilon_{\mathbf{G}^{(i)}}(k_i) \geq C$ for some $k_i \in [n]$. In this case, we have

$$\sum_{s=1}^{k} |\mathcal{M}_{\sigma^\star,s} \triangle \mathcal{C}_s| = O\left(\frac{k}{\max\{k_i\} \cdot \sum_{i=1}^{\ell} k_i} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \cdot n,$$

and completes the proof.

□

C.3 Proof of Lemma 5.4

Lemma C.7. *Let $\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'})$ be the cost returned by a k'-means algorithm with approximation ratio APT, where $k' \in [k, k^*]$. Then, it holds that*

$$\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'}) \leq \mathsf{APT} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof of Lemma 5.4. The property of the k-means objective is that the optimal cost is monotonically non-increasing with respect to the number of clusters. Since $k' \geq k$, we have

$$\mathsf{OPT}(k') \leq \mathsf{OPT}(k).$$

where $\mathsf{OPT}(k')$ and $\mathsf{OPT}(k)$ denote the minimum cost for k'-means and k-means for the set $\mathcal{V}$, respectively. Then, the Lemma 5.2 gives that

$$\mathsf{OPT}(k') \leq \mathsf{OPT}(k) \leq \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Recall that $\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'})$ be the cost returned by a k'-means algorithm with approximation ratio APT, we have

$$\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'}) \leq \mathsf{APT} \cdot \mathsf{OPT}(k').$$

This gives

$$\mathsf{COST}(\mathcal{R}_1, \ldots, \mathcal{R}_{k'}) \leq \mathsf{APT} \cdot \sum_{i=1}^{\ell} \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

This completes the proof. □

C.4 Proof of Lemma 5.5

In this subsection we present the detailed proof of Lemma 5.5. Before that, we first illustrate some properties of modality-specific embedding with respect to each graph as shown in Lemmas C.8 and C.12, which is useful for proving Lemma 5.5. Let $\{\mathcal{S}_j^{(i)}\}_{j=1}^{k_i}$ be the optimal clusters for $\mathbf{G}^{(i)}$ for $i \in [\ell]$, we define the vector $\mathbf{s}_j^{(i)} \in \mathbb{R}^{k_i}$ for every $\mathcal{S}_j^{(i)}$ by

$$\mathbf{s}_j^{(i)}(r) = \frac{1}{\sqrt{\left|\mathcal{S}_j^{(i)}\right|}} \left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_j^{(i)} \right\rangle.$$

The vectors $\{\mathbf{s}_j\}_{j=1}^{k_i}$ are denoted as the approximate center of the point embedding from the optimal clusters $\{\mathcal{S}_j^{(i)}\}_{j=1}^{k_i}$. Based on this, the following lemma gives an upper bound of the k_i-means clustering cost for finding the optimal clusters $\{\mathcal{S}_j^{(i)}\}_{j=1}^{k_i}$.

Lemma C.8. *It holds that*

$$\sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \left\| F^{(i)}(u) - \mathbf{s}_j^{(i)} \right\|^2 \leq \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof. This part mainly follows the proof pipeline of Lemma C.1. By definition of k_i-means and applying the transformation, e.g., $\left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_j^{(i)} \right\rangle \mathbf{g}_j^{(i)}(u) = \sum_{x=1}^{k_i} \left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_x^{(i)} \right\rangle \mathbf{g}_x^{(i)}(u) = \hat{\mathbf{g}}_r^{(i)}(u)$ for $u \in \mathcal{S}_j^{(i)}$, we have that

$$\begin{aligned}
\sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \left\| F^{(i)}(u) - \mathbf{s}_j^{(i)} \right\|^2 &= \sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \left[\sum_{r=1}^{k_i} \left(\mathbf{f}_r^{(i)}(u) - \frac{\left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_j^{(i)} \right\rangle}{\sqrt{\left|\mathcal{S}_j^{(i)}\right|}} \right)^2 \right] \\
&= \sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \sum_{r=1}^{k_i} \left(\mathbf{f}_r^{(i)}(u) - \left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_j^{(i)} \right\rangle \mathbf{g}_j^{(i)}(u) \right)^2 \\
&= \sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \sum_{r=1}^{k_i} \left(\mathbf{f}_r^{(i)}(u) - \hat{\mathbf{g}}_r^{(i)}(u) \right)^2 \\
&= \sum_{r=1}^{k_i} \left(\mathbf{f}_r^{(i)} - \hat{\mathbf{g}}_r^{(i)} \right)^2 \\
&\leq \frac{k_i}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},
\end{aligned}$$

where the last inequality follows by the second statement of Lemma 4.1. □

Next, we give the lower and upper bound of $\left\| \mathbf{s}_j^{(i)} \right\|^2$ for $i \in [\ell]$ and $j \in [k_i]$ as follows:

Lemma C.9. *It holds for any $j \in [k_i]$ and $i \in [\ell]$ that*

$$\frac{1}{\left|\mathcal{S}_j^{(i)}\right|} \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right) \leq \left\| \mathbf{s}_j^{(i)} \right\|^2 \leq \frac{1}{\left|\mathcal{S}_j^{(i)}\right|}.$$

Proof. This part mainly follows the proof pipeline of Lemma C.2. By definition of $\mathbf{s}_j^{(i)}$, we have

$$\left|\mathcal{S}_j^{(i)}\right| \left\| \mathbf{s}_j^{(i)} \right\|^2 = \sum_{r=1}^{k_i} \left\langle \mathbf{f}_r^{(i)}, \mathbf{g}_j^{(i)} \right\rangle^2 = 1 - \left\| \hat{\mathbf{f}}_j^{(i)} - \mathbf{g}_j^{(i)} \right\|^2.$$

By using the first statement of Lemma 4.1, we have

$$\frac{1}{\left|\mathcal{S}_j^{(i)}\right|}\left(1-\frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \leq \left\|\mathbf{s}_j^{(i)}\right\|^2 \leq \frac{1}{\left|\mathcal{S}_j^{(i)}\right|}.$$

□

Lemma C.10. *It holds for any different $j, j' \in [k_i]$ and $i \in [\ell]$ that*

$$\left\|\sqrt{\left|\mathcal{S}_j^{(i)}\right|}\cdot\mathbf{s}_j^{(i)} - \sqrt{\left|\mathcal{S}_{j'}^{(i)}\right|}\cdot\mathbf{s}_{j'}^{(i)}\right\|^2 \geq 2-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof. This part mainly follows the proof pipeline of Lemma C.3. We have

$$\begin{aligned}
\left\|\sqrt{\left|\mathcal{S}_j^{(i)}\right|}\cdot\mathbf{s}_j^{(i)} - \sqrt{\left|\mathcal{S}_{j'}^{(i)}\right|}\cdot\mathbf{s}_{j'}^{(i)}\right\|^2 &= \sum_{r=1}^{k_i}\left(\left\langle\mathbf{f}_r^{(i)},\mathbf{g}_j^{(i)}\right\rangle - \left\langle\mathbf{f}_r^{(i)},\mathbf{g}_{j'}^{(i)}\right\rangle\right)^2 \\
&= \left(\sum_{r=1}^{k_i}\left\langle\mathbf{f}_r^{(i)},\mathbf{g}_j^{(i)}\right\rangle^2\right) + \left(\sum_{r=1}^{k_i}\left\langle\mathbf{f}_r^{(i)},\mathbf{g}_{j'}^{(i)}\right\rangle^2\right) - 2\sum_{r=1}^{k_i}\left\langle\mathbf{f}_r^{(i)},\mathbf{g}_j^{(i)}\right\rangle\left\langle\mathbf{f}_r^{(i)},\mathbf{g}_{j'}^{(i)}\right\rangle \\
&\geq \left\|\hat{\mathbf{f}}_j^{(i)}\right\|^2 + \left\|\hat{\mathbf{f}}_{j'}^{(i)}\right\|^2 - 2\left|\left(\hat{\mathbf{f}}_j^{(i)}\right)^\top\hat{\mathbf{f}}_{j'}^{(i)}\right| \\
&\geq 2\cdot\left(1-\frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) - 2\left|\left(\hat{\mathbf{f}}_j^{(i)}\right)^\top\hat{\mathbf{f}}_{j'}^{(i)}\right|.
\end{aligned}$$

For the second term, we transfer it as the same way shown in Lemma C.3, we have

$$\begin{aligned}
2\left|\left(\hat{\mathbf{f}}_j^{(i)}\right)^\top\hat{\mathbf{f}}_{j'}^{(i)}\right| &= 2\left|\left(\mathbf{g}_j^{(i)}+\hat{\mathbf{f}}_j^{(i)}-\mathbf{g}_j^{(i)}\right)^\top\left(\mathbf{g}_{j'}^{(i)}+\hat{\mathbf{f}}_{j'}^{(i)}-\mathbf{g}_{j'}^{(i)}\right)\right| \\
&= 2\left|\left\langle\mathbf{g}_j^{(i)},\hat{\mathbf{f}}_{j'}^{(i)}-\mathbf{g}_{j'}^{(i)}\right\rangle + \left\langle\mathbf{g}_{j'}^{(i)},\hat{\mathbf{f}}_j^{(i)}-\mathbf{g}_j^{(i)}\right\rangle + \left\langle\hat{\mathbf{f}}_j^{(i)}-\mathbf{g}_j^{(i)},\hat{\mathbf{f}}_{j'}^{(i)}-\mathbf{g}_{j'}^{(i)}\right\rangle\right| \\
&\leq \frac{6}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.
\end{aligned}$$

This gives

$$\left\|\sqrt{\left|\mathcal{S}_j^{(i)}\right|}\cdot\mathbf{s}_j^{(i)} - \sqrt{\left|\mathcal{S}_{j'}^{(i)}\right|}\cdot\mathbf{s}_{j'}^{(i)}\right\|^2 \geq 2-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

which proves the statement. □

Lemma C.11. *It holds for any different $j, j' \in [k_i]$ and $i \in [\ell]$ that*

$$\left\|\frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|} - \frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|}\right\|^2 \geq 2-\frac{20}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

Proof. This part mainly follows the proof pipeline of Lemma C.4. In the same way, without loss of generality, we assume that

$$\sqrt{\left|\mathcal{S}_j^{(i)}\right|}\left\|\mathbf{s}_j^{(i)}\right\| \leq \sqrt{\left|\mathcal{S}_j^{(i)}\right|}\left\|\mathbf{s}_{j'}^{(i)}\right\|$$

Let $\mathbf{a}_j = \sqrt{\left|\mathcal{S}_j^{(i)}\right|} \cdot \mathbf{s}_j^{(i)}$, and $\mathbf{a}_{j'} = \sqrt{\left|\mathcal{S}_{j'}^{(i)}\right|} \cdot \mathbf{s}_{j'}^{(i)}$, implying that $\|\mathbf{a}_s\| \leq \|\mathbf{a}_z\| \leq 1$. Combing Lemmas C.9 and C.10, we have

$$\begin{aligned}
\left\| \frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|} - \frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|} \right\| &\geq \left\| \mathbf{a}_j - \frac{\|\mathbf{a}_j\|}{\|\mathbf{a}_{j'}\|} \mathbf{a}_{j'} \right\| \\
&\geq \|\mathbf{a}_j - \mathbf{a}_{j'}\| - (\|\mathbf{a}_{j'}\| - \|\mathbf{a}_j\|) \\
&\geq \sqrt{2 - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} - \left(\sqrt{\left|\mathcal{S}_{j'}^{(i)}\right|} \left\|\mathbf{s}_{j'}^{(i)}\right\| - \sqrt{\left|\mathcal{S}_j^{(i)}\right|} \left\|\mathbf{s}_j^{(i)}\right\| \right) \\
&\geq \sqrt{2 - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} + \sqrt{1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}} - 1 \\
&\geq \sqrt{2} \left(1 - \frac{4}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) + \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) - 1 \\
&\geq \sqrt{2} - \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},
\end{aligned}$$

where the second inequality holds by the triangle inequality, and the third and fourth use Lemma C.10. We also use the fact that for $0 \leq x \leq 1$, it is the case that $\sqrt{1-x} \geq 1 - x$. This gives

$$\left\| \frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|} - \frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|} \right\|^2 \geq \left(\sqrt{2} - \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)^2 \geq 2 - \frac{14\sqrt{2}}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \geq 2 - \frac{20}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}.$$

which proves the statement. □

The next lemma proves that, for different j and j', $\mathbf{s}_j^{(i)}$ and $\mathbf{s}_{j'}^{(i)}$ are far from each other.

Lemma C.12. *It holds for any different $j, j' \in [k_i]$ that:*

$$\left\|\mathbf{s}_j^{(i)} - \mathbf{s}_{j'}^{(i)}\right\|^2 \geq \frac{1}{\min\left\{\left|\mathcal{S}_j^{(i)}\right|, \left|\mathcal{S}_{j'}^{(i)}\right|\right\}} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

Proof. This part mainly follows the proof pipeline of Lemma C.5. In the same way, we assume that $\left\|\mathbf{s}_j^{(i)}\right\| \geq \left\|\mathbf{s}_{j'}^{(i)}\right\|$. Then, let $\left\|\mathbf{s}_j^{(i)}\right\| = \alpha \left\|\mathbf{s}_j^{(i)}\right\|$ for $\alpha \in [0, 1]$. By Lemma C.9, it holds that

$$\left\|\mathbf{s}_j^{(i)}\right\| \geq \frac{1}{\min\left\{\left|\mathcal{S}_j^{(i)}\right|, \left|\mathcal{S}_{j'}^{(i)}\right|\right\}} \left(1 - \frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

By using the proof of Lemma C.11 and the fact that $x + y \leq \sqrt{2}$ if $x^2 + y^2 = 1$, we have

$$\left\langle \frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|}, \frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|} \right\rangle \leq \sqrt{2} - \frac{1}{2} \left\| \frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|} - \frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|} \right\| \leq \frac{\sqrt{2}}{2} + \frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)},$$

Then,

$$\begin{aligned}\left\|\mathbf{s}_j^{(i)}-\mathbf{s}_{j'}^{(i)}\right\|^2 &= \left\|\mathbf{s}_j^{(i)}\right\|^2+\left\|\mathbf{s}_{j'}^{(i)}\right\|^2-2\left\langle\frac{\mathbf{s}_j^{(i)}}{\left\|\mathbf{s}_j^{(i)}\right\|},\frac{\mathbf{s}_{j'}^{(i)}}{\left\|\mathbf{s}_{j'}^{(i)}\right\|}\right\rangle\left\|\mathbf{s}_j^{(i)}\right\|\left\|\mathbf{s}_{j'}^{(i)}\right\| \\ &\geq (1+\alpha)\left\|\mathbf{s}_j^{(i)}\right\|^2-\left(\sqrt{2}+\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\alpha\left\|\mathbf{s}_j^{(i)}\right\|^2 \\ &\geq \left(1-(\sqrt{2}-1)\alpha-\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\left\|\mathbf{s}_j^{(i)}\right\|^2 \\ &\geq \frac{1}{\min\left\{\left|\mathcal{S}_j^{(i)}\right|,\left|\mathcal{S}_{j'}^{(i)}\right|\right\}}\left(\frac{1}{2}-\frac{7}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right)\left(1-\frac{1}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right) \\ &\geq \frac{1}{\min\left\{\left|\mathcal{S}_j^{(i)}\right|,\left|\mathcal{S}_{j'}^{(i)}\right|\right\}}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right),\end{aligned}$$

which proves the statement. □

After introducing the above Lemma, we analyze the total $(k-1)$-means cost. Let $\{\mathcal{Q}_r\}_{r=1}^{k-1}$ be the optimal $(k-1)$ clustering on $\mathbf{H}$, and $\mathbf{q}_r\in\mathbb{R}^t$ the optimal cluster center for cluster $\mathcal{Q}_r$. We prove that the resultant total $(k-1)$-means cost of the embedding points can be lower bounded as follows:

Lemma C.13. *It holds that*

$$\sum_{r=1}^{k-1}\sum_{u\in\mathcal{Q}_r}\|F(u)-\mathbf{q}_r\|^2\geq\sum_{i=1}^{\ell}\left(\frac{k_i^2}{16}-\frac{2k_i^2}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right).$$

Proof. Let $\{\mathcal{Z}_r\}_{r=1}^{k-1}$ be an arbitrary $(k-1)$ clustering on $\mathbf{H}$, and the vector $\mathbf{z}_r\in\mathbb{R}^t$ be the cluster center for every cluster $\mathcal{Z}_r$. We partition the k_i optimal clusters determined by $\mathbf{G}^{(i)}$ into $\{\mathcal{Z}_r\}_{r=1}^{k-1}$. We denote the $\mathcal{B}_{rj}^{(i)}=\mathcal{Z}_r\cap\mathcal{S}_j^{(i)}$ as a subset, in which the vertices are mapped from $\mathcal{S}_j^{(i)}$ to $\mathcal{Z}_r$. We decompose $\mathbf{z}_r$ into $\{\mathbf{z}_r^{(i)}\}_{i=1}^{\ell}$, and $\mathbf{z}_r^{(i)}\in\mathbb{R}^{k_i}$ is the cluster center for $\mathcal{Z}_r$ on $\mathbf{G}^{(i)}$. We have that

$$\sum_{u\in\mathcal{B}_{rj}^{(i)}}\left\|F^{(i)}(u)-\mathbf{z}_r^{(i)}\right\|^2\geq\sum_{u\in\mathcal{B}_{rj}^{(i)}}\frac{\left\|\mathbf{s}_j^{(i)}-\mathbf{z}_r^{(i)}\right\|^2}{2}-\sum_{u\in\mathcal{B}_{rj}^{(i)}}\left\|F^{(i)}(u)-\mathbf{s}_j^{(i)}\right\|^2,$$

where the inequality holds by $\|\mathbf{a}-\mathbf{b}\|^2\geq\|\mathbf{b}-\mathbf{c}\|^2/2-\|\mathbf{a}-\mathbf{c}\|^2$. By taking subsets $\{\mathcal{B}_{rj}^{(i)}\}_{j=1}^{k_i}$ into consideration, we have

$$\sum_{j=1}^{k_i}\sum_{u\in\mathcal{B}_{rj}^{(i)}}\left\|F^{(i)}(u)-\mathbf{z}_r^{(i)}\right\|^2\geq\sum_{j=1}^{k_i}\left|\mathcal{B}_{rj}^{(i)}\right|\frac{\left\|\mathbf{s}_j^{(i)}-\mathbf{z}_r^{(i)}\right\|^2}{2}-\sum_{j=1}^{k_i}\sum_{u\in\mathcal{B}_{rj}^{(i)}}\left\|F^{(i)}(u)-\mathbf{s}_j^{(i)}\right\|^2.$$

Notice that we can order the distance $\left\|\mathbf{s}_j^{(i)}-\mathbf{z}_r^{(i)}\right\|^2$ for $j\in[k_i]$ from largest to lowest, and denote the first and second lowest distances as $\left\|\mathbf{s}_x^{(i)}-\mathbf{z}_r^{(i)}\right\|^2$ and $\left\|\mathbf{s}_y^{(i)}-\mathbf{z}_r^{(i)}\right\|^2$. By applying the triangle inequality and AM-GM inequality, we have

$$\begin{aligned}\sum_{j=1}^{k_i}\left\|\mathbf{s}_j^{(i)}-\mathbf{z}_r^{(i)}\right\|^2 &\geq \frac{k_i}{2}\cdot\left(\left\|\mathbf{s}_x^{(i)}-\mathbf{z}_r^{(i)}\right\|^2+\left\|\mathbf{s}_y^{(i)}-\mathbf{z}_r^{(i)}\right\|^2\right) \\ &\geq \frac{k_i}{4}\cdot\left\|\mathbf{s}_x^{(i)}-\mathbf{s}_y^{(i)}\right\|^2 \\ &\geq \frac{k_i}{4}\cdot\frac{1}{\min\left\{\left|\mathcal{S}_x^{(i)}\right|,\left|\mathcal{S}_y^{(i)}\right|\right\}}\left(\frac{1}{2}-\frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)}\right),\end{aligned}$$

where the third inequality holds by Lemma C.12. By the assumption that the clusters in every $\mathbf{G}^{(i)}$ is almost balanced in size, i.e., $\left|\mathcal{S}_j^{(i)}\right| \in (n/(2k_i), (2n)/k_i)$ holds for every $i \in [\ell]$ and $j \in [k_i]$, we have

$$\sum_{j=1}^{k_i} \left\| \mathbf{s}_j^{(i)} - \mathbf{z}_r^{(i)} \right\|^2 \geq \frac{k_i^2}{8n} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right).$$

Thus, we have

$$\sum_{j=1}^{k_i} \sum_{u \in \mathcal{B}_{rj}^{(i)}} \left\| F^{(i)}(u) - \mathbf{z}_r^{(i)} \right\|^2 \geq \sum_{j=1}^{k_i} \left| \mathcal{B}_{rj}^{(i)} \right| \frac{k_i^2}{8n} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right) - \sum_{j=1}^{k_i} \sum_{u \in \mathcal{B}_{rj}^{(i)}} \left\| F^{(i)}(u) - \mathbf{s}_j^{(i)} \right\|^2.$$

By inductively performing above analysis on the cases $\{\mathcal{B}_{rj}^{(i)}\}_{j=1}^{k_i}$ for $r \in [k-1]$ and summarizing together, we have

$$\begin{aligned}
\sum_{r=1}^{k-1} \sum_{u \in \mathcal{Z}_r} \left\| F^{(i)}(u) - \mathbf{z}_r^{(i)} \right\|^2 &= \sum_{r=1}^{k-1} \sum_{j=1}^{k_i} \sum_{u \in \mathcal{B}_{rj}^{(i)}} \left\| F^{(i)}(u) - \mathbf{z}_r^{(i)} \right\|^2 \\
&\geq \sum_{r=1}^{k-1} \sum_{j=1}^{k_i} \left| \mathcal{B}_{rj}^{(i)} \right| \frac{k_i^2}{8n} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right) - \sum_{r=1}^{k-1} \sum_{j=1}^{k_i} \sum_{u \in \mathcal{B}_{rj}^{(i)}} \left\| F^{(i)}(u) - \mathbf{s}_j^{(i)} \right\|^2 \\
&\geq \frac{k_i^2}{8} \left(\frac{1}{2} - \frac{8}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right) - \sum_{j=1}^{k_i} \sum_{u \in \mathcal{S}_j^{(i)}} \left\| F^{(i)}(u) - \mathbf{s}_j^{(i)} \right\|^2.
\end{aligned}$$

By Lemma C.8, we have

$$\sum_{r=1}^{k-1} \sum_{u \in \mathcal{Z}_r} \left\| F(u) - \mathbf{z}_r \right\|^2 = \sum_{i=1}^{\ell} \sum_{r=1}^{k-1} \sum_{u \in \mathcal{Z}_r} \left\| F^{(i)}(u) - \mathbf{z}_r^{(i)} \right\|^2 \geq \sum_{i=1}^{\ell} \left(\frac{k_i^2}{16} - \frac{2k_i^2}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right).$$

It should be noticed that the output of optimal $(k-1)$ clustering is one of the cases that arbitrarily partitions $\mathbf{H}$ into $(k-1)$ clusters, and has the lowest $(k-1)$-means cost among all possible cases. Thus, we have

$$\sum_{r=1}^{k-1} \sum_{u \in \mathcal{Q}_r} \left\| F(u) - \mathbf{q}_r \right\|^2 \geq \sum_{i=1}^{\ell} \left(\frac{k_i^2}{16} - \frac{2k_i^2}{\Upsilon_{\mathbf{G}^{(i)}}(k_i)} \right),$$

which proves the statement. □

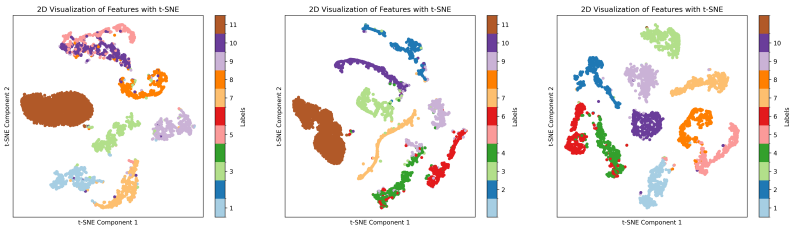
D Omitted Details from Section 6

D.1 Omitted Details for t-SNE Visualization Results

To further qualitatively demonstrate the effectiveness of our approach, we visualize both the modality-specific spectral embedding and modality-joint spectral embedding (used in our algorithm) with t-SNE (Maaten & Hinton, 2008) in Figures 4, 5, and 6. The results show that applying every single modality alone fails to identify the full set of ground-truth clusters present in the multi-modal data, exhibiting blurred cluster boundaries. In contrast, our algorithm reveals more compact and distinct cluster structure. This observation is consistent with our theoretical analysis.

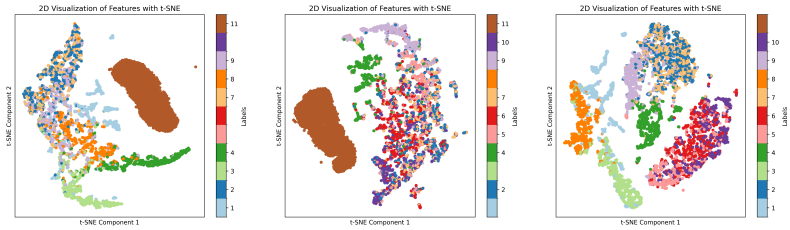
D.2 Omitted Details for Multi-modal Datasets

In this section, we present the details of three widely used multi-modal datasets and their specific configurations in our experiments:



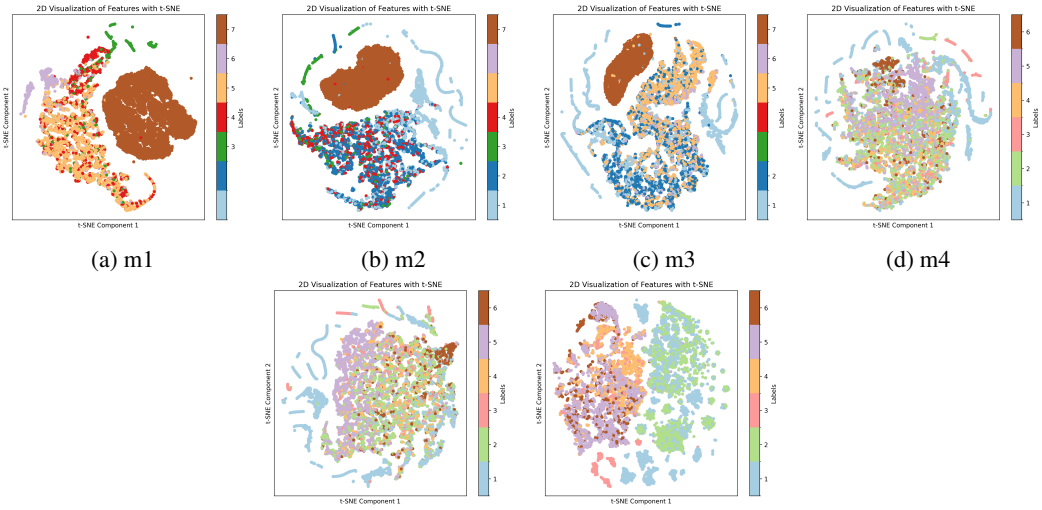
(a) m1 (b) m2 (c) joint

Figure 4: Comparison of t-SNE visualization results on the MNIST-USPS dataset. We use m1, m2, and m3 to represent three different modalities, and use "joint" to represent our algorithm. In our settings, modality m1 merges original clusters {6, 4, 2} into a single cluster 11, m2 combines clusters {5, 1, 8} as cluster 11.



(a) m1 (b) m2 (c) joint

Figure 5: Comparison of t-SNE visualization results on the nuswide dataset. We use m1, m2, and m3 to represent three different modalities, and use "joint" to represent our algorithm. In our settings, the {6, 10, 5} and {3, 8, 1} clusters are respectively merged into single clusters in the first and second modalities as cluster 11.



(a) m1 (b) m2 (c) m3 (d) m4

(e) m5 (f) joint

Figure 6: Comparison of t-SNE visualization results on the Reuters dataset. We use m1 − m5 to represent five different modalities, and use "joint" to represent our algorithm. In our settings, modality m1 merges original clusters {1, 2} into a single cluster, m2 combines clusters {5, 6} as a cluster, and m3 aggregates clusters {3, 4} as a cluster.

Table 2: Details of the three datasets

Datasets	# samples	# modalities	# modality-specific clusters	# modality-specific features	# total clusters
MNIST-USPS	5,000	2	8; 8	784; 784	10
nuswide	9,000	2	8; 8	4,096; 300	10
Reuters	18,758	5	5; 5; 5; 6; 6	300; 300; 300; 300; 300	6

- MNIST-USPS (Peng et al., 2019): This dataset combines handwritten digits from The MNIST [4] and USPS [5], treating them as two distinct modalities of different styles. It contains 5,000 samples across 10 categories, each represented by 784-dimensional features. The {6, 4, 2} clusters are merged in the first modality, while the {5, 1, 8} clusters are consolidated in the second.
- nuswide (Zhen et al., 2019): This collection comprises 9,000 image-tag pairs from 10 categories. Samples consist of 4,096-dimensional deep visual features and 300-dimensional text features. The {6, 10, 5} and {3, 8, 1} clusters are respectively merged into single clusters in the first and second modalities.
- Reuters (Amini et al., 2009): It is a large-scale multilingual dataset containing news articles in five languages (English, French, German, Italian, and Spanish). To further reduce the input dimensions, we preprocess the Reuters by selecting the top 300 columns with the highest variance, turning all input dimensions to 300. In our settings, modality m1 merges original clusters {1, 2} into a single cluster, m2 combines clusters {5, 6} as a cluster, and m3 aggregates clusters {3, 4} as a cluster.

We summarized the statistical information for the above three datasets in Table 2

[4] http://yann.lecun.com/exdb/mnist

[5] http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html