Divide and Conquer: Coordinating Multiplex Mixture of Graph Learners to Handle Multi-Omics Analysis

Zhihao Wu[1] , **Jielong Lu**[1] , **Jiajun Yu**[1] , **Sheng Zhou**[1] , **Yueyang Pi**[2] and **Haishuai Wang**[1*]
[1]Zhejiang Key Laboratory of Accessible Perception and Intelligent Systems,
College of Computer Science and Technology, Zhejiang University
[2]College of Computer and Data Science, Fuzhou University
zhihaowu1999@gmail.com, jielonglu2022@163.com, jiajunyu1999@gmail.com,
piyueyangcc@163.com, {zhousheng_zju, haishuai.wang}@zju.edu.cn

Abstract

Graph learning has shown significant advantages in organizing and leveraging complex data, making it promising for numerous real-world applications with heterogeneous information, particularly multi-omics data analysis. Despite its potential in such scenarios, existing methods are still in their infancy, lacking architectural potential and struggling to handle such complex data. In this paper, we propose the Multiplex Mixture of Graph Learners (MMoG) framework. MMoG first conducts fine-grained processing of consensus and unique information, constructing consistent features and multiplex graph structures. Then, a macroscopically shared group of sub-GNNs with diverse orders and architectures synergistically learn representations, providing a foundation for strong interaction between different views. Inspired by the mixture of experts (MoE), each sample in different omics adaptively determines the neighborhood ranges and architectures for information aggregation, while blocking unsuitable sub-GNNs. MMoG treats the complex multi-omics analysis as a multi-view learning problem, and essentially decomposes it into multiple sub-problems, allowing each omics/view to solve intersecting yet unique sub-problem groups. Additionally, we introduce mutual information-driven orthogonal loss and balancing loss to avoid view collapse. Extensive experiments on multi-omics data across multiple cancer types highlight MMoG's superiority.

1 Introduction

The proliferation of heterogeneous data has become a pervasive characteristic of many real-world applications [Wang *et al.*, 2021b; Zhou *et al.*, 2025]. These data often contain multiplex information collected from diverse analytical techniques, modalities, or perspectives, each providing a unique yet potentially consensual description of the underlying phenomena [Wang *et al.*, 2021a; Zhang *et al.*, 2025c]. This characteristic gives rise to a powerful paradigm–multi-view learning [Wan *et al.*, 2023], which models multiplex information as several views. By effectively leveraging complementary and shared information across views, multi-view learning has shown significant promise in integrating and analyzing such heterogeneous data. A representative example of multi-view learning is found in biological research [Zhu *et al.*, 2024], particularly in the study of cancer. Cancer is an inherently complex and heterogeneous disease, with significant molecular and phenotypic diversity both between and within tumors. This variability, driven by genetic, epigenetic, and proteomic alterations, complicates the accurate classification of cancer subtypes, which is crucial for personalized treatment strategies but still challenging. Recently, advancements in medical technologies like high-throughput sequencing and mass spectrometry have made it possible to integrate different layers of biological data. In cancer research, genome-wide multi-omics data, including transcriptomics (e.g., mRNA expression), epigenomics (e.g., DNA methylation), and genomics (e.g., copy number variation) etc., have emerged as powerful tools for addressing the complexity of tumor biology. As illustrated in Figure 1(a), large-scale projects such as The Cancer Genome Atlas (TCGA) have systematically collected multi-omics data across various cancer types. These diverse layers of biological information provide a more comprehensive and fine-grained view of the molecular landscape of cancer, enabling integrative analyses for subtype classification, biomarker discovery, and personalized therapy.

Multi-omics data captures more complete and nuanced biological information than typical data processed by multi-view learning. Effectively integrating this information from different omics is crucial for a comprehensive understanding of cancer subtypes. The more general form of this problem has been systematically discussed in multi-view learning, known as the principles of consistency and complementarity. For instance, Yang *et al.*; Sun *et al.*; Qin *et al.* boosted cross-view consistent and complementary information via different manners. However, these methods may struggle to handle cases with high view-heterogeneity, which is precisely a prominent characteristic of multi-omics data arising from varying measurement techniques and biological processes. Graphs hold promise for bridging this gap, as they are innately capable of modeling data from arbitrary distributions, thus aligning views with large discrepancies. Therefore, graph-based multi-view learning approaches [Li *et al.*, 2023;

*Corresponding author.

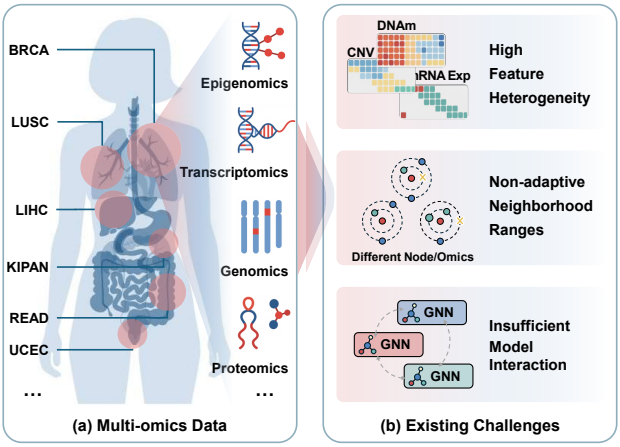

Figure 1: (a) Multi-omics data across multiple cancer types; (b) Challenges for graph-based multi-omics analysis model.

Zhang *et al.*, 2025a] have gained widespread adoption. Despite these advancements, they rely on classical shallow models, which may not perform well in complex multi-view scenarios like multi-omics cancer subtype inference.

Beyond general multi-view learning tasks, multi-omics data presents unique challenges due to its complexity. Due to variations in data scales, measurement techniques, and biological processes, multi-omics data suffer from erratic sample sizes as well as high-dimensional, multi-scale, and inherently heterogeneous features that are difficult to process using the traditional graph-based model, leading to a shift to Graph Neural Networks (GNNs). Recent works have attempted to apply GNNs to establish new frameworks. Chen *et al.*; Lu *et al.* explored using different GNN architectures. In particular, Wu *et al.*; Wang *et al.* specifically focused on multi-omics data. However, the above characteristics of multi-omics data and limitations of existing methods pose critical challenges: 1) Pronounced feature complexity and heterogeneity require greater model capacity, but architectures in hand offer limited representational power due to their design; 2) Topological differences across views and samples demand flexible neighborhoods, which current architectures lack; 3) Existing methods typically configure independent GNNs for each view, integrating with pre- or post-fusion, which architecturally limits intrinsic cross-view interactions. These challenges highlight the architectural limitations of current multi-view or multi-omics models, which remain relatively coarse for handling such complex problems. This prompts the question: *Can we design a more fine-grained graph learning architecture for such complex problems?*

Motivated by this question, we propose a novel framework based on the *divide-and-conquer* principle, termed Multiplex Mixture of Graph Learners (MMoG). First, the highly heterogeneous raw features are treated as unique information, and view-specific projection modules map them into same-dimensional latent spaces. The graph structures, modeling data from disparate distributions, are shared across all views. Each view adaptively integrates the graph structure to construct its multiplex graph. Through finer-grained management of shared and unique information, these establish a solid foundation for intrinsic interactions between views. As the core of MMoG, multiplex graph learners receive the constructed features and multiplex graphs, and shape representations through a group of collaborative sub-GNNs with multiple orders and parameters. These sub-GNNs are shared across views as a whole, integrating multiple hypotheses to enhance capacity, while fundamentally ensuring strong inter-view interactions. Inspired by Mixture of Experts (MoE), the sparse gating mechanism allows each sample in each view to flexibly select the optimal combination of parameters and neighborhood ranges, while blocking inappropriate sub-GNNs. In essence, this manner subtly divides the complex task of multi-view learning into interdependent subproblems. To mitigate view collapse, we introduce a balancing loss and an orthogonal loss driven by mutual information to constrain the sub-GNNs. Finally, we treat each view as an expert with self-gating to form the final representation. The procedure of MMoG is depicted in Figure 2. In summary, we make the following contributions:

- Taking advantages of MoE, we design MMoG to overcome the key gaps in existing algorithms, i.e., the limited architectural potential in complex multi-view scenarios.
- MMoG allows samples to adaptively aggregate long- or short-range information on suitable sub-GNNs.
- Propose the mutual information-driven orthogonal loss and balancing loss to mitigate view collapse.
- Comprehensive experiments on multi-omics cancer datasets demonstrate the superiority of MMoG.

2 Related Work

2.1 Graph Neural Networks

GNNs emerged as a powerful framework for learning on graph-structured data [Liu *et al.*, 2024a; Yu *et al.*, 2024; Liu *et al.*, 2024b; Wu *et al.*, 2023b]. They propagate information across graphs using message passing, allowing each node to update its representation by incorporating neighborhood information. The famous GNN was proposed by Kipf and Welling in the form of spectral-based GNNs. Over time, GNNs have evolved with the development of various architectures [Xu *et al.*, 2019; Veličković *et al.*, 2018; Zhuang *et al.*, 2025; Zheng *et al.*, 2022]. Recently, several studies [Li *et al.*, 2023; Wu *et al.*, 2023a] integrated knowledge from multiple relationships by combining several GNNs. As a powerful integration paradigm, MoE has shown promise in large language models [Shazeer *et al.*, 2017], but its application to GNNs remains underexplored. Few studies have addressed single-view problems with MoE [Wang *et al.*, 2023; Ma *et al.*, 2024].

2.2 Graph-based Multi-view Learning

Graph-based multi-view learning aims to leverage the various advantages of graphs to perform complex multi-view learning. Early works [Li *et al.*, 2025; Zhuang *et al.*, 2024; Wen *et al.*, 2023] were built on traditional optimization-based models, such as spectral clustering, and primarily focused

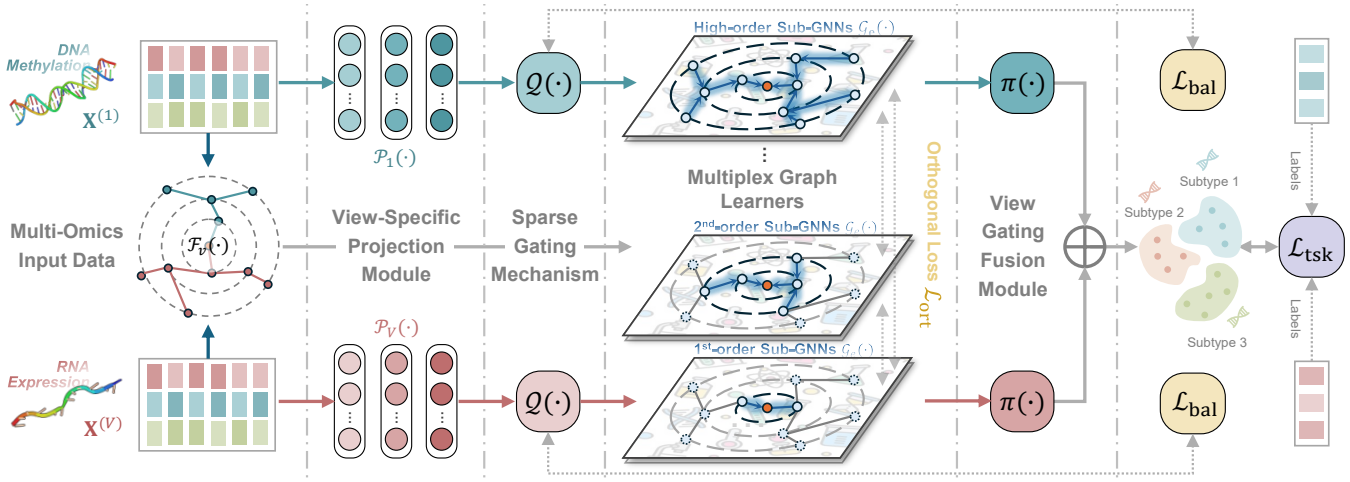

Figure 2: The illustration of the proposed MMoG framework.

on unsupervised learning tasks. In recent years, there has been a shift towards GNNs [Lin *et al.*, 2024; Chen *et al.*, 2024], which provide a deeper understanding of graphs and can effectively propagate label information. This has sparked the growth of semi-supervised learning approaches, where GNNs can enhance label propagation across views. Given the challenges posed by complex multi-view scenarios, such as multi-omics data, some methods have been specifically designed to address multi-omics integration [Wu *et al.*, 2024a; Xiao *et al.*, 2025]. However, these approaches typically treat the GNN as a monolithic module. A finer-grained model remains an area for further exploration.

2.3 Multi-Omics Analysis

Multi-omics analysis integrates data from multiple molecular layers, such as genomics, transcriptomics, epigenomics, and proteomic, providing a holistic view of biological processes and diseases. In cancer research, multi-omics integration has become crucial for accurately classifying cancer subtypes, identifying biomarkers, and understanding tumor heterogeneity. Studies like The Cancer Genome Atlas (TCGA) pan-can analysis [Weinstein *et al.*, 2013] have demonstrated that integrating multi-omics genomic data with clinical information improves the accuracy of cancer subtype prediction and patient stratification. For example, [Vasaikar *et al.*, 2018; Berg *et al.*, 2017; Zhang *et al.*, 2025b] demonstrated how integrating various omic layers, such as genomics, transcriptomics, and proteomics, can reveal molecular relationships and enhance our understanding of cancer biology. With the advancements, GNN-based methods have gained traction for multi-omics cancer studies [Schulte-Sasse *et al.*, 2021; Wang *et al.*, 2021d; Xiao *et al.*, 2023].

3 Proposed Method

3.1 Preliminary

Let $\mathcal{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_V\}$ be the multi-view or multi-omics input set, where V is the number of views and $\boldsymbol{X}_v \in \mathbb{R}^{n \times m_v}$ is the v-th input feature. Note that n denotes the number of samples and m_v is the feature dimension for $v \in [V]$. By constructing graphs over all the views based on certain similarities, we have the graph structures formed as adjacency matrices $\mathcal{A} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_V\}$, where $\boldsymbol{A}_v \in \{0,1\}^{n \times n}$. Only the undirected graph is discussed, i.e., $\boldsymbol{A}^v_{ij} = 1$ when there exists a link between sample i and j in the v-th view and $\boldsymbol{A}^v_{ij} = 0$ otherwise. Let $\boldsymbol{Y}_{i,:} = \{0,1\}^c$ be the one-hot label vector of sample i and $\mathcal{Y} = \{\boldsymbol{Y}_{i,:} : \forall i \in \mathcal{L}\}$ denotes all target labels, where $\mathcal{L}$ is the index set of labeled samples. For classification tasks, our goal is to learn a model from $\mathcal{X}$, $\mathcal{A}$, and $\mathcal{Y}$ to predict unlabeled samples.

3.2 Multiplex Mixture of Graph Learners

For each view in multi-view data, there are a couple of inputs $\boldsymbol{X}_v$ and $\boldsymbol{A}_v$. We then define distinct processing for each of them. First, due to the complexity of scenarios such as multi-omics analysis, the features of each view exhibit high and varying dimensionalities, carrying semantically significant differences. Therefore, we consider a group of view-specific encoders to project the features into a latent space sharing the same dimensionality, which provides the foundation for subsequent cross-view interactions, that is

$$\boldsymbol{Z}_v = \mathcal{P}_v\left(\boldsymbol{X}_v; \phi_v\right), \tag{1}$$

where $\mathcal{P}_v : \mathbb{R}^{m_v} \to \mathbb{R}^m$ denotes the projection module parameterized by ϕ_v and $\boldsymbol{Z}_v \in \mathbb{R}^{n \times m}$ is the learned features with m shared dimensions. Due to the properties of graphs that model arbitrarily distributed features, graph structures can be easily shared across views. For convenience, we denote $\phi = \{\phi_v : v \in [V]\}$. But within each view, a trade-off is required for view-wise graph structures. For this, we calculate the view-specific multiplex graph

$$\boldsymbol{S}_v = \mathcal{F}_v\left(\mathcal{A}; \psi_v\right), \tag{2}$$

where $\mathcal{F}_v : \mathcal{A} \to \mathbb{R}^{n \times n}$ is a fusion module with trainable parameters ψ_v and $\boldsymbol{S}_v \in \mathbb{R}^{n \times n}$ is the learned multiplex graph structure. For example, $\mathcal{F}_v$ can simply be a weighted sum on structures from all views. Similarly, we denote $\psi = \{\psi_v : v \in [V]\}$. Note that, in terms of graph theory,

the fused multiplex graph may still aggregate conflicting contents, such as heterophilic and homophilic edges, or graphs with large spectral differences. This issue can be addressed in the subsequent multiplex graph learners consisting of tailored sub-GNN models, such as those based on adaptive filtering.

After the distinct initial processing of features and structures, we introduce GNNs as the further representation learner. Modern GNNs are built upon the message-passing schema, which recursively propagates information across the graph structure and finally encodes non-Euclidean graphs into low-dimensional embeddings in Euclidean space. Due to the high heterogeneity of multi-omics data, each view may adapt to completely different models. For example, the neighborhood ranges required for DNA methylation and mRNA expression vary significantly and lead to entirely different parameter updates. From the lens of uniqueness preserving, we first consider building a separate GNN for each view

$$\boldsymbol{H}_v = \mathcal{G}_v\left(\boldsymbol{S}_v, \boldsymbol{Z}_v; \mathcal{W}_v\right), \tag{3}$$

where $\mathcal{G}_v : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times c}$ denotes a GNN model with a set of learnable weights $\mathcal{W}_v$ and $\boldsymbol{H}_v \in \mathbb{R}^{n\times c}$ is the yielded graph embedding. Typically, each layer in GNNs can be summarized as a unified formula, which contains an aggregation with a combination step

$$\begin{aligned} \boldsymbol{m}_i^{(l)} &= \mathrm{AGGREGATE}^{(l)}\left(\left\{\boldsymbol{h}_j^{(l-1)} : j \in \mathcal{N}(i)\right\}\right), \\ \boldsymbol{h}_i^{(l)} &= \mathrm{COMBINE}^{(l)}\left(\boldsymbol{h}_i^{(l-1)}, \boldsymbol{m}_i^{(l)}\right), \end{aligned} \tag{4}$$

where $\boldsymbol{m}_i^{(l)}$ is the message from neighborhood and $\boldsymbol{h}_i^{(l)}$ is the representation of sample i at the l-th layer. Note that we only consider the single-view situation in Eqn (4) for brevity. The composition of messages, along with the aggregation and combination operations, is crucial and depends on the specific design of different GNNs. In addition to these, GNNs often include learnable weights for linear projections and nonlinear transformations through activation functions. It is evident that constructing an appropriate GNN for each view involves specifying the number of message-passing layers L and training the corresponding parameters. Nevertheless, in the context of multi-omics or other complex multi-view data, limited model capacity hinders effective parameter optimization, while determining the number of layers or orders requires substantial prior knowledge. Moreover, this paradigm essentially integrates the GNN into the multi-view learning framework in a simplistic manner, lacking the intrinsic interaction between views.

Therefore, the design of framework should be considered at a finer granularity, ensuring that while associations between views are preserved, it can also handle the heterogeneity of each view. We will discuss the design philosophy in the context of specific scenarios. Considering multi-view classification tasks, the process of learning representation in each view and contributing to the final prediction can be viewed as solving a complex problem. Since the learned representations from different views should reach a consensus for the classification, the problem-solving process across views should be seen as interrelated. However, due to the complexity of each problem, it is difficult to explicitly define this relationship. Therefore, we divide each complex problem into multiple sub-problems. In this way, the associations between views can be viewed as intersections or overlaps between the subproblems. Guided by this idea, we advocate for an ensemble of models, where each model is designed to solve a simpler problem. Then each view adaptively selects the appropriate combination to learn representations that possess both uniqueness and potential correlations. Motivated by this analysis, we formulate the framework as

$$\begin{aligned} &\mathcal{M}(\boldsymbol{S}_v, \boldsymbol{Z}_v; \theta, \mathcal{W})_{i,:} = \sum_{e\in[E]} \mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,e}\, \mathcal{G}_e\left(\boldsymbol{S}_v, \boldsymbol{Z}_v; \mathcal{W}\right)_{i,:} \\ &\text{subject to } \sum_{e\in[E]} \mathbb{I}\{\mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,e} \neq 0\} = K \quad \forall i \in [n], \end{aligned} \tag{5}$$

where we have $\mathcal{M} : \mathbb{R}^m \to \mathbb{R}^d$, E is the number of sub-GNNs, and $\mathcal{Q} : \mathbb{R}^m \to \mathbb{R}^E$ is the sparse gating function, which outputs vectors $\mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,:} \in \mathbb{R}^E$ for each sample i to decide the gating value of every sub-GNN $e \in [E]$. Note that $\mathcal{Q}$ satisfies the sparse constraint that $\|\mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,:}\|_0 = K$. For this, we adopt the top-K strategy, i.e.,

$$\mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,e} = \frac{\exp(\mathcal{H}(\boldsymbol{Z}_v; \theta)_{i,e})}{\sum_{j\in\mathcal{T}(\boldsymbol{Z}_v; K)_i} \exp(\mathcal{H}(\boldsymbol{Z}_v; \theta)_{i,j})} \tag{6}$$

when $e \in \mathcal{T}(\boldsymbol{Z}_v)_i$ and $\mathcal{Q}(\boldsymbol{Z}_v; \theta)_{i,e} = 0$ otherwise, where $\mathcal{H} : \mathbb{R}^m \to \mathbb{R}^E$ denotes the gating network that computes weights of each sub-GNNs, and $\mathcal{T}(\boldsymbol{Z}_v; K)_i \subseteq [E]$ selects the indices of top-K largest logits from $\mathcal{H}(\boldsymbol{Z}_v; \theta)_{i,:}$. For the sub-GNNs, we first define the maximum layers/orders, and then construct several GNNs with different parameters for each number of layers/orders. The message-passing rules for each GNN depend on the specific model chosen.

The design of MMoG is inspired by the idea of MoE, containing an elaborate gating mechanism and multiple sub-GNNs as experts. In short, the input $\boldsymbol{Z}_v$ of each view is first processed by a gating network $\mathcal{H}$, which maps the features of each sample to weights for each sub-GNN. The gating function then selects the top-K sub-GNNs for each sample based on the weights. Each sample is routed to sub-GNNs with appropriate parameters, where it obtains neighborhood information at required orders. Finally, the representations learned from the selected top-K modules are integrated according to their weights to produce the final representation for each sample. In essence, MMoG divides the complex problems of each view into simpler sub-problems, each handled by a dedicated expert, i.e., a sub-GNN. This divide-and-conquer strategy effectively addresses data complexity and heterogeneity. Moreover, the varying numbers of layers/orders of sub-GNNs enable highly flexible selection of the neighborhood ranges. Macroscopically, MMoG serves as a shared representation learner across all views, which means the interrelations between complex problems of views, thereby guaranteeing strong inter-view interactions and successfully realizing our envisioned goal. After processing all views, it is critical to synthesize the opinions across them. To this end, we propose a view-expert fusion module, which conceptualizes the outputs from each view as distinct experts.

$$\boldsymbol{H}_{i,:} = \sum_{v\in[V]} \pi(\boldsymbol{Z}_v; \varphi)_i\, \mathcal{M}(\boldsymbol{S}_v, \boldsymbol{Z}_v; \theta, \mathcal{W})_{i,:}, \tag{7}$$

where $\boldsymbol{H}$ is the fused representation, $\pi : \mathbb{R}^d \to \mathbb{R}$ is a self-gating network, parameterized with φ. Leveraging the shared π, each sample dynamically determines an optimal combination of representations from the various views.

3.3 Mitigating View Collapse

Despite its well-established architecture, MMoG still faces a significant risk—view collapse—arising from its view-shared nature. View collapse refers to the phenomenon where, during the process of representation learning, each view loses its unique semantics from the original input and only retains similar information. In the context of "problem division", it means that each view focuses solely on the overlapping sub-problems, leading to trivial solutions. This risk can arise from two factors. The gating mechanism introduces sparsity and results in faster inference speeds, but it may lead to an imbalanced training process across sub-GNNs. Therefore, we need to introduce a balancing loss. For view v, we define a random variable ω^v with observations

$$\omega_e^v = \sum_{i\in[n]} \mathcal{Q}(\boldsymbol{Z}_v;\theta)_{i,e}. \tag{8}$$

It actually represents the total weight obtained by a specific sub-GNN for the v-th view across all samples, representing its importance. We aim to let each expert be similarly important, and thus introduce the following balancing loss:

$$\mathcal{J}_{\text{bal}}(\phi,\psi,\theta) = \frac{1}{V}\sum_{v\in[V]} \text{CV}(\omega^v)^2, \tag{9}$$

where CV is the coefficient of variation. Apart from the imbalance introduced by the gating mechanism, the sub-GNNs themselves can also cause view collapse. This arises because the information included in representations from different sub-GNNs may exhibit substantial overlap, which can be seen as a degradation of MMoG. To mitigate this, we incorporate the Mutual Information (MI) and expect to minimize the MI between distinct sub-GNNs to foster diversity:

$$\underset{\phi,\psi,\mathcal{W}}{\text{minimize}}\ I(\boldsymbol{U}_e^v, \boldsymbol{U}_{e'}^v), \tag{10}$$

where $\boldsymbol{U}_e^v$ and $\boldsymbol{U}_{e'}^v$ denote representations generated by two different sub-GNNs $\mathcal{G}_e$ and $\mathcal{G}_{e'}$, and $I(\boldsymbol{U}_e^v, \boldsymbol{U}_{e'}^v)$ is the MI between them. Specifically,

$$I(\boldsymbol{U}_e^v, \boldsymbol{U}_{e'}^v) = \int\int p(\boldsymbol{u}_e^v, \boldsymbol{u}_{e'}^v) \log \frac{p(\boldsymbol{u}_e^v, \boldsymbol{u}_{e'}^v)}{p(\boldsymbol{u}_e^v)p(\boldsymbol{u}_{e'}^v)} d\boldsymbol{u}_e^v d\boldsymbol{u}_{e'}^v, \tag{11}$$

where $p : \mathbb{R}^d \to \mathbb{R}_+$ is the probability density function. Problem (10) implies that $p(\boldsymbol{u}_e^v, \boldsymbol{u}_{e'}^v)$ and $p(\boldsymbol{u}_e^v)p(\boldsymbol{u}_{e'}^v)$ should be as similar as possible for a small $\log \frac{p(\boldsymbol{u}_e^v, \boldsymbol{u}_{e'}^v)}{p(\boldsymbol{u}_e^v)p(\boldsymbol{u}_{e'}^v)}$, which means de-correlation between the two variables $\boldsymbol{u}_e^v$ and $\boldsymbol{u}_{e'}^v$. Inspired by [Zhang *et al.*, 2023], we formulate that by minimizing the covariance

$$\underset{\phi,\psi,\mathcal{W}}{\text{minimize}}\ \text{Cov}([\boldsymbol{U}_e^v]_{:,i}, [\boldsymbol{U}_{e'}^v]_{:,j}), \tag{12}$$

which drives features (columns) of $\boldsymbol{U}_e^v$ and $\boldsymbol{U}_e^v$ to be as independent as possible. Further, we have

$$\text{Cov}([\boldsymbol{U}_e^v]_{:,i}, [\boldsymbol{U}_{e'}^v]_{:,j}) = \frac{1}{n}\left[\boldsymbol{U}_e^{v\top}\boldsymbol{U}_{e'}^v\right]_{i,j} - \mu_i\mu'_j, \tag{13}$$

where $\mu_i = \frac{1}{n}\sum_{t=1}^n [\boldsymbol{U}_e^v]_{t,i}$ and $\mu'_j = \frac{1}{n}\sum_{t=1}^n [\boldsymbol{U}_{e'}^v]_{t,j}$ are the mean value of $[\boldsymbol{U}_e^v]_{:,i}$ and $[\boldsymbol{U}_{e'}^v]_{:,j}$ respectively. Let the mean values to be zero, and we only focus on the first term $[\boldsymbol{U}_e^{v\top}\boldsymbol{U}_{e'}^v]_{i,j}$ on the right side of Eqn (13). That can be achieved by an additional centering step. Considering all pairs (i,j), the aim is to let $\boldsymbol{U}_e^{v\top}\boldsymbol{U}_{e'}^v$ be a zero matrix:

$$\underset{\phi,\psi,\mathcal{W}}{\text{minimize}}\ \left\|\boldsymbol{U}_e^{v\top}\boldsymbol{U}_{e'}^v\right\|_0, \tag{14}$$

where the ℓ_0-norm is adopted. Since Problem (14) is NP hard, we relax the ℓ_0-norm to ℓ_1-norm and obtain the orthogonal loss between any two sub-GNNs

$$\mathcal{J}_{\text{ort}}(\phi,\psi,\mathcal{W}) = \sum_{v\in[V]}\sum_{e<e'} \left\|\boldsymbol{U}_e^{v\top}\boldsymbol{U}_{e'}^v\right\|_1. \tag{15}$$

In each view, we constrain the mutual information between different sub-GNNs by decorrelating any two representations. The combination of balancing loss $\mathcal{J}_{\text{bal}}$ and orthogonal loss $\mathcal{J}_{\text{ort}}$ ensures that the framework maintains a larger hypothesis space, thereby preventing view collapse. The detailed derivations of this subsection are deferred to the Appendix.

3.4 Training Detail

To perform multi-omics cancer subtype classification, we adopt the cross-entropy loss to minimize the divergence between predictions and ground-truth labels

$$\mathcal{J}_{\text{tsk}}(\phi,\psi,\theta,\mathcal{W},\varphi) = -\sum_{i\in\mathcal{L}}\sum_{j\in[c]} \boldsymbol{Y}_{i,j}\log\tilde{\boldsymbol{Y}}_{i,j}, \tag{16}$$

where we recall that $\mathcal{L}$ is the index set of labeled sample and c is the number of classes. $\tilde{\boldsymbol{Y}}$ is the obtained via performing the softmax on logtis $\boldsymbol{H}$. The toatal objecitve is to

$$\underset{\phi,\psi,\theta,\mathcal{W},\varphi}{\text{minimize}}\ \mathcal{J} := \mathcal{J}_{\text{tsk}} + \alpha\mathcal{J}_{\text{bal}} + \beta\mathcal{J}_{\text{ort}}, \tag{17}$$

where α and β are trade-off parameters.

We give the computational complexity of MMoG, that is $\mathcal{O}(nmC + nKLF^2 + nKLBF)$, where C is the maximum input dimension, B is the number of edges in the multiplex graph, and L and F are the maximum orders and hidden units in each sub-GNN. It shows that appropriately adding more sub-GNNs expands the model's capacity without imposing a significant computational burden. Details are in Appendix.

4 Experiment

In this section, we examine the performance of MMoG to answer four research questions, that is:

- RQ1: How does MMoG perform in the real-world multi-omics analysis task?
- RQ2: How does the balancing loss $\mathcal{J}_{\text{bal}}$ and orthogonal loss $\mathcal{J}_{\text{ort}}$ work?
- RQ3: How do hyperparameters α, β, Orders and K affect the performance of MMoG?

Metric	Meth/DS	BRCA	KIPAN	LGG	UCEC	CRC	GBMLGG	TCGA
ACC	SVM	0.702 ± 0.000	0.939 ± 0.000	0.600 ± 0.000	0.725 ± 0.000	0.763 ± 0.000	0.520 ± 0.000	0.680 ± 0.000
	RF	0.802 ± 0.000	0.939 ± 0.000	0.498 ± 0.000	0.725 ± 0.000	0.776 ± 0.000	0.546 ± 0.000	0.670 ± 0.000
	DeepMO	0.814 ± 0.021	0.941 ± 0.009	0.642 ± 0.031	0.802 ± 0.031	0.754 ± 0.047	0.535 ± 0.024	0.730 ± 0.069
	Moanna	0.775 ± 0.018	0.925 ± 0.009	0.613 ± 0.018	0.773 ± 0.030	0.726 ± 0.042	0.531 ± 0.049	0.587 ± 0.022
	MOGONET	0.716 ± 0.015	0.914 ± 0.005	0.623 ± 0.018	0.754 ± 0.019	0.759 ± 0.000	0.492 ± 0.015	0.421 ± 0.007
	MoGCN	0.736 ± 0.004	0.915 ± 0.004	0.511 ± 0.000	0.716 ± 0.000	0.779 ± 0.012	0.530 ± 0.063	0.739 ± 0.004
	MOSGAT	0.730 ± 0.032	0.911 ± 0.004	0.585 ± 0.013	0.821 ± 0.019	0.757 ± 0.011	0.496 ± 0.012	-
	Ours	**0.851 ± 0.014**	**0.952 ± 0.004**	**0.690 ± 0.013**	**0.847 ± 0.012**	**0.865 ± 0.026**	**0.579 ± 0.008**	**0.752 ± 0.011**
F1	SVM	0.436 ± 0.000	0.929 ± 0.000	0.600 ± 0.000	0.280 ± 0.000	0.433 ± 0.000	0.394 ± 0.000	0.614 ± 0.000
	RF	0.688 ± 0.000	0.928 ± 0.000	0.496 ± 0.000	0.287 ± 0.000	0.489 ± 0.000	0.470 ± 0.000	0.520 ± 0.000
	DeepMO	0.764 ± 0.049	0.925 ± 0.008	0.637 ± 0.038	0.538 ± 0.019	0.676 ± 0.052	0.511 ± 0.022	0.646 ± 0.069
	Moanna	0.699 ± 0.020	0.903 ± 0.017	0.608 ± 0.019	0.534 ± 0.020	0.688 ± 0.040	0.481 ± 0.030	0.517 ± 0.023
	MOGONET	0.589 ± 0.026	0.905 ± 0.005	0.618 ± 0.024	0.437 ± 0.006	0.432 ± 0.000	0.436 ± 0.027	0.385 ± 0.004
	MoGCN	0.553 ± 0.007	0.914 ± 0.004	0.338 ± 0.000	0.280 ± 0.000	0.677 ± 0.046	0.415 ± 0.084	0.667 ± 0.004
	MOSGAT	0.578 ± 0.053	0.908 ± 0.005	0.550 ± 0.029	0.523 ± 0.041	0.478 ± 0.031	0.467 ± 0.015	-
	Ours	**0.825 ± 0.014**	**0.940 ± 0.004**	**0.688 ± 0.014**	**0.552 ± 0.005**	**0.760 ± 0.072**	**0.541 ± 0.008**	**0.672 ± 0.011**

Table 1: Classification results (mean and standard deviation) of all comparative algorithms on the seven multi-omics cancer subtype datasets, with 10% labeled samples, where the best results are highlighted with **pink** and the second-best results are highlighted with **cyan**.

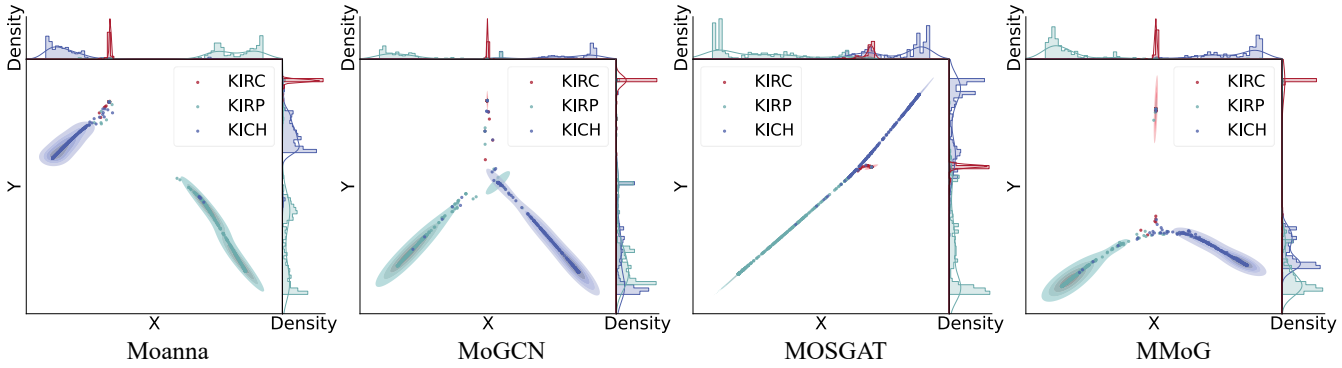

Figure 3: The visualization of the representations of Moanna, MoGCN, MOSGAT, and MMoG on the KIPAN dataset.

4.1 Experimental Setting

Datasets: We evaluate MMoG on 7 multi-omics datasets: BRCA, KIPAN, LGG, UCEC, CRC, GBMLGG, and TCGA, and we also use some classical multi-view datasets like Animals. We adopt a semi-supervised classification setting, where only 10% of the samples are used as the training set, while the remaining 90% are reserved for testing.

Compared Methods: For comparison, we employ GNN-free methods: SVM, RF, Moanna [Lupat *et al.*, 2023], DeepMO [Lin *et al.*, 2020], and GNN-based methods: MOGONET [Wang *et al.*, 2021c], MoGCN [Li *et al.*, 2022], and MOSGAT [Wu *et al.*, 2024b]. All the above methods are set as default following the original paper. To evaluate performance, we conduct the experiments on all datasets, reporting the mean values and standard deviations (10 runs). More experiments and setting details are deferred to Appendix.

4.2 Cancer Subtype Classification (RQ1)

In this subsection, we perform semi-supervised classification of multi-omics cancer subtypes and report the results w.r.t. the accuracy (ACC) and the macro-F1 score (F1) in Table 1. From the table, we derive that: Benefiting from its fine-grained design, MMoG consistently outperforms or closely matches the best performance across multiple datasets, particularly excelling in ACC and F1 metrics on BRCA, KIPAN, LGG, and TCGA datasets.

$\mathcal{L}_{\text{bal}}$	$\mathcal{L}_{\text{ort}}$	TCGA		Animals	
		ACC	F1	ACC	F1
✓	✓	0.752 ± 0.011	0.672 ± 0.011	0.851 ± 0.014	0.788 ± 0.005
	✓	0.744 ± 0.017	0.655 ± 0.017	0.838 ± 0.008	0.760 ± 0.008
✓		0.744 ± 0.012	0.655 ± 0.012	0.810 ± 0.028	0.753 ± 0.028
		0.705 ± 0.015	0.648 ± 0.015	0.796 ± 0.023	0.722 ± 0.023

Table 2: Ablation study of MMoG on datasets TCGA and Animals. ✓ indicates the corresponding loss is used.

4.3 Representation Visualization (RQ1)

To further validate the effectiveness of the proposed method, we visualized the representations generated by different approaches on the KIPAN dataset, as shown in Figure 3. The figure demonstrates that the representations obtained using

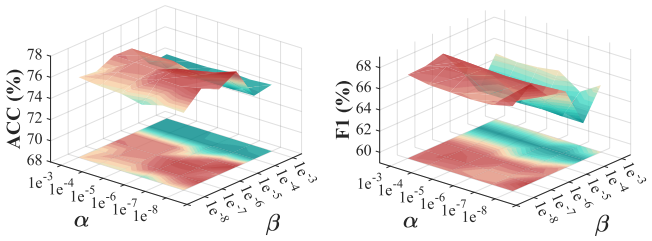

Figure 4: The classification performance of MMoG w.r.t. hyperparameters α and β on the TCGA and Animals dataset.

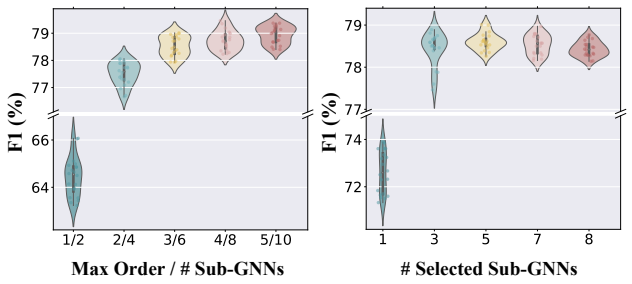

Figure 5: The classification performance of MMoG w.r.t. the max orders, the number of sub-GNNs and selected sub-GNNs.

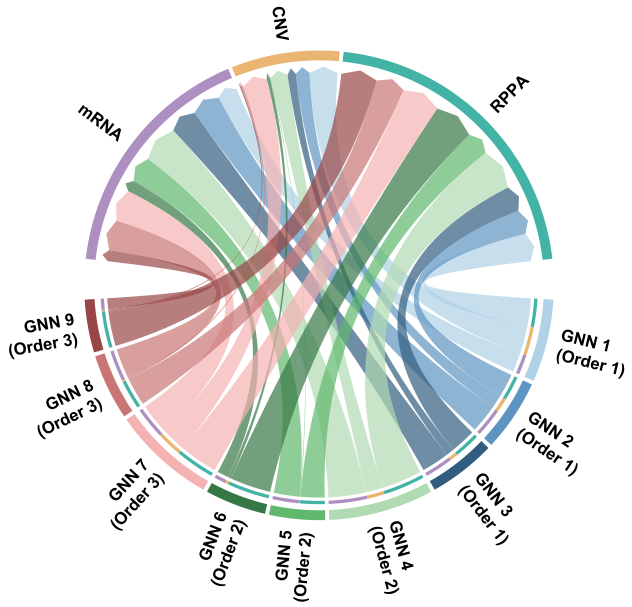

Figure 6: Visualization of omics-sub-GNN selection in MMoG on the BRCA dataset. Arrows indicate selection strengths.

the MMoG method exhibit clear and well-separated category distributions. In contrast, other methods exhibit substantial overlap between categories. For example, both Moanna and MOSGAT struggle to distinguish between KIRC and KIRP, while MoGCN performs better overall, but still shows pronounced overlap between KIRP and KICH. These results highlight MMoG's ability to capture inter-category feature differences better, thereby producing representations with stronger discriminative power. These results indicate that MMoG is capable of learning more discriminative representations, which may be attributed to the model capacity provided by the MoE architecture.

4.4 Ablation and Parameter Analysis (RQ3)

To validate the effectiveness of the proposed module, we conduct an ablation study by progressively removing the balancing loss $\mathcal{J}\mathrm{bal}$ and the orthogonal loss $\mathcal{J}\mathrm{ort}$. The results in Table 2 show that excluding both losses leads to the lowest performance, while including either loss improves results, and the best performance is achieved when both are used together. This underscores the complementary roles and necessity of both loss terms in our model. Additionally, we examine the sensitivity of the model to hyperparameters α and β (see Figure 4). Performance peaks when $\alpha = 1\mathrm{e}{-5}$ and β is set to $1\mathrm{e}{-6}$ or $1\mathrm{e}{-8}$. However, when β becomes too large, performance becomes unstable, suggesting that excessive orthogonality constraints may harm the learning process.

4.5 Expert Effect Analysis (RQ4)

To further investigate the MoE-based architecture, we fix the ratio of maximum order to the number of experts (sub-GNNs) and the ratio of selected samples, both at 1/2, and repeat each setting 20 times. As shown in Figure 5, performance improves and then plateaus as the maximum order and total number of experts increase. Similarly, with 10 experts, using only one expert yields poor performance, while increasing to three brings significant gains. These results indicate that using sub-GNNs with diverse orders and parameters notably enhances overall performance, while the number of selected experts can be sparse but should not be too low. Based on the statistics of expert selection for each sample on each view, Figure 6 further visualizes selection patterns across omics. Different omics prefer sub-GNNs with different parameters and orders, demonstrating that MMoG effectively supports fine-grained multi-omics analysis. This also suggests that the proposed loss functions regularize the model effectively.

5 Conclusion

In this paper, we proposed the MMoG framework to address the challenges of multi-omics and complex multi-view data. MMoG overcame the limitations of existing methods by enabling stronger collaboration between views through fine-grained processing, scale-consistent features, and multiplex graph structures. Inspired by the MoE approach, The framework leveraged a shared group of diverse GNNs, adapting neighborhood ranges and architectures for each sample in each view. We also introduced mutual information-driven orthogonal loss and balancing loss to prevent view collapse. Experimental results on multi-omics and typical multi-view datasets demonstrated MMoG's superiority in handling highly heterogeneous multi-view data and outperforming current methods. In summary, MMoG effectively enhanced graph-based multi-view learning, providing a robust solution to the complexities of multi-omics analysis and setting a new benchmark for future work. One possible limitation is that MMoG on data with both highly heterogeneous graph structures and features has not yet been explored.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62202422 and 62372408).

Contribution Statement

Zhihao Wu and Jielong Lu contributed equally to this work.

References

[Berg *et al.*, 2017] Kaja CG Berg, Peter W Eide, Ina A Eilertsen, Bjarne Johannessen, Jarle Bruun, Stine A Danielsen, Merete Bjørnslett, Leonardo A Meza-Zepeda, Mette Eknæs, Guro E Lind, et al. Multi-omics of 34 colorectal cancer cell lines-a resource for biomedical studies. *Molecular Cancer*, 16:1–16, 2017.

[Chen *et al.*, 2023] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95:109–119, 2023.

[Chen *et al.*, 2024] Man-Sheng Chen, Xi-Ran Zhu, Jia-Qi Lin, and Chang-Dong Wang. Contrastive multiview attribute graph clustering with adaptive encoders. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–12, 2024.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Li *et al.*, 2022] Xiao Li, Jie Ma, Ling Leng, Mingfei Han, Mansheng Li, Fuchu He, and Yunping Zhu. Mogcn: A multi-omics integration method based on graph convolutional network for cancer subtype analysis. *Frontiers in Genetics*, 13, 2022.

[Li *et al.*, 2023] Xingfeng Li, Yinghui Sun, Quansen Sun, Zhenwen Ren, and Yuan Sun. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100:101941, 2023.

[Li *et al.*, 2025] Xingfeng Li, Yuangang Pan, Yuan Sun, Yinghui Sun, Quansen Sun, Zhenwen Ren, and Ivor W Tsang. Scalable unpaired multi-view clustering with bipartite graph matching. *Information Fusion*, 116:102786, 2025.

[Lin *et al.*, 2020] Yuqi Lin, Wen Zhang, Huanshen Cao, Gaoyang Li, and Wei Du. Classifying breast cancer subtypes using deep neural networks based on multi-omics data. *Genes*, 11:888, 2020.

[Lin *et al.*, 2024] Jia-Qi Lin, Man-Sheng Chen, Xi-Ran Zhu, Chang-Dong Wang, and Haizhang Zhang. Dual information enhanced multiview attributed graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–12, 2024.

[Liu *et al.*, 2024a] Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. Arc: A generalist graph anomaly detector with in-context learning. In *NeurIPS*, 2024.

[Liu *et al.*, 2024b] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Siwei Wang, Ke Liang, Wenxuan Tu, and Liang Li. Simple contrastive graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13789–13800, 2024.

[Lu *et al.*, 2024] Jielong Lu, Zhihao Wu, Zhaoliang Chen, Zhiling Cai, and Shiping Wang. Towards multi-view consistent graph diffusion. In *ACM MM*, pages 186–195, 2024.

[Lupat *et al.*, 2023] Richard Lupat, Rashindrie Perera, Sherene Loi, and Jason Li. Moanna: multi-omics autoencoder-based neural network algorithm for predicting breast cancer subtypes. *IEEE Access*, 11:10912–10924, 2023.

[Ma *et al.*, 2024] Li Ma, Haoyu Han, Juanhui Li, Harry Shomer, Hui Liu, Xiaofeng Gao, and Jiliang Tang. Mixture of link predictors on graphs. In *NeurIPS*, 2024.

[Qin *et al.*, 2023] Yang Qin, Yuan Sun, Dezhong Peng, Joey Tianyi Zhou, Xi Peng, and Peng Hu. Cross-modal active complementary learning with self-refining correspondence. In *NeurIPS*, 2023.

[Schulte-Sasse *et al.*, 2021] Roman Schulte-Sasse, Stefan Budach, Denes Hnisz, and Annalisa Marsico. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nature Machine Intelligence*, 3(6):513–526, 2021.

[Shazeer *et al.*, 2017] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *ICLR*, 2017.

[Sun *et al.*, 2024] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):9150–9162, 2024.

[Vasaikar *et al.*, 2018] Suhas V Vasaikar, Peter Straub, Jing Wang, and Bing Zhang. Linkedomics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Research*, 46(D1):D956–D963, 2018.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

[Wan *et al.*, 2023] Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. Auto-weighted multi-view clustering for large-scale data. In *AAAI*, pages 10078–10086, 2023.

[Wang *et al.*, 2021a] Qianqian Wang, Jiafeng Cheng, Quanxue Gao, Guoshuai Zhao, and Licheng Jiao. Deep multi-view subspace clustering with unified and discriminative learning. *IEEE Transactions on Multimedia*, 23:3483–3493, 2021.

[Wang *et al.*, 2021b] Qianqian Wang, Huanhuan Lian, Gan Sun, Quanxue Gao, and Licheng Jiao. icmsc: Incomplete

cross-modal subspace clustering. *IEEE Transactions on Image Processing*, 30:305–317, 2021.

[Wang *et al.*, 2021c] Tongxin Wang, Wei Shao, Zhi Huang, Haixu Tang, Jie Zhang, Zhengming Ding, and Kun Huang. Mogonet integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature Communications*, 12:3445, 2021.

[Wang *et al.*, 2021d] Yi Wang, Zhongyue Zhang, Hua Chai, and Yuedong Yang. Multi-omics cancer prognosis analysis based on graph convolution network. In *IEEE International Conference on Bioinformatics and Biomedicine*, pages 1564–1568, 2021.

[Wang *et al.*, 2023] Haotao Wang, Ziyu Jiang, Yuning You, Yan Han, Gaowen Liu, Jayanth Srinivasa, Ramana Kompella, and Zhangyang "Atlas" Wang. Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling. In *NeurIPS*, pages 50825–50837, 2023.

[Weinstein *et al.*, 2013] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113–1120, 2013.

[Wen *et al.*, 2023] Yi Wen, Suyuan Liu, Xinhang Wan, Siwei Wang, Ke Liang, Xinwang Liu, Xihong Yang, and Pei Zhang. Efficient multi-view graph clustering with local and global structure preservation. In *ACM MM*, pages 3021–3030, 2023.

[Wu *et al.*, 2023a] Zhihao Wu, Xincan Lin, Zhenghong Lin, Zhaoliang Chen, Yang Bai, and Shiping Wang. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE Transactions on Multimedia*, pages 1–14, 2023.

[Wu *et al.*, 2023b] Zhihao Wu, Zhao Zhang, and Jicong Fan. Graph convolutional kernel machine versus graph convolutional networks. In *Advances in Neural Information Processing Systems*, volume 36, pages 19650–19672, 2023.

[Wu *et al.*, 2024a] Jiecheng Wu, Zhaoliang Chen, Shunxin Xiao, Genggeng Liu, Wenjie Wu, and Shiping Wang. Deepmoic: multi-omics data integration via deep graph convolutional networks for cancer subtype classification. *BMC Genomics*, 25(1):1–13, 2024.

[Wu *et al.*, 2024b] Wenhao Wu, Shudong Wang, Yuanyuan Zhang, Wenjing Yin, Yawu Zhao, and Shanchen Pang. Mosgat: Uniting specificity-aware gats and cross modal-attention to integrate multi-omics data for disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[Xiao *et al.*, 2023] Shunxin Xiao, Huibin Lin, Conghao Wang, Shiping Wang, and Jagath C Rajapakse. Graph neural networks with multiple prior knowledge for multi-omics data analysis. *IEEE Journal of Biomedical and Health Informatics*, 27(9):4591–4600, 2023.

[Xiao *et al.*, 2025] Yuang Xiao, Dong Yang, Jiaxin Li, Xin Zou, Hua Zhou, and Chang Tang. Dual alignment feature embedding network for multi-omics data clustering. *Knowledge-Based Systems*, 309:112774, 2025.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.

[Yang *et al.*, 2023] Xihong Yang, Jin Jiaqi, Siwei Wang, Ke Liang, Yue Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En Zhu. Dealmvc: Dual contrastive calibration for multi-view clustering. In *ACM MM*, pages 337–346, 2023.

[Yu *et al.*, 2024] Jiajun Yu, Zhihao Wu, Jinyu Cai, Adele Lu Jia, and Jicong Fan. Kernel readout for graph neural networks. In *IJCAI*, pages 2505–2514, 2024.

[Zhang *et al.*, 2023] Lei Zhang, Lele Fu, Tong Wang, Chuan Chen, and Chuanfu Zhang. Mutual information-driven multi-view clustering. In *CIKM*, pages 3268–3277, 2023.

[Zhang *et al.*, 2025a] Chao Zhang, Deng Xu, Chunlin Chen, and Huaxiong Li. Semi-supervised multi-view clustering with active constraints. In *SIGKDD*, pages 1903–1912, 2025.

[Zhang *et al.*, 2025b] Chao Zhang, Deng Xu, Chunlin Chen, Min Zhang, and Huaxiong Li. Multi-relational multi-view clustering and its applications in cancer subtype identification. *Information Fusion*, 117:102831, 2025.

[Zhang *et al.*, 2025c] Yunhe Zhang, Jinyu Cai, Zhihao Wu, Pengyang Wang, and See-Kiong Ng. Mixture of experts as representation learner for deep multi-view clustering. In *AAAI*, pages 22704–22713, 2025.

[Zheng *et al.*, 2022] Yizhen Zheng, Shirui Pan, Vincent CS Lee, Yu Zheng, and Philip S Yu. Rethinking and scaling up graph contrastive learning: An extremely efficient approach with group discrimination. In *NeurIPS*, pages 10809–10820, 2022.

[Zhou *et al.*, 2025] Sheng Zhou, Hongjia Xu, Zhuonan Zheng, Jiawei Chen, Zhao Li, Jiajun Bu, Jia Wu, Xin Wang, Wenwu Zhu, and Martin Ester. A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions. *ACM Computing Surveys*, 57(3):69:1–69:38, 2025.

[Zhu *et al.*, 2024] Yanran Zhu, Xiao He, Chang Tang, Xinwang Liu, Yuanyuan Liu, and Kunlun He. Multi-view adaptive fusion network for spatially resolved transcriptomics data clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8889–8900, 2024.

[Zhuang *et al.*, 2024] Shuman Zhuang, Sujia Huang, Wei Huang, Yuhong Chen, Zhihao Wu, and Ximeng Liu. Enhancing multi-view graph neural network with cross-view confluent message passing. In *ACM MM*, pages 10065–10074, 2024.

[Zhuang *et al.*, 2025] Shuman Zhuang, Zhihao Wu, Zhaoliang Chen, Hong-Ning Dai, and Ximeng Liu. Refine then classify: Robust graph neural networks with reliable neighborhood contrastive refinement. In Toby Walsh, Julie Shah, and Zico Kolter, editors, *AAAI*, pages 13473–13482, 2025.