# Advanced unsupervised learning: a comprehensive overview of multi-view clustering techniques

Abdelmalik Moujahid[1] · Fadi Dornaika[2,3]

## Abstract

Machine learning techniques face numerous challenges to achieve optimal performance. These include computational constraints, the limitations of single-view learning algorithms and the complexity of processing large datasets from different domains, sources or views. In this context, multi-view clustering (MVC), a class of unsupervised multi-view learning, emerges as a powerful approach to overcome these challenges. MVC compensates for the shortcomings of single-view methods and provides a richer data representation and effective solutions for a variety of unsupervised learning tasks. In contrast to traditional single-view approaches, the semantically rich nature of multi-view data increases its practical utility despite its inherent complexity. This survey makes a threefold contribution: (1) a systematic categorization of multi-view clustering methods into well-defined groups, including co-training, co-regularization, subspace, deep learning, kernel-based, anchor-based, and graph-based strategies; (2) an in-depth analysis of their respective strengths, weaknesses, and practical challenges, such as scalability and incomplete data; and (3) a forward-looking discussion of emerging trends, interdisciplinary applications, and future directions in MVC research. This study represents an extensive workload, encompassing the review of over 140 foundational and recent publications, the development of comparative insights on integration strategies such as early fusion, late fusion, and joint learning, and the structured investigation of practical use cases in the areas of healthcare, multimedia, and social network analysis. By integrating these efforts, this work aims to fill existing gaps in MVC research and provide actionable insights for the advancement of the field.

✉ Fadi Dornaika
fadi.dornaika@ehu.eus

Abdelmalik Moujahid
abdelmalik.moujahid@unir.net

[1] Universidad Internacional de La Rioja (UNIR), Logroño, Spain

[2] University of the Basque Country (UPV/EHU), San Sebastián, Spain

[3] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

# 1 Introduction

Multi-view learning focuses on using information from different sets of features or representations, called *views*, to improve learning performance. The basic idea is that different views of the same data provide unique and complementary insights, so it is beneficial to consider them simultaneously. The field addresses challenges such as dealing with missing or noisy views, ensuring alignment between views, and selecting appropriate integration techniques.

In multi-view clustering, the way in which the information from multiple views is integrated has a significant impact on the clustering results. Depending on when and how the views are combined, the methods for multi-view clustering can be broadly categorized into three primary integration strategies: *early fusion*, *late fusion* and *joint learning*. The three categories serve as an overarching framework for understanding multi-view clustering (MVC) approaches, offering a conceptual foundation from which various methodologies can be explored. While these categories provide a general overview, the subsequent sections of this paper delve into specific techniques and advancements within each category, examining how recent innovations and refinements have addressed the limitations and enhanced the performance of multi-view clustering methods across a variety of applications. This structure aims to provide readers with both a broad understanding and a detailed, in-depth perspective on the evolving landscape of multi-view clustering. The three primary categories are as follows:

- Early Fusion: Early fusion methods combine the features from multiple views at the input level by merging the feature representations from all views into a single, unified feature model (Liang et al. 2022, 2019). This approach allows the model to learn from all available data simultaneously, processing the combined information early in the learning process. One of the main advantages of early fusion is its computational efficiency, as it allows a single model to work with the combined feature set. However, early fusion assumes that all views are perfectly aligned and equally informative, which is often not the case in practice. In real-world applications, the views may be noisy or contain incomplete information, which can negatively impact the performance of early fusion methods. In addition, this approach cannot handle heterogeneity between views (e.g. different data modalities) as effectively as other strategies.
- Late Fusion: Late fusion approaches take a different strategy by training separate models for each view independently and combining their results - such as predictions or cluster assignments - at a later stage. In this way, each view can be processed individually, providing flexibility in dealing with heterogeneous data types or situations where some views might be missing (Li et al. 2024; Bruno and Marchand-Maillet 2009). For example, in a clustering problem with multiple views, the individual clustering results from each view could be merged into a final consensus clustering. While late fusion is highly adaptive and robust to incomplete or varying data types, it cannot capture complex dependencies and interactions between the different views. Since each model is trained independently, the relationships between views are not always well utilized, which can lead to suboptimal integration of the data.
- Joint Learning: Joint learning techniques aim to simultaneously learn a common representation or task-specific features for all views, often by mapping each view into a com-

mon latent space. These methods explicitly capture the interactions and complementarities between views during the learning process, promoting an integrated understanding of the data (Li et al. 2022; Huang et al. 2021). Joint learning approaches can be particularly powerful when the views are complementary, as they allow the model to exploit the full potential of the available information. By learning a common representation or model, these techniques improve the matching and integration of data from different views, which can lead to better performance than early or late fusion methods. However, joint learning is often more computationally intensive and requires more sophisticated optimization techniques and larger computational resources. It can also depend on the choice of model architecture, and the learning process can be more complex and prone to overfitting, especially when the data is noisy or sparse.

Each of these strategies has its own strengths and limitations, making them suitable for different application scenarios. For example, early fusion methods are effective when the views are well aligned and have similar characteristics, such as in multi-channel image processing. Late fusion is often used in scenarios with heterogeneous data types, such as in bioinformatics, where each view represents a different biological modality. Joint learning is particularly useful in applications such as sensor networks or multitask learning, where capturing interactions between different views is crucial.

This review systematically examines these strategies and provides a comparative analysis of the underlying principles, strengths and limitations of early fusion, late fusion and joint learning methods in multi-view clustering. The goal is to provide researchers with a comprehensive understanding of the multi-view clustering landscape and help them identify the most appropriate integration strategy for specific applications.

The emerging field of multi-view clustering techniques specifically aims to seamlessly integrate this heterogeneous data landscape and extract meaningful patterns from it. By considering multiple data perspectives, these techniques have the potential to outperform traditional single-view clustering approaches and lead to more accurate and robust cluster mappings, as highlighted in studies by Xu et al. (2015) and Kumar et al. (2011) Xin and Tian (2015); Kumar and Udupa (2011). Furthermore, the application of multi-view clustering goes beyond improved clustering accuracy. It enables a deeper understanding of complex data by revealing hidden structures, relationships and insights that may remain undetected when analyzing single perspectives in isolation, as Cai et al. Cai et al. (2011) emphasize. This synergy between multi-view learning and clustering techniques opens up avenues for improved data analysis and knowledge extraction in various real-world applications.

## 1.1 Addressing the importance of multi-view clustering

Multi-view clustering has become a central approach in the field of unsupervised learning, driven by the increasing prevalence of heterogeneous and multimodal datasets in modern applications. In contrast to single-view learning methods that rely on a single representation or perspective of the data, multi-view clustering utilizes multiple complementary views to provide a more holistic understanding of the underlying structures. This paradigm is particularly important for several reasons:

1. Overcoming limitations in data representation: Single-view methods often fail to capture the variety and richness of data coming from different sources or modalities. Multi-view clustering addresses this limitation by integrating complementary information from multiple views, resulting in a more robust and comprehensive data representation. In biomedical research, for example, combining genomic data (one view) with proteomic data (another view) leads to a more comprehensive understanding of biological processes and improves the clustering of disease subtypes.

2. Handling heterogeneous datasets: In real-world scenarios, datasets often comprise different domains, each of which provides unique insights. In multimedia analysis, for example, a dataset may contain text descriptions, audio recordings and visual features of videos. Multi-view clustering processes these heterogeneous modalities to effectively group similar multimedia content, outperforming single-view approaches that only consider one modality at a time.

3. Reducing bias through single-view dependency: Dependence on a single view can introduce bias and limit generalizability, as the selected view may not fully represent the variability of the data. Multi-view clustering synthesizes information across multiple views to mitigate this problem. For example, when analyzing social networks, a single-view approach may only consider user interactions (e.g. likes or shares), while multi-view clustering integrates other perspectives such as user profiles and content preferences, resulting in more accurate community detection.

By systematically addressing these constraints, multi-view clustering provides a transformative framework that meets the needs of modern machine learning tasks. Its ability to fuse complementary information, align different data modalities, and improve clustering performance highlights its crucial role in advancing unsupervised learning methods.

## 1.2 Challenges

Multi-view clustering addresses the challenge of heterogeneity in real-world datasets by integrating different types of information and perspectives into a unified framework (Xin and Tian 2015). By bridging the semantic gap between simple and high-level features, it enables the discovery of meaningful patterns, especially in multimedia data (Cai et al. 2011). However, despite its advantages, multi-view clustering faces several interrelated challenges that need to be overcome for effective application:

One key challenge in clustering with multiple views is the high-dimensional feature space that results from the integration of multiple views. While additional information improves clustering, it also introduces the "curse of dimensionality", which can degrade performance and increase computational complexity. Dimensionality reduction techniques, such as principal component analysis (PCA) (Greenacre et al. 2022) and t-SNE (van der Maaten and Hinton 2008), reduce the dimensionality of individual views. Advanced methods such as multi-view subspace learning (Xu et al. 2013) uncover common latent representations across all views and thus improve the efficiency of clustering.

To efficiently process large datasets, anchor-based methods provide a scalable approach by selecting a small set of representative data points, called anchors, to approximate the original dataset. These methods effectively reduce computational complexity while preserving the structural relationships within the data. Anchor Graph Regularization (AGR) (Yang

et al. 2022) uses these anchors to construct similarity graphs that enable efficient computations and preserve the integrity of the data structure. In the context of multi-view clustering, anchor-based techniques extend this principle by identifying anchors that capture common structures across multiple views (Qiyuan et al. 2020; Zhou et al. 2025; Li et al. 2024; Wang et al. 2022), which promotes the discovery of common patterns and ensures computational efficiency.

Apart from dimensionality, the diversity and heterogeneity of multi-view datasets require more sophisticated methods to uncover complex relationships. Tensor factorization models have proven useful for clustering and merging data across modalities (Chen et al. 2019, 2021; Xia et al. 2022). Techniques such as Tucker decomposition (Tucker Sep. 1966) and Kolda and Bader's framework (Kolda and Bader 2009) reduce dimensionality while identifying common latent structures. Sequentially Truncated Higher-Order Singular Value Decomposition (ST-HOSVD) (Fang et al. 2019) provides robust data reconstruction and clustering, while hybrid methods combining Dynamic Mode Decomposition (DMD) with Graph Laplacian techniques (Zhu et al. 2022) are promising for large-scale multi-view clustering. These tensor-based approaches complement anchor-based methods and provide powerful solutions to the challenges posed by high-dimensional, heterogeneous data.

In addition to the challenges associated with dimensionality and heterogeneity, several other issues need to be addressed for the effective application of multi-view clustering

- Model selection and scalability: Choosing an appropriate multi-view learning model and fusion strategy is a non-trivial task as it depends on the specific problem and the dataset. In addition, some multi-view learning algorithms have difficulty scaling effectively to large datasets, which limits their practical utility (Liu et al. 2013).
- Merging clustering results: A key challenge is to merge the clustering results from different views in a natural and coherent way. Since each view provides a different perspective on the data, integrating these results requires novel clustering target functions that can capture the diverse information from multiple sources.
- Determining the importance of views: Not all views contribute equally to the clustering process. One of the challenges of clustering with multiple views is determining the relative importance of the individual views. This requires mechanisms that weight the views appropriately and ensure that the more informative views have a greater influence on the clustering result
- Heterogeneity of data with multiple views: In real-world applications, multi-view data often exhibits significant heterogeneity in terms of scale, modality and quality. This heterogeneity poses a major challenge in multi-view learning (Yan et al. 2016). In particular, evaluating the correlation and redundancy between the different views is crucial, as overly correlated views can lead to suboptimal clustering results (Cai et al. 2013)
- Incomplete Multi-view Clustering (IMC): Addressing the challenge of missing data has led to the emergence of Incomplete Multi-view Clustering (IMC). However, most existing multi-view clustering methods assume complete data, which makes them less effective in real-world scenarios where data is often incomplete. Developing methods that can deal with missing data, especially with a high rate of missing views, remains an ongoing challenge (Liu 2020; Lv et al. 2022).

Table 1 provides an overview of key surveys on multi-view clustering (MVC), summarizing their main focus and contributions. Existing surveys often focus on specific subfields, applications, or disciplinary boundaries, leading to gaps in coverage and overlooking new interdisciplinary approaches. Each of these surveys presents a distinct perspective on MVC techniques, ranging from deep matrix factorizations and generative versus discriminative approaches to the categorization of algorithms and their evaluation in real-world applications. The surveys address various challenges in MVC, such as scalability and noise management, and propose future research directions to advance the field. Consequently, there is a compelling need for a systematic and up-to-date survey that bridges these gaps, offering a comprehensive overview of multi-view clustering methods and their applications across diverse domains. This table highlights the breadth of approaches and provides a comparative look at the state-of-the-art in multi-view clustering.

### 1.3 Motivation

The motivation for conducting this survey lies in the recognition of multi-view learning as a central and effective approach to overcoming the limitations of single-view methods. Multi-view learning not only compensates for these limitations, but also provides a more comprehensive representation of the data and thus solutions to various machine learning challenges. The versatility of multi-view learning makes it an invaluable tool that can be used in various domains and applications.

In contrast to traditional data representations that represent objects from a single view, multi-view data is semantically rich, which increases its practical utility despite its higher inherent complexity. In light of these considerations, this survey aims to provide a systematic overview of the main clustering methods documented in the scientific literature for processing multi-view data. Moreover, it contributes to a better understanding of the landscape of multi-view learning methods, their applications and their potential importance in addressing the challenges faced by today's machine learning techniques.

On the other hand, the lack of comprehensive survey papers in the literature addressing multi-view clustering arises from the diverse methodological landscape, rapid evolution of techniques, application-specific focus, and the overall growth of the field. Existing surveys (De Handschutter et al. 2021; Chao et al. 2021; Yang and Wang 2018; Lele et al. 2020) often focus on specific subfields, applications or disciplinary boundaries, leading to gaps in coverage and overlooking new interdisciplinary approaches. Consequently, there is a compelling need for a systematic and up-to-date survey that bridges these gaps, offering a comprehensive overview of multi-view clustering methods and their applications across diverse domains.

### 1.4 Structure of the survey

This review is systematically organized to provide a comprehensive overview of the multi-view clustering landscape, structured as follows:

1. *Introduction* The opening section highlights the growing importance of multi-view clustering in modern machine learning and data analytics. It discusses the significance, challenges, and motivations driving this area of research.

**Table 1** Summary of existing surveys on multi-view clustering methods, highlighting their focus on specific subfields, applications, or methodologies

| Year | Refs. | Title | Main focus | Key contributions |
|---|---|---|---|---|
| 2025 | Qin et al. (2025) | A Survey on Representation Learning for Multi-view Data | Focuses on multi-view clustering, self-supervised multi-view clustering and incomplete multi-view clustering | Provides a novel survey on multi-view clustering by organizing existing algorithms into two distinct categories: non-self-supervised and self-supervised multi-view clustering, addressing the gap left by previous surveys that overlooked simultaneous consideration of both |
| 2024 | Haris et al. (2024) | Breaking down multi-view clustering: A comprehensive review of multi-view approaches for complex data structures | Provides a comprehensive classification of Multi-View Clustering (MVC) methods, categorizing them into generative and discriminative approaches, with a focus on deep learning-based techniques | Classifies MVC methods into generative and discriminative categories, emphasizing deep learning's role in complex data structures. Provides a systematic comparison and identifies research gaps |
| 2024 | Raya et al. (2024) | Multi-modal data clustering using deep learning: A systematic review | Focuses on introducing a novel taxonomy for deep learning-based multi-modal clustering | Explores CNNs, Autoencoders, RNNs, and GCNs, identifies gaps in multi-modal clustering research, and suggests future research directions |
| 2023 | Fang et al. (2023) | A Comprehensive Survey on Multi-View Clustering | The survey categorizes current MVC approaches into two main technical Mechanisms: heuristic-based multi-view clustering (HMVC) and neural network-based multi-view clustering (NNMVC) | Explores key approaches within HMVC, including nonnegative matrix factorization, graph learning, and tensor learning, as well as deep representation learning and deep graph learning in NNMVC |
| 2022 | Wen et al. (2022) | A Survey on Incomplete Multiview Clustering | Focuses on incomplete multi-view clustering (IMC). IMC is particularly relevant for practical applications such as disease diagnosis, multimedia analysis, and recommendation systems, where incomplete data is common | The survey unifies IMC methods under common frameworks, conducts a comparative analysis of representative approaches, and highlights open problems to guide future research |
| 2021 | De Handschutter et al. (2021) | A survey on deep matrix factorizations | Focuses on deep matrix factorization techniques and their applications | Explores advanced matrix factorization methods in deep learning, highlights applications in multidimensional data analysis, and identifies future challenges |
| 2021 | Chao et al. (2021) | A Survey on multiview clustering | Review of multiview clustering methods, proposing a novel taxonomy based on generative and discriminative approaches | Classifies multiview clustering approaches into generative and discriminative categories; links MVC with representation, ensemble clustering, multi-task learning |

**Table 1** (continued)

| Year | Refs. | Title | Main focus | Key contributions |
|------|-------|-------|-----------|-------------------|
| 2020 | Lele et al. (2020) | An overview of recent multi-view clustering | Review of recent multiview clustering algorithms with an experimental focus | Divides algorithms into three main categories; conducts extensive experiments on seven datasets using accuracy, NMI, and purity metrics; proposes future directions |
| 2018 | Yang and Wang (2018) | Multi-view Clustering: A Survey | Review of multiview clustering algorithms, organized according to mechanisms and underlying principles | Proposes a taxonomy in five categories: co-training, multi-kernel learning, graph-based clustering, subspace clustering, and multitask multiview clustering |

The table underscores the need for a recent and comprehensive survey, like ours, to address gaps in interdisciplinary approaches and provide a unified perspective on the evolving landscape of multi-view clustering techniques and applications

2. *Fundamentals of Multi-view Clustering* (Sect. 2) This section lays the groundwork by explaining the key principles of multi-view clustering. It presents a taxonomy of multi-view approaches to help contextualize the methods discussed later.

3. *Classical Methods for Multi-view Clustering* (Sect. 3) We categorize and explain seminal approaches in multi-view clustering, focusing on classical techniques that have shaped the field.

4. *Exploring Graph-based Multi-view Clustering* (Sect. 4) We provide a comprehensive overview of novel graph-based algorithms tailored to the challenges of multi-view clustering, such as handling missing data, noise reduction, and efficient computation.

5. *Multi-view Clustering with Missing or Incomplete Data* (Sect. 5) This section addresses the recent challenges and methods for clustering with incomplete data, a common issue in multi-view settings.

6. *Formal Review of Typical Approaches* (Sect. 6) A rigorous mathematical examination of the typical multi-view clustering approaches, focusing on the formal framework and methods used.

7. *Datasets for Multi-view Clustering* (Sect. 3) We provide an overview of key datasets used to evaluate multi-view clustering methods, detailing common feature extraction techniques and considerations for selecting the optimal number of views.

8. *Conclusion* This section summarizes the main findings of the review and reinforces the importance of multi-view clustering in modern machine learning.

9. *Future Directions* (Sect. 9) The final section discusses the potential future developments and research opportunities in multi-view clustering.

## 2 Fundamentals of multi-view clustering

Multi-view learning has gained much attention due to its potential to improve the performance of models by leveraging multiple data perspectives. Integrating information from different views increases prediction accuracy and provides a more comprehensive understanding of the data structure, improving the robustness of the clustering process (Xu et al. 2013). By capturing the underlying structures from different views, multi-view methods can often perform better than single-view approaches (Kumar and Udupa 2011). Furthermore, learning from multiple views is crucial for domain adaptation, i.e. the alignment of data representations from different sources or domains (Sun et al. 2013).

Despite its advantages, multi-view clustering faces several challenges. The heterogeneity of data in terms of scale, modality and quality can make it difficult to effectively merge different views. In addition, overly correlated views can affect the performance of clustering algorithms and lead to suboptimal results. High-dimensional feature spaces resulting from the integration of multiple views require the use of dimensionality reduction techniques to avoid problems such as the "curse of dimensionality" (Sindhwani et al. 2008). In addition, model selection and scalability remain a challenge, as the choice of an appropriate multi-view learning model depends on the specific problem and the dataset, and some models do not scale well to large datasets (Liu et al. 2013).

In the field of multi-view clustering, established methods can be systematically categorized into different groups, each tailored to a specific strategy for merging information from

different views. We emphasize that the boundaries between the categories may be blurred, as a given clustering approach may belong to more than one category.

## 2.1 Taxonomy of multi-view clustering approaches

Multi-view clustering methods are diverse and employ a wide range of strategies to integrate information from different views. Below, we categorize these methods into distinct groups based on their underlying principles and techniques. While these categories provide a structured overview, it is important to note that certain approaches may overlap multiple categories.

- Co-training: This method starts by clustering data from a single view and iteratively refines the results across other views. It is advantageous because it allows for the enhancement of initial clustering through additional views. However, it requires a strong initial clustering and may struggle when views are highly inconsistent or noisy (Kumar and Udupa 2011; Tan et al. 2020).
- Co-regularized multi-view spectral clustering: This technique integrates multi-view learning with spectral clustering by introducing a co-regularization term, harmonizing clusterings across different views. The benefit is that it can effectively combine the views' information in an unsupervised manner. However, the choice of the appropriate regularization term can be challenging and may affect the clustering performance significantly (Kumar and Udupa 2011).
- Kernel-based multi-view clustering: Kernel methods map data into high-dimensional spaces to handle non-linearity, making it easier to cluster data that are not linearly separable. This method is useful for addressing the diversity in shapes across views. However, learning optimal kernels and performing dimensionality reduction are computationally expensive tasks, and the risk of overfitting increases as the number of views grows (Li et al. 2016; Chen et al. 2019; Wang et al. 2021).
- Subspace multi-view clustering: Subspace multi-view clustering aims to learn a unified feature representation by assuming that all views share this representation. It can be divided into subspace-based methods (Wang et al. 2015; Chen et al. 2020) and matrix factorization approaches (Khan et al. 2022), both of which are designed to analyze low-dimensional representations embedded in multiple views. However, conventional methods often struggle to capture high-dimensional information from nonlinear subspaces or overlook the high-level relationships between fundamental partitions obtained by clustering in a single view. These relationships could bridge the gap between heterogeneous feature spaces and improve clustering performance. Moreover, existing multi-view ensemble clustering methods often overlook the noise that arises in the data generation phase. To overcome this challenge, recent approaches have introduced new methods that incorporate regularization techniques and denoising strategies. For example, Zheng et al. Zheng et al. (2024) proposed a multiview subspace clustering method that combines hypergraph p-Laplacian regularization with low-rank subspace learning. This method was developed to capture complex hierarchical structures in the data while effectively mitigating the effects of noise. By integrating ensemble strategies, noise reduction and weight adaptation, these techniques improve the robustness of multi-view clustering and lead to better performance in heterogeneous and noisy environments. Similar ap-

proaches that fuse regularization, denoising, and low-rank learning in multi-view environments can be found in recent studies such as Pu et al. (2023), which propose a robust low-rank graph multi-view clustering method that integrates spectral embedding, non-convex low-rank approximation and noise handling.

● Deep learning based approach: The integration of deep learning with multi-view clustering leverages neural networks to create joint representations, significantly improving clustering performance in datasets with multiple views. However, current models have been criticized as shallow, as they directly map multi-view data to low-dimensional space, often neglecting essential nonlinear structure information within each view (Liu et al. 2021). To approximate these limitations, the authors in Gupta et al. (2024) propose DeConFCluster, an unsupervised multi-view clustering framework based on Deep Convolutional Transform Learning (CTL). By eliminating the need for an additional decoder network during training, DeConFCluster reduces overfitting in data-constrained scenarios, a common drawback of encoder-decoder-based methods. Furthermore, the model incorporates a K-Means-inspired loss function, enhancing representation learning for clustering tasks. The framework outperforms state-of-the-art multi-view deep clustering techniques on five benchmark datasets, showcasing its efficacy in capturing both joint representations and nonlinear structural information.

● Graph-based multi-view clustering: Graph-based methods integrate multiple views by constructing a unified graph from individual similarity matrices. The relationship between the data points of each view is represented as a graph, and the fusion of these graphs helps to capture complex interactions between the views (Wang et al. 2018; Tang et al. 2018). This approach is very powerful for managing diverse relationships, but it also brings computational challenges, especially in graph construction and matrix fusion, especially for large datasets (Zhang et al. 2021). A key problem is the selection of relevant views for graph creation, as not all views contribute equally to clustering. Some views may cause noise or redundancy, so it is important to identify and prioritize the most informative views. Methods such as adaptive graph learning and regularization techniques have been proposed to address these issues and optimize both computational efficiency and clustering accuracy (Wang et al. 2018; Zhang et al. 2021).

● Anchor-based methods: meet the main objectives of multi-view clustering as they provide a scalable approach to processing large data sets. By selecting representative samples (anchors) to approximate the entire dataset, scalability challenges are effectively addressed while preserving essential data structures. This approach improves scalability by reducing computational complexity and promotes efficiency when processing diverse, large-scale data sources. Anchors also contribute to robustness by capturing important patterns in different views, ensuring consistent performance even in the presence of noise or incomplete data. In addition, they enable flexibility as they can be seamlessly adapted to different clustering frameworks and methods. Ultimately, anchor-based methods promote effectiveness by preserving the quality of clustering results while optimizing resource utilization. A notable technique in this area is Anchor Graph Regularization (AGR) (Yang et al. 2022), which constructs a similarity graph using the selected anchors and enables efficient computations. In multi-view clustering, these methods are extended by selecting anchors that encapsulate common information across multiple views, which improves computational efficiency and data representation. Recent research, including work by Qiyuan et al. (2020); Zhou et al. (2025); Li et al.

(2024); Wang et al. (2022), has further explored anchor-based approaches, highlighting their ability to identify common patterns across different views and provide practical solutions for large-scale clustering tasks.

- *Other approaches* address the challenges of multi-view clustering by tackling specific limitations in traditional methods. For instance, the work (Yi and Niu 2024) focuses on two significant issues: (1) the inability of hard clustering techniques to capture uncertainty between samples and clusters, and (2) the challenge of effective incremental learning when the number of views increases. To address these, the study introduces a three-way fuzzy spectral clustering algorithm that generates soft clustering results, effectively modeling uncertainty. Furthermore, it incorporates an incremental learning mechanism based on sequential decision-making to handle dynamically increasing views. By combining these advancements, the proposed multi-view clustering algorithm based on sequential three-way decision-making achieves enhanced clustering accuracy and efficiency, as validated through experimental evaluations.

## 2.2 Practical applications of multi-view clustering

Multi-view clustering has found wide application in a number of areas where the integration of multiple data views enables a more comprehensive understanding of the problem at hand. This section highlights the main areas where multi-view clustering methods have been used effectively:

- Healthcare and medical diagnostics: Multiview clustering methods have proven their effectiveness on various types of data, including medical imaging, multi-omics data and physiological signals such as EEG. In medical imaging, different imaging modalities (e.g. MRI, CT scans, PET scans) are combined to improve diagnostic accuracy. Similarly, when integrating multi-omics data, consistent clustering patterns (common to all omics levels) can be identified alongside differential patterns (specific to individual omics types), helping to uncover biologically meaningful correlations and insights into disease mechanisms (Al-kuhali et al. 2022). Multi-view clustering has also been successfully applied to EEG signals and enables the classification of brain activity into different patterns, such as seizures and non-seizures, thus improving diagnostic precision (Zhan and Hu 2020). In addition, these techniques are used to identify disease subtypes, support personalized treatment planning and predict patient outcomes by integrating multiple clinical and biological data sources (Pfeifer et al. 2023; Feng et al. 2022). Al-kuhali et al. (2022).
- Social network analysis: In social networks, multi-view clustering can be used to analyze data from different sources such as social interactions, user behavior and content analysis. By combining multiple views, e.g. interactions between users, posts and metadata, multi-view clustering enables more accurate identification of communities, influencers or trends within the network (Kim and Sohn 2017; Lan et al. 2017; Dong et al. 2013).
- Computer vision and image processing: Multi-view clustering has been widely applied in computer vision, especially in object recognition and scene analysis. These tasks benefit from the ability to integrate information from multiple views or modalities to improve the accuracy of the models. One notable application is the tracking of pedes-

trians in crowded environments where targets are temporarily blocked and reappear. To solve this problem, a novel multiview clustering method has been developed that improves tracking accuracy by utilizing the correlation of fusion features within a multiview system. This method is particularly beneficial in cases where pedestrians are occluded or reappear after disappearing briefly (Chen et al. 2024). In object detection and scene analysis, multi-view clustering enhances recognition accuracy by integrating multiple views of the same scene. Techniques employing multi-view learning, such as those using deep neural networks, effectively capture inter-view relationships, leading to improved performance in complex tasks like 3D object recognition and scene segmentation. These methods have proven successful in dynamic environments, offering significant advancements in scene understanding and recognition accuracy (Teepe et al. 2024; Li et al. 2023).

- Natural language processing: In natural language processing, multi-view clustering has been successfully applied to document clustering and sentiment analysis. These applications often involve integrating different views of text data, such as syntactic and semantic information, to improve the quality and relevance of clustering results. Multi-view clustering techniques have shown effectiveness in these tasks by capturing diverse features from the text, which helps in grouping documents with similar content or sentiment (Zhou and Wang 2007; Zhan and Hu 2020; Serra et al. 2019; Hsueh et al. 2020; Yu et al. 2023).

- Recommendation systems and personalized medicine: Multi-view clustering plays a crucial role in data fusion as it enables the integration of different data views for more informed decision making. In recommendation systems, multi-view clustering methods enable the combination of different user data sources (e.g. behavior, preferences and demographics) to provide personalized recommendations. Multi-view clustering has also been used in personalized medicine to identify disease subtypes by combining different medical views (e.g. imaging data and genetic information). For example, a multi-view learning approach has been developed to identify imaging-related subtypes in mild cognitive impairment (MCI). The approach uses techniques such as Deep Generalized Canonical Correlation Analysis (DGCCA) to learn low-dimensional correlated embeddings, which significantly advances personalized medicine and medical diagnostics (Feng et al. 2022).

The inclusion of these practical applications emphasizes the versatility and impact of multi-view clustering techniques in practice and demonstrates their potential in various domains. This section provides readers with both an overview of existing approaches and practical insights into how these methods effectively address complex, real-world challenges.

In the following sections, we review different methodological approaches to multi-view clustering (MVC), each addressing distinct aspects of the problem rather than forming a step-by-step optimization process. Section 3 presents classical MVC methods that serve as the foundation for more advanced techniques. Section 4 focuses on graph-based approaches, which incorporate graph structures to enhance clustering performance by capturing relationships across views. Finally, Sect. 5 explores MVC techniques designed to handle missing or incomplete data, addressing a critical challenge in real-world applications. These three sections provide complementary perspectives on MVC, highlighting different methodological advancements tailored to specific challenges in multi-view learning.

# 3 Classical methods for multi-view clustering

Classical machine learning approaches for multi-view learning encompass diverse strategies. The co-training approach, as proposed by Blum et al. Blum and Mitchell (1998), uses multiple data views to improve model performance. Each view is trained independently and the information between the views is iteratively exchanged and refined. Figure 1 shows a graphical illustration of this idea.

Co-training is particularly beneficial in scenarios where there is little labeled data and allows the effective use of unlabeled data. However, practical considerations such as selecting features, addressing class imbalances and defining a robust matching criterion for the selection of instances to be labeled are essential. Co-training has applications in various fields, including natural language processing, computer vision and bioinformatics.

The co-regularized approach for semi-supervised learning from multiple views presented by Kumar et al. (2011) provides a powerful technique. This method combines the principles of co-regularization and semi-supervised learning to jointly learn representations for each view and make predictions in an unsupervised manner. Co-regularization enforces a constraint that ensures that the representations of the same instance from different views are close to each other, which promotes robust and discriminative representations.

Multi-kernel spectral clustering (MKSC) integrates kernel features from different views into a unified kernel matrix and uses spectral clustering for effective multi-view clustering (Xia et al. 2014). Canonical correlation analysis (CCA) is another approach that identifies correlated subspaces between views, as described in Chaudhuri et al. (2009); Feng et al. (2022).

Alternative strategies include the Joint Non-Negative Matrix Factorization approach, where data matrices from different views are factorized into non-negative matrices to capture common and view-specific information. Li et al. (2018) demonstrate spectral clustering on the learned representations for multi-view clustering.

Multi-View Spectral Embedding (MVSE) constructs individual similarity graphs for each view and merges them into a unified graph, which is subjected to spectral embedding for low-dimensional representations and subsequent clustering (Xingyu et al. 2019). Chen et al. Chen and Liu (2022) propose an improved spectral multiview clustering method that incorporates tissue-like P-systems and optimizes the similarity matrices through a weighted
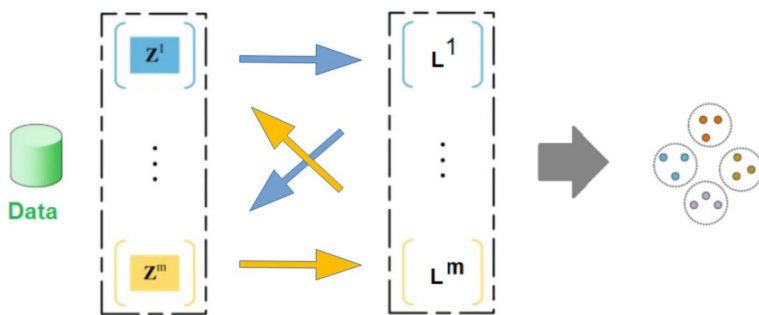


**Fig. 1** A graphic illustration of the co-training scheme with several views. The matrices $\{Z^1, ..., Z^m\}$ refer to the different data matrices corresponding to each view and $\{L^1, ..., L^m\}$ are the co-trained models. In this scheme, the information obtained from the individual views is systematically refined and iteratively exchanged, which promotes joint learning from the different views

iterative process. The approach combines the K-nearest neighbor algorithm, a weighted fusion operation, and an iterative update to obtain non-negative embedding matrices and achieve clustering results. Figure 2 shows the flow of chart of this approach. The method first determines the similarity matrices of each view (Z) by the K-nearest neighbor algorithm and then fuses all views into a unified matrix P by a weighted fusion operation. The unified matrix P in turn updates the similarity matrix for each view. Through an iterative updating process, the algorithm then takes the updated similarity matrices obtained in the previous step as input and combines the spectral clustering algorithm and the symmetric non-negative matrix factorization algorithm to obtain the non-negative embedding matrix M to directly output the clustering results.

Recent advances include joint frameworks that incorporate the orthonormality constraint and co-regularization (Cai et al. 2020). This improves class differences and feature scale consistency, which contributes to improved multiview clustering performance. In addition, Hajjar et al. El Hajjar et al. (2021) constraints to maintain consistent smoothing and enforce orthogonality, which improves the robustness and stability of clustering results. This method overcomes the limitations of conventional spectral clustering and provides more reliable and accurate results.

## 3.1 Kernel-based approaches

Multi-view spectral clustering algorithms are designed to effectively handle data with multiple views, exploiting the complementary information present in each view to improve the quality of clustering results. In their pursuit of optimal performance, many *multikernel-based methods* have also been developed with the main goal of obtaining a unified kernel
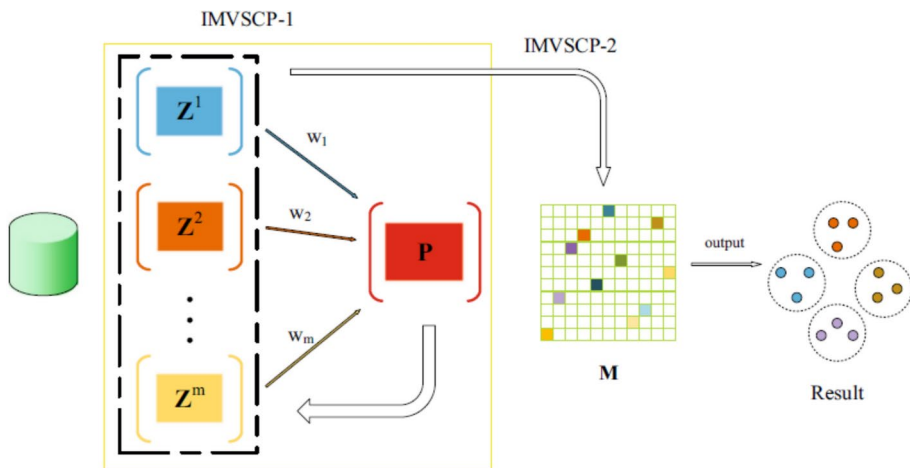
**Fig. 2** The flow chart of the Multi-view Spectral Clustering (Chen and Liu 2022). First, similarity matrices of the individual views (Z) are created using the K-nearest neighbor algorithm. Then, all views are combined into a unified matrix P using a weighted fusion operation. This unified matrix P is used to update the similarity matrix for each view. The algorithm iteratively refines the similarity matrices obtained in the previous step. It integrates the spectral clustering algorithm and the symmetric non-negative matrix factorization algorithm to generate the non-negative embedding matrix M, which leads directly to the clustering results

by specifying a predefined kernel for each different view and then combining all kernels linearly or nonlinearly.

Kernel functions are an essential part of these methods. They define a measure of similarity or affinity between data points in the transformed feature space. Commonly used kernel functions include the linear kernel, which computes the dot product between data points in the original feature space and thus measures their linear similarity, and the radial basis function (RBF) kernel, which measures similarity based on the Euclidean distance between data points in the transformed space and thus captures nonlinear relationships. Figure 3 gives a graphical illustration of the flow chart of Kernel-based approaches for multi view clustering.

For example, in Shudong et al. (2019), the authors discuss two kernel-based graph learning multi-view clustering methods with automatic weights. These two methods are used to map the data into a space where they are linearly separable. The first method uses a single kernel per view, while the second method uses a combination of multiple kernel matrices to improve the utilization of the input kernel matrix. Moreover, these methods simultaneously estimate the unified similarity matrix, the consistent spectral projection matrix, and the weight of each view without additional parameters.

Many of these approaches follow a two-step process to achieve clustering results. The initial step involves learning the joint affinity matrix, while the second step utilizes a hard clustering method like k-means clustering to obtain the final result. To address the inconsistency issue arising from the fact that the primary objective of the first step is not optimal clustering performance, a novel method is introduced.

In Zhu et al. (2018), the authors present a method called One-step Multi-view Spectral Clustering (OMSC), which integrates the steps of learning the affinity matrix of each view and the joint affinity matrix learned from the low-dimensional space of the data, as well as the step of k-means clustering in one framework. The joint affinity matrix is considered as the final clustering assignment. Moreover, the weighting of each view is automatically learned to reduce the impact of noisy views.

In Ren et al. (2020), the authors jointly estimate an optimal graph and an appropriate consensus kernel for clustering by forcing the global kernel matrix to be a convex combination of a set of basis kernels. Their proposed model enforces a regularization of the unified graph and the final kernel matrix.

However, the performance of multikernel-based methods strongly depends on the predefined kernel, including the kernel type (e.g. Gaussian kernel, linear kernel and polyno-
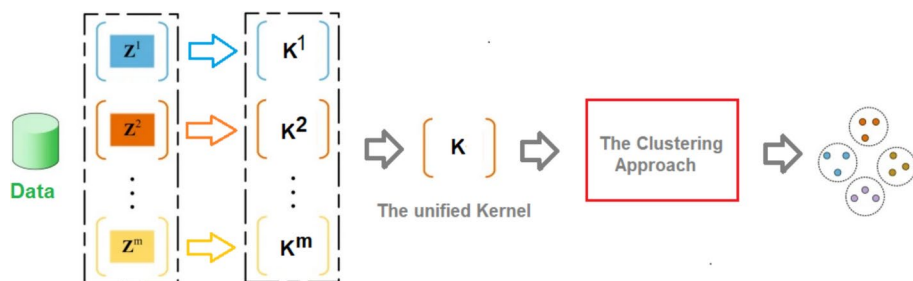


**Fig. 3** The flowchart of kernel-based multi-view spectral clustering. For multi-view data, we denote $(Z^1, Z^2, ..., Z^m)$ as the data matrix with $m$ views. Then an individual kernel is constructed for each view so that we have $(K^1, K^2, ..., K^m)$. A specific clustering method is then selected to cluster the data based on the unified kernel K

mial kernel) and the corresponding parameters. In addition, most of these methods do not exploit the label space of the data (i.e. the soft mappings of the cluster) and only extract information from the data space. New approaches have been developed to utilise the information in the label space. For instance, in El Hajjar et al. (2022), the authors develop a new method that incorporates the non-negative embedding matrix, which can be used as a cluster indicator matrix to perform the final cluster assignment without further post-processing steps such as k-means or spectral rotation. Unlike other multi-view clustering methods, this method can create a new graph based on the soft cluster labels. Moreover, this method smoothes the cluster-label indices over the data graphs and the label graph, which improves the performance.

## 3.2 Subspace-based approaches

Subspace clustering has been extensively researched over the last ten years due to its promising performance. Basically, it assumes that the data points come from multiple low-dimensional subspaces, with each cluster fitting into one of these low-dimensional subspaces. Considerable progress has been made in uncovering these underlying subspaces. In this context, several innovative methods have been proposed to overcome the challenges associated with multi-view data. In this section, four notable works are presented, each of which presents a different approach for clustering subspaces from multiple perspectives.

The work of Zhang et al. (2017) provides a method called Latent Multi-view Subspace Clustering (LMSC). LMSC introduces a unique perspective by clustering data points with latent representations while exploring complementary information from multiple views. Unlike traditional single-view subspace clustering, LMSC searches for an underlying latent representation, resulting in more accurate and robust subspace representations. The method is intuitive and is efficiently optimized by the Augmented Lagrangian Multiplier with Alternating Direction Minimization (ALM-ADM) algorithm, which has been confirmed by extensive experiments on benchmark datasets.

Khan et al. Ghufran (2023), however, addresses a limitation in existing multiview subspace clustering approaches by considering the structure of self-representation through consistent and specific representations. Based on low-rank sparse representations, the method uncovers global common representation structures among views and preserves geometric structures based on consistent and specific representations.

Other (Khan and Maji 2021) focuses on integrative clustering for high-dimensional data with multiple views based on manifold optimization. To identify consensus clusters, the algorithm constructs a joint graph Laplacian and optimizes a joint clustering target while minimizing the disagreements between individual and joint views. This optimization is alternatively performed over k-means and Stiefel manifolds (James 1976), modeling nonlinearities and differential clusters within individual views. The convergence of the algorithm is demonstrated over the manifold, and experimental results on benchmark and multi-omics cancer datasets show its superiority over existing multi-view clustering approaches.

The work of Wang et al. (2022) addresses the efficiency challenges in multi-view subspace clustering by proposing a Fast Parameter-free Multi-view Subspace Clustering with Consensus Anchor Guidance (FPMVS-CAG). This method performs anchor selection and subspace graph construction in a unified optimization formulation, thus promoting clustering

quality. FPMVS-CAG has linear time complexity with respect to the number of samples and automatically learns an optimal anchor subspace graph without additional hyperparameters.

### 3.3 Integration of deep learning with multi-view clustering

Integrating deep learning with multi-view clustering is a modern approach that uses neural networks to create joint representations for different views of data. This strategy improves clustering performance by learning feature representations that capture both intra-view and inter-view relationships. However, despite the growing interest and proliferation of algorithms based on various theories, many existing models are superficial and map multi-view data directly to low-dimensional spaces.

To address the limitation of shallow models, deep multi-view clustering algorithms have emerged as a promising solution. Du et al. (2021) propose a methodology utilizing multiple autoencoders with a layer-wise approach to capture nonlinear structural information within each view. This model incorporates local invariance within a view and integrates consistent, complementary information between views. The algorithm surpasses shallow models by unifying representation learning and clustering in a cohesive framework, enabling joint optimization of both tasks.

In a similar context, Wang et al. (2024) combines the flexibility of deep learning with the statistical benefits of data-driven and knowledge-driven feature selection, providing interpretable results. It learns nonlinear relationships in multi-view data by using deep neural networks to create low-dimensional, view-independent embeddings, while imposing a regularization penalty on the reconstructed data. The method uses the normalized Laplacian of a graph to model bilateral relationships between variables within each view, promoting the selection of related variables.

Other approaches (Wang et al. 2024) attempt to overcome the limitations of traditional deep multi-view subspace clustering approaches by using a decomposed optimization strategy that involves three stages: Pre-training with auto-encoders to extract multiscale features, fine-tuning to learn consensus self-expressions and generate high-quality pseudo-labels, and re-training with self-label supervision for robust clustering. This approach improves clustering performance by addressing issues such as skew in pipelined methods and complex parameter optimization in end-to-end methods.

- Deep subspace approaches: Multiview subspace clustering (MVSC) uses complementary information from multiple views to achieve better clustering compared to single-view methods. However, it often struggles with high-dimensional and noisy raw data. To overcome this, the self-guided deep multiview subspace clustering (SDMSC) model (Li et al. 2022) combines deep feature embedding with subspace analysis to uncover a reliable consensus data affinity relationship across views and embedding spaces. By using raw feature affinities as monitoring signals, SDMSC controls the embedding process itself, reduces the risk of poor local minima, and improves clustering performance. Similarly, Zhu et al. (2024) address the major limitations of traditional MVSC methods, such as insufficient multiview integration and lack of end-to-end learning, with the Multiview Deep Subspace Clustering Network (MvDSCN). This approach uses two sub-networks: Dnet for view-specific representations and Unet for joint representations. Deep convolutional autoencoders are used to construct a multiview self-representation

matrix in an end-to-end method.

- Advances in graph neural networks: Recent advancements in graph neural networks (GNNs) have provided new methods for multi-view learning. GNNs have garnered significant attention in recent years (Zonghan et al. 2021; Jiang 2023; Rebafka 2024), with their primary idea being the embedding of node representations by capturing and aggregating information from local neighborhoods. This technique has shown great promise, especially in tasks such as social network analysis, recommendation systems, anomaly detection, and the analysis of physiological data (Liu et al. 2018; Wu et al. 2019; Pucci et al. 2006; Manuel and Igone 2023). GNNs integrate information from multiple views, either through graph structures or self-supervised learning, leading to improved predictions at both the node and graph levels, and efficiently utilizing unlabeled data.

- Generative models for multi-view learning: In parallel, generative models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) have been adapted to handle multi-view data (Zhang et al. 2023; Goodfellow et al. 2020; Wang et al. 2018). These models enable the generation of data samples from multiple views, offering advantages such as data augmentation and enhanced model robustness. Adversarial multi-view learning techniques, in particular, align representations across different views while preserving view-specific characteristics, thus addressing the challenges posed by view heterogeneity. For instance, Tang et al. (2023) propose Consistent and Diverse Deep Latent Representations (CDDLR), which integrates K-means with spectral clustering to enforce consistency and diversity in the latent representations. This approach employs Laplacian-regularized deep neural networks to maintain diversity and consistency, providing a robust solution for multi-view subspace clustering.

- Solutions for deep multi-view subspace clustering: To address challenges in deep multi-view subspace clustering, Shi and Zhao (2023) introduce a novel solution known as Deep Multi-View Clustering based on Reconstructed Self-Expressive Matrix (DCRSM). This approach utilizes a reconstruction module to approximate self-expressive coefficients and resolves scalability issues by employing a minimal number of training samples. It effectively combines common and specific layers, facilitating the fusion of consistent and view-specific information, leading to improved clustering outcomes. Additionally, Wang et al. (2024) propose Decomposed Deep Multi-View Subspace Clustering with Self-Labeling Supervision (D2MVSC), which addresses traditional challenges in deep multi-view subspace clustering through a decomposed optimization strategy. This method includes a three-stage training process, adaptive fusion, and structure supervision, improving clustering accuracy and robustness.

- Further contributions to multi-view learning: Wang et al. (2023) contribute to subspace-based multi-view learning by proposing Multi-view Orthonormalized Partial Least Squares (MvOPLSs). This framework incorporates regularization techniques for model parameters, decision values, and latent projected points. By extending the framework with nonlinear transformations using deep networks, MvOPLSs achieves superior performance across diverse multi-view datasets compared to existing methods. These innovations mark significant progress in deep multi-view subspace clustering and contribute to the ongoing evolution of multi-view learning techniques.

Recent advancements in multi-view learning underscore the growing potential of multi-view data to enhance model performance, robustness, and generalization. The choice of

algorithm often depends on the specific characteristics of the data and the nature of the problem. Approaches such as deep learning and generative techniques provide comprehensive solutions, enabling the effective handling of complex multi-view data and improving clustering outcomes across a variety of applications.

## 4 Exploring graph-based multi-view clustering

Graph-based multi-view clustering, an essential aspect of data analysis, has made significant progress with the introduction of various methods. This section provides an overview of the latest approaches developed over the last five years, summarizing key contributions, including novel algorithms and methods that address the main challenges. These contributions provide valuable insights into the evolving landscape of multi-view clustering, offering significant contributions to the broader field of machine learning and data analytics. Researchers working on multi-view clustering will find this compilation valuable for understanding and navigating these developments.

- Multi-view graph clustering: The goal is to find a fusion graph across all views and apply graph-cut algorithms (e.g., spectral clustering) on the fusion graph to produce the final clustering result. We focus on graph-based clustering, where each view's data points are represented by a graph. These graphs are combined to facilitate clustering of the entire dataset, utilizing complementary information from different views to improve cluster mapping and provide a more comprehensive understanding of complex data structures.
- Schematic representation of graph-based methods: Fig. 4 illustrates the process. Here, the data matrix from each view is converted into a graph matrix, followed by a fusion method applied to create a unified graph. The fusion process derives the weights for each view automatically using a novel multi-view fusion technique. The unified graph similarity matrix is learned jointly with the individual graph similarity matrices, representing pairwise similarities between data samples.
- Limitations and Challenges in Graph-based Multi-view Clustering: Traditional graph-based methods face several challenges that impact their performance. Issues such as
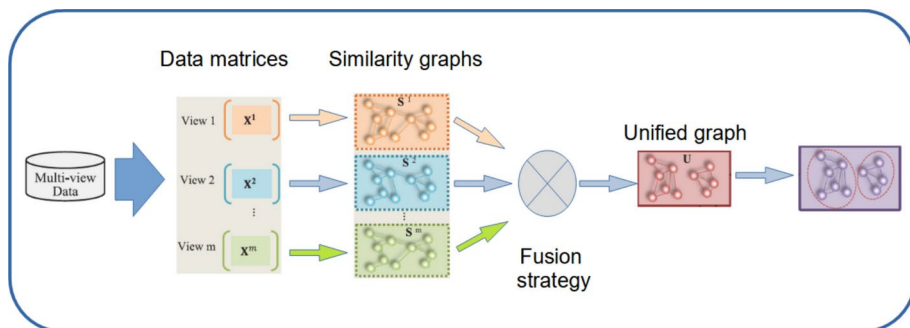


**Fig. 4** The flowchart of graph-based multi-view clustering, adapted from Wang et al. (2020). In this scheme, the data matrix of each view is converted into a graph matrix, followed by applying a fusion method across all views to create a unified graph

sample selection bias, additional clustering steps, and variations in similarity metrics complicate the clustering process (Wang et al. 2020). Furthermore, specific challenges in graph-based methods include:

– Post-processing necessity: Final clustering results often require methods like K-Means or spectral rotation to achieve consistent spectral embeddings, introducing uncertainties due to initialization.
– Parameter selection: Many graph-based models introduce additional parameters, complicating the task, which is inherently unsupervised.
– High computational costs: Spectral multi-view clustering methods require eigenvalue decomposition with a computational complexity of $O(n^3)$, and multi-view subspace clustering requires matrix inversion with similar computational complexity, making these methods particularly challenging for large-scale data (Zhanxuan et al. 2020).

Approaches like ASMV (Adaptive Similarity Metric Fusion for Multi-view Clustering) (Zhang et al. 2017) and GMC (Graph-based Multi-view Clustering) (Wang et al. 2020) have been proposed to mitigate these limitations and improve the effectiveness of graph-based multi-view clustering.

● A novel solution - MCGLSR: In response to these challenges, a new method called *Single Phase Multi-view Clustering using Unified Graph Learning and Spectral Representation (MCGLSR)* (Dornaika and El Hajjar 2023) is proposed. Unlike conventional methods that directly integrate similarity matrices (which may introduce noise), MCGLSR generates similarity graphs and a joint similarity matrix using a unified global objective function. This ensures that similarity matrices from different views align effectively, mitigating noise and promoting a more coherent data structure. The method outputs individual graphs, a joint graph, a joint spectral representation, and cluster mappings, eliminating the need for an external clustering algorithm.
● Incomplete data handling: In the next section, we address the challenges of multi-view clustering with incomplete data. Li et al. (2024) presents a breakthrough solution to the limitations of subspace clustering in the context of incomplete data. This approach uses a multiple kernel completion scheme to detect intact kernels and ensures the learning of complete, low-redundancy representations. By integrating multi-view subspaces with a weighted tensor low-rank constraint, the method explores higher-order relationships between views and assigns appropriate weights to each view. The result is a unified model that learns low-redundancy representations, view-specific subspaces, and their low-rank tensor structure, significantly improving subspace clustering in incomplete data scenarios.

## 5 Multi-view clustering with missing or incomplete data

The rapid development of technology has led to a massive growth in data, with multi-view data offering complementary information from different sources. This makes multi-view data more informative than single-view data. However, in real-world applications, multi-

view data is often incomplete. For example, a video without sound or data points with missing features represent incomplete views. Traditional multi-view clustering techniques cannot handle such incomplete data (Wen et al. 2022).

Dealing with incomplete multiview data has become an important research topic that has attracted much attention in recent years (Wen et al. 2020; Xia et al. 2022; Diallo et al. 2023). Several methods have been proposed to solve this problem:

- Filling missing views: A common approach is to fill missing views with zeros or the average values of the available instances for that view (Shao et al. 2015, 2016). In addition, weighted non-negative matrix factorization with $\ell_{2,1}$ regularization was used to improve clustering performance on incomplete views. However, this method often results in clustered data that lacks meaningful information, as filled values may cluster unrelated samples together.
- Consensus-based anchor guidance: A more reasonable approach proposed in Wang et al. (2022) emphasizes the use of available information in incomplete multiview data. This method uses consensus-based anchor guidance to achieve faster and more robust clustering, especially in image processing applications.
- Nonnegative matrix factorization (NMF): In Li et al. (2014), NMF was used to construct a latent subspace from the available multiview data, ensuring correct alignment between the corresponding instances across the different views. In Zhao et al. (2016), graph embedding techniques were applied to capture global structure in heterogeneous data with missing values and improve clustering of multimodal visual data.
- Hypergraph-based approach: To address the challenges of processing incomplete multiview data, the authors in Chen et al. (2025) propose a hypergraph-based approach. In contrast to traditional methods that focus on pairwise relations, their approach captures the higher order relations in the data using a hypergraph. This enables a more comprehensive exploration of both local and global data structures. Furthermore, the method combines non-negative matrix factorization with orthogonality constraints and K-means clustering, eliminating the need for post-processing.

Apart from these approaches, additional techniques such as sparsity constraints, weighted learning, graph learning, local embedding structures, and global embedding structures have been integrated into Incomplete Multi-View Clustering (IMVC) models to further improve clustering performance (Zhou et al. 2023; Yuan et al. 2021; Wang et al. 2019).

Despite the success of these techniques, there are still challenges to be addressed. Many models treat the contributions of all views equally, ignoring the varying discriminative information each view may provide. Additionally, IMVC methods are prone to misclassification, often resulting in fewer clusters than actually exist. Furthermore, there is a risk of imbalanced clustering, where some clusters are over-partitioned, while others are underrepresented, even when the data is balanced.

The challenge of clustering multi-view data with missing information remains an active area of research. As summarized in Table 2, which provides an overview of multi-view clustering methods developed in the last five years—including approaches developed to deal with incomplete data—future work should refine existing methods, explore new techniques, and address specific application areas to enhance the handling of incomplete multi-view data.

# 6 Formal review of typical approaches

Graph-based multi-view clustering (MVC) methods have made significant progress in recent years and have overcome important challenges in this area. The main ones are constructing reliable similarity graphs, mastering the computational complexity of large datasets, ensuring consistency between different views, and optimizing the integration of information from multiple views.

A critical problem with MVC is the construction of the similarity graph for each view, which often requires large amounts of memory and computational resources. Many methods have developed more efficient techniques for constructing graphs, either by integrating multiple views into a common framework or by using scalable algorithms that reduce computational costs without sacrificing accuracy.

Another major challenge is effectively assigning appropriate weights to views, which may have different importance in different datasets. Several methods have addressed this problem by proposing automatic or adaptive weight assignment mechanisms that eliminate the need for manual tuning and improve the robustness of the clustering process.

Consistency across multiple views is another key concern in MVC. Since different views may represent different perspectives or aspects of the same data, it is critical that the clustering results remain consistent across views. Newer approaches address this issue by explicitly modeling both consistent and inconsistent information, improving the overall accuracy and reliability of the clustering process.

Finally, the integration of various components involved in clustering, such as similarity matrices, spectral representations, and soft cluster assignments, has traditionally been handled sequentially. However, newer methods take a more integrated approach by optimizing these components together in a single framework, thus improving the efficiency and consistency of clustering results.

By overcoming these challenges, the methods described in the following sections contribute to scalable, robust and effective multi-view clustering and push the boundaries of what is possible with large, complex datasets.

To begin, we introduce the primary notation used in this section, where matrices are shown in bold capital letters and vectors in bold lowercase letters.

Let $X^{(v)}$ be a data matrix $(x_1^{(v)}, x_2^{(v)}, ..., x_n^{(v)}) \in \mathbb{R}^{n \times d^{(v)}}$ where $n$ is the number of data instances, $d_v$ the number of features in view $v$ where $v = 1, ..., V$. Given a matrix A, its trace is denoted by $\mathrm{Tr}(A)$ and its transpose by $A^T$. $A_{ij}$ is an element of the matrix A.

## 6.1 Auto-weighted multi-view clustering via kernelized graph learning (MVCSK)

Graph-based approaches have been widely adopted for multi-view clustering due to their ability to reveal hidden structures within data. However, a major challenge in these methods is the construction of accurate similarity graphs, which can be influenced by several factors, including the scale of the data, neighborhood size, choice of similarity metric, and the presence of noise and outliers.

To overcome these limitations, Shudong et al. (2019) propose a novel approach to multi-view clustering that simultaneously performs the clustering task and learns similarity relationships in kernel spaces. The key feature of this method is its ability to construct

a similarity graph that can be directly partitioned into $c$ connected components, where $c$ is the number of clusters. This ensures that the clustering process is more accurate and aligned with the underlying data structure.

One of the most significant innovations of this approach is its automatic weight assignment for each view. Unlike traditional methods, which either require manual assignment of weights or introduce additional parameters, the model in Shudong et al. (2019) learns the optimal weight for each view as part of the clustering process. This automatic weighting improves the robustness and adaptability of the clustering method, as it allows the model to adjust to the unique characteristics of the data without requiring prior knowledge of the view importance.

Additionally, the proposed model incorporates multiple kernel learning to address the sensitivity to the input kernel matrix. The inclusion of this extension enables the model to better handle nonlinear relationships in the data and ensures that it can capture complex patterns that would otherwise be missed by traditional graph-based clustering methods.

The model operates within a joint learning framework, which simultaneously solves three subtasks:

- Constructing the most accurate similarity graph,
- Automatically assigning optimal weights to each view,
- Finding the cluster indicator matrix.

By solving these subtasks jointly, each task enhances the others, leading to a more robust and effective clustering outcome. Experimental results on benchmark datasets show that the method proposed by Shudong et al. (2019) outperforms several state-of-the-art multi-view clustering algorithms, demonstrating its effectiveness in handling complex, incomplete, and noisy multi-view data.

The objective function of MVCSK is:

$$
\min_{S, P} \sum_{v=1}^{V} \sqrt{Tr\left(K_v - 2\,K_v\,S + S^T\,K_v\,S\right)} + \mu\,||S||_2^2 + \lambda\,Tr\left(P^T\,L\,P\right) \tag{1}
$$
$$
s.t.\ S \geq 0,\ P^T P = I,
$$

where $S$ is the unified similarity matrix, $K_v$ is the kernel matrix of each view, $P$ is the unified spectral projection matrix from which the final clustering is estimated using an extra step.

The objective function is designed to achieve two main goals: (1) to construct an accurate and smooth similarity matrix $S$ that reflects the true relationships between data points, and (2) to learn the optimal projection matrix $P$ that enables the effective clustering of the data. The first term in the objective function minimizes the distance between the kernel matrices of each view and the unified similarity matrix, promoting consistency between the different views. The second term regularizes $S$ by enforcing its smoothness, while the third term regularizes the projection matrix $P$ by encouraging it to preserve the underlying graph structure as encoded in the Laplacian matrix $L$.

The constraints $S \geq 0$ and $P^T P = I$ ensure that the similarity matrix is non-negative and that the projection matrix is orthonormal, respectively. Together, these terms and constraints guide the model toward a robust and effective clustering solution.

**Table 2** Compilation of multi-view clustering methods developed in the last five years

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Zhou et al. (2025) | 2025 | ✓ Graph learning<br>✓ Anchor-based<br>✓ Bipartite graph | · Integrate anchor graph learning and subspace graph construction into a unified optimization framework based on a bipartite graph.<br>· It enhances clustering by jointly optimizing the projection matrix, consensus anchor matrix, and similarity matrix, ensuring connectivity constraints to form clusters directly | · Image datasets<br>· Multi-view datasets |
| Chen et al. (2025) | 2025 | ✓ Hypergraph<br>✓ Non-negative matrix factorization<br>✓ Tensor Schatten $p$-norm | · Reconstructed missing views using a hypergraph, capturing both local structures and higher-order relationships.<br>· It integrates representation learning and clustering into a one-step framework, avoiding suboptimal results from two-step approaches | · Incomplete multi-view datasets |
| Dou et al. (2025) | 2025 | ✓ Multi-level graphs<br>✓ Deep non-negative matrix factorization | · Addresses the challenge of balancing diversity and consistency across multiple views.<br>· It integrates feature learning, multi-level topology representation, and clustering into a unified framework. Specifically, it uses deep non-negative matrix factorization (DNMF) to learn multi-level (hierarchical) representations of objects | · Image datasets<br>· Benchmark datasets |
| Zhou and Zhang (2024) | 2024 | ✓ Sparse graph learning | · Addresses the challenge of allocating contributions of different views by assigning view-specific weights instead of equal weights.<br>· The method directly obtains cluster indicators by applying low-rank constraints, eliminating the need for post-processing | · Caltech101-07, Caltech101<br>· UCI-digit, Mfeat, STL10 |
| Zhao et al. (2024) | 2024 | ✓ Graph learning<br>✓ Bipartite graph<br>✓ Dynamic adaptation | · A learnable graph filter that dynamically refines the original feature space, progressively filtering out noise and producing a smooth, clustering-friendly representation.<br>· A unified bipartite graph that combines multi-granular structural information from different views, capturing both distinct and shared features across views | · Multiple benchmark datasets |
| Wang et al. (2024) | 2024 | ✓ Deep learning | · Combines the flexibility of deep learning with the statistical benefits of data-driven and knowledge-driven feature selection, providing interpretable results.<br>· It learns nonlinear relationships in multi-view data by using deep neural networks to create low-dimensional, view-independent embeddings, while imposing a regularization penalty on the reconstructed data. The method uses the normalized Laplacian of a graph to model bilateral relationships between variables within each view, promoting the selection of related variables | · Holm Breast Cancer, Study, LGG Dataset (grade 2 and 3)<br>· Shear Transformed, MNIST Dataset |
| Li et al. (2024) | 2024 | ✓ Graph-based<br>✓ Subspace-based<br>✓ Kernel-based | · Incomplete multi-view subspace clustering based on multiple kernel completion, low-redundant representation learning, and weighted tensor low-rank constraint.<br>· The unified objective function combines the intact view-specific subspaces and the hidden low-rank tensor | · BBC-Sport, MSRCv1,<br>· 100Leaves, NGs,<br>· 3Sources, ORL |

**Table 2** (continued)

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Chao et al. (2023) | 2023 | ✓ Graph-based ✓ Subspace-based ✓ Anchor-based | · Anchor-based Multi-View Subspace Clustering with Graph Learning. · Integrate anchor learning and the construction of coefficient matrices through a unified optimization procedure that exploits the global and local structure of the samples and the learned anchors | · Caltech101-7, Caltech101-20, · SUN-RGBD, Animal, AwA, · NUSWIDEOBJ, YoutubeFace |
| Pan and Kang (2023)[1] | 2023 | ✓ Graph-based ✓ Anchor-based | · High-order multi-view clustering based on graph filtering, intrinsic relationships up to infinity order, adaptive graph fusion, and anchors selected by high-order structure | · ACM, DBLP, IMDB, · Amazon Photos/Computers |
| Zhong and Pun (2023) | 2023 | ✓ Graph-based | · Integrate adaptive weighting, Laplacian embedding (spectral embedding), consensus graph learning and discrete indicator matrix learning into a unified framework. · Outputs clustering results directly without the need for post-processing | · Leaves100, COIL100, NGs, · BBCSport, HW, ORL, · Mfeat, ALOI |
| Dornaika and El Hajjar (2023) | 2023 | ✓ Graph-based ✓ Kernel-based | · Unified single-phase multi-view clustering with consensus graph learning and spectral representation. Jointly generates similarity graphs of the views and their joint similarity matrix using a unified global objective function. · It takes as input a kernelized representation of the features and directly returns the individual graphs, the joint graph, the joint spectral representation and the cluster assignments | · ORL, COIL20, BBCSport, · MSRCv1, MNIST-25000 |
| Dornaika and El Hajjar (2023) | 2023 | ✓ Graph-based ✓ Kernel-based | · Multi-view clustering via kernelized graph and nonnegative embedding. · It is based on a single global criterion that jointly provides the consistent similarity matrix for all views, the consistent spectral representation, the soft cluster assignments and the view weights | · COIL20, ORL, Out-Scene, · MNIST, BBCSport, MSRCv1, · Caltech101-7, Extended-Yale |
| Jie et al. (2023) | 2023 | ✓ Low-rank tensors | · An approach for learning low-rank tensors that can provide a consensus low-dimensional embedding matrix for incomplete multiview clustering. · It involves learning individual low-dimensional embedding matrices from incomplete multiview data, utilizing the self-expression property of high-dimensional data | · Reuters, O-Scene, · Handwritten, COIL-20, · ProteinFold, Flower17, · SUN-RGBD, · 100leaves, Caltech101 |
| Luo and Cao (2022) | 2022 | ✓ Subspace Dual Clustering | · Combines dual-clustering and multiview subspace learning to simultaneously discover consensus representation and dual-clustering structure using alternating optimization. · A unified framework is developed to jointly explore clustering and subspace learning | · Real-world multiview dual · single-clustering datasets |
| Liu and Song (2022) | 2022 | ✓ Virtual-label Guided Matrix Factorization (VLMF) | · Utilizes graph regularization to capture geometric structure, and a virtual-label guided matrix factorization to recover and learn consensus latent representations. · The approach integrates clustering and latent representation learning into a joint optimization process | · Incomplete multi-view datasets |

**Table 2** (continued)

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Wang et al. (2022) | 2022 | ✓ Double embedding-transfer-based Multi-view Spectral Clustering (DETMSC) | · Incorporates two types of embeddings: consistency embedding and feature embedding. · Knowledge transfer between these embeddings is achieved via bipartite graph co-clustering, which improves clustering accuracy by learning both consistency across views and diversity of features. · Robustness to noisy data is enhanced through sparse constraints | · Real-world benchmark datasets |
| Yin et al. (2022) | 2022 | ✓ Anchor-based Incomplete Multi-view Spectral Clustering (AIMSC) | · Uses anchor points to connect instances from each view and recover missing data. · The similarities between data points are derived from their relationship with anchor points. · The method then applies anchor-based spectral clustering to generate accurate clustering results | · Multiple benchmark datasets |
| Chen et al. (2022)[2] | 2022 | ✓ Graph-based ✓ Anchor-based | · A unified framework for anchor learning, graph construction and partitioning, while keeping the complexity almost linear. · Through mutual improvement, the model achieves a more discriminative and flexible anchor representation and cluster indicator | · Notting-Hill, · Caltech101-20, · VGGFace2-50, YTF |
| Yuan et al. (2022) | 2022 | ✓ Graph-based ✓ Kernel-based | · A robust self-tuning multi-view clustering method. · It solves the problems of existing multi-view clustering methods, such as initialization sensitivity, fixing the number of clusters, and reducing the influence of outliers | · Advertisement, 3source, Flowers · Caltech101, ImageS, Cornell, · Texas, Washingdon, Wisconsion |
| Wang et al. (2022) | 2022 | ✓ Graph-based ✓ Subspace-based ✓ Anchor-based | · Fast parameter-free multiview subspace clustering with consensus anchor guidance. · A subspace clustering method with linear time complexity, joint anchor selection and graph construction, and parameter-free characteristics for large-scale applications | · Caltech101-20, Caltech101-all, · CCV, SUN-RGBD, NUS-WIDE, · AWA, MNIST, YoutubeFace |
| Kang et al. (2021)[3] | 2021 | ✓ Graph-based ✓ Anchor-based ✓ Bipartite-graph | · A scalable graph learning framework that incorporates anchor points and the concept of a bipartite graph. · In contrast to conventional, a bipartite graph is constructed to illustrate the relationship between samples and anchor points. · A connectivity constraint is used to represent clusters directly by connected components | · Caltech101-7, · Citeseer, · NUS |
| El Hajjar et al. (2021)[4] | 2021 | ✓ Graph-based | · Multi-view spectral clustering via constrained nonnegative embedding, overcoming limitations of traditional spectral clustering by integrating constraints for smoother nonnegative embedding and orthogonal columns | · COIL201, ORL, Out-Scene, · BBCSport, Caltech101, MSRCv1, · Extended-Yale, MNIST-10000 |

**Table 2** (continued)

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Xiao et al. (2021) | 2021 | ✓ Graph-based<br>✓ Subspace-based | · An approach to similarity merging in multi-view spectral clustering that addresses the challenge of assigning uniform weights to all samples within a view.<br>· It deals with biased or missing elements in incomplete views, using sparse subspace clustering to form initial similarity matrices | · BBCSport, ORL,<br>· Still DB, MSRC,<br>· UCI, NUS, 3-Sources |
| Horie and Kasai (2021) | 2021 | ✓ Graph-based<br>✓ Subspace-based | · An approach to merge similarities in multi-view spectral clustering.<br>· Addresses the challenge of assigning uniform weights to all samples within a view, manages missing elements in incomplete views,<br>· Uses sparse subspace clustering to form initial similarity matrices | · BBC,<br>· NGs,<br>· WebKB,<br>· 100leaves |
| Yin et al. (2021) | 2021 | ✓ Graph-based<br>✓ Subspace-based<br>✓ Kernel-based | · A one-step multi-view spectral clustering method that addresses the problem of inconsistency.<br>· It splits the non-negative embedding matrix into two matrices: the joint non-negative embedding matrix, which represents the joint cluster structure, and the specific non-negative embedding matrix, which represents the specific cluster structure for each view | · COIL20, MSRC-v1,<br>· Caltech101-7,<br>· Caltech101-20,<br>· ORL, 3-Sources |
| Chen et al. (2021) | 2021 | ✓ Graph-based<br>✓ Subspace-based<br>✓ Graph filtering | · A smoothed multiview subspace clustering that preserves the geometric features of the graph through graph filtering, which simplifies the subsequent clustering process | · Handwritten, Citeseer<br>· Caltech101-7 |
| Niu et al. (2021) | 2021 | ✓ Graph-based<br>✓ Subspace-based | · A one-step multiview subspace clustering method with incomplete views.<br>· It uses low-rank matrix factorization to learn a consensus representation matrix, and then combines it with the objective function of non-negative embedding and spectral embedding subspace clustering | · BBCSport, NGs, WebKB,<br>· BUAA, Orl, Yale,<br>NUS-WIDE,<br>· Caltech101, CCV |
| Xie et al. (2020) | 2020 | ✓ Graph-based<br>✓ Subspace-based | · A subspace learning based multiview clustering method.<br>· It derives a joint latent representation from the latent subspace rather than from the original data space by linear transformation.<br>· The latent representation has a low-rank structure, which reduces computational complexity. The similarity matrix is then dynamically learned from this latent representation using manifold learning | · MSRC-v1,<br>· UCI Digits,<br>· NUS-WIDE,<br>· Scene15 |
| Zhao et al. (2020) | 2020 | ✓ Graph-based | · An innovative multiview spectral clustering model capable of performing graph fusion and spectral clustering simultaneously.<br>· The fusion graph approximates the original graph from each individual view while maintaining a clear and distinct cluster structure | · BBC, Reuters,<br>· Digits, Caltech101-20 |

**Table 2** (continued)

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Wang et al. (2020) | 2020 | ✓ Graph-based | · The approach can jointly estimate the similarity graph matrix, the unified graph matrix, and the final cluster assignment by an innovative multi-view fusion technique.<br>· This method imposes a rank constraint on the Laplacian matrix, ensuring the accurate derivation of clusters | Two toy data sets<br>· Two-Moon,<br>· Three-Ring |
| Zhanxuan et al. (2020) | 2020 | ✓ Graph-based | · A simple and efficient approach to multiview spectral clustering that aims to learn a sparse similarity matrix that is consistent across all views.<br>· The advantage of the model lies in its ability to directly obtain a closed-form solution without the need for iteration.<br>· The method introduces only one additional parameter, which depends on the way the similarity matrix is constructed and can be practically set to a small value | · Caltech101,<br>· MSRCv1,<br>· NUS-WIDE,<br>· ORL,<br>· 3-Sources |
| Zhou et al. (2020) | 2020 | ✓ Subspace-based | · An approach that learns the joint information to leverage the underlying correlations across multiple views while capturing view-specific details to represent specific features for each independent view.<br>· This is achieved without being affected by redundancy or the high dimensionality of the data | · Yale, MSRCV1,<br>· Caltech101-7,<br>· BBCSport, CMU-PIE |
| Shudong et al. (2019)[5] | 2019 | ✓ Graph-based<br>✓ Kernel-based | · An innovative multiview learning model is presented that is capable of simultaneously performing a multiview clustering task and capturing similarity relationships in kernel spaces.<br>· The model autonomously assigns optimal weights to each view without the need for additional parameters | · Texas, Cornell,<br>· Washington,<br>· Winconsin,<br>· BBC, BBCSport, NUS-WIDE |
| Guoqing et al. (2019) | 2019 | ✓ Co-clustering | · The method addresses the problem of missing data in multiview clustering. In contrast to existing multiview co-clustering approaches that struggle with incomplete data, especially when different patterns of missing data are present.<br>· This method uses an indicator matrix. The indicator matrix highlights which data elements are present and clustering performance is measured solely on the observed values | · Clinical data from the heroin treatment study |
| Wang et al. (2019) | 2019 | ✓ Graph-based<br>✓ Consensus learning<br>✓ Perturbation risk | · Addresses clustering of multi-view data with missing instances using spectral perturbation theory to construct similarity matrices and learn a consensus Laplacian matrix | · 100Leaves, Flowers17, Mfeat,<br>· ORL, 3Sources, BBCSport |
| Peng et al. (2019) | 2019 | ✓ Graph-based | · A method that estimates the number of clusters by providing cross-view consensus on view-specific similarity graphs instead of relying on view-specific data representations.<br>· this end, a novel objective function is used to project the raw data into a space where the projection accounts for geometric consistency and cluster assignment consistency | · Caltech101,<br>· LandUse-21,<br>· Scene-15,<br>· Still-DB |

**Table 2** (continued)

| Refs. | Year | Methods | Multi-view clustering strategy | Dataset |
|---|---|---|---|---|
| Xing et al. (2019) | 2019 | ✓ Subspace-based | · A correntropy-based method for dealing with noise in multiview clustering using a view-specific embedding from an information-theoretic perspective. · The objective function uses the Frobenius norm to efficiently estimate the dense connections between points that lie in the same subspace | · UCI Digits, 3-Sources, · Movies617, BBC, · Cora, Washington |
| Zhan et al. (2018) | 2018 | ✓ Graph-based | · A graph learning framework, optimizing initial graphs from different views using low-rank constraint Laplacian matrix. · It leads to a global unified graph estimation | · UCI Digits, · Caltech101, · Notting-Hill, COIL–20 |

[1] See Sect. 6.4 for the formal definition of the method

[2] See Sect. 6.7 for the formal definition of the method

[3] See Sect. 6.5 for the formal definition of the method

[4] See Sect. 6.3 for the formal definition of the method

[5] See Sect. 6.1 for the formal definition of the MVCSK method

## 6.2 Consistency-aware and inconsistency-aware graph-based multi-view clustering (CI-GMVC)

In Horie and Kasai (2021), the authors propose a novel graph-based multi-view clustering method called CI-GMVC, which addresses a major limitation of existing approaches, including GMVC. Although conventional methods use a unified graph matrix for clustering multi-view data, they do not consider the inconsistent parts of the input graph matrices. This omission can lead to suboptimal clustering performance. The CI-GMVC method explicitly separates the consistent and inconsistent parts of the graph matrices and thus improves the robustness and accuracy of the clustering process when analyzing multi-view data.

The proposed objective function aims to jointly estimate the consensus graph, the spectral representation of the data, and the consistent graphs in each view. This is given by:

$$
\min_{\mathrm{A}_v=1,\ldots,V,\ \mathrm{U},\ \mathrm{F},\ \alpha} \sum_{v=1}^{V} \alpha_v \, ||\mathrm{U} - \mathrm{A}_v||_2^2
$$

$$
+ 2\lambda \, Tr\left(\mathrm{F}^T \, \mathrm{L}_U \, \mathrm{F}\right) + \sum_{v,w=1}^{V} b_{vw} \, Tr\left((\mathrm{S}_v - \mathrm{A}_v)(\mathrm{S}_w - \mathrm{A}_w)^T)\right) \tag{2}
$$

$$
\text{subject to } U_{ij} \geq 0, \quad \sum_j U_{ij} = 1, \quad \mathrm{F}^T \mathrm{F} = \mathrm{I}, \quad \mathrm{S}_v \geq \mathrm{A}_v \geq 0
$$

where $\mathrm{S}_v$ is the graph matrix of the *v*-th view (known input), U is the consensus graph, F is the common spectral representation of the data, and $\mathrm{A}_v$ is the consistent graph matrix of view *v*.

The objective function consists of several key terms, each with its own role in the clustering process:

- $\alpha_v ||\mathrm{U} - \mathrm{A}_v||_2^2$: This term measures the difference between the consensus graph U and the consistent graph matrix $\mathrm{A}_v$ for each view $v$. The parameter $\alpha_v$ controls the importance of this term for each view. Minimizing this term encourages the consensus graph to be close to the consistent graph for each individual view, ensuring that the consensus graph captures the consistent parts of the multi-view data. This helps to identify the parts of the data that are reliably shared across all views.
- $2\lambda \, Tr(\mathrm{F}^T \, \mathrm{L}_U \, \mathrm{F})$: This term regularizes the common spectral representation F by enforcing smoothness on the graph defined by $\mathrm{L}_U$, the graph Laplacian associated with the consensus graph. The parameter $\lambda$ controls the strength of the regularization, encouraging the spectral representation to preserve the structure of the consensus graph. This helps to capture the global structure of the data and ensures that the learned representation aligns with the consensus information across all views.
- $\sum_{v,w=1}^{V} b_{vw} \, Tr\left((\mathrm{S}_v - \mathrm{A}_v)(\mathrm{S}_w - \mathrm{A}_w)^T\right)$: This term captures the pairwise relationships between the consistent graph matrices of different views. The matrices $\mathrm{S}_v$ and $\mathrm{S}_w$ represent the graph matrices for views $v$ and $w$, and the term $(\mathrm{S}_v - \mathrm{A}_v)$ represents the inconsistent part of the graph for view $v$. Minimizing this term encourages the consistent parts of the graph from different views to be similar to each other. The parameter $b_{vw}$ controls the weight of the relationship between views $v$ and $w$. This term is critical

for handling the inconsistencies in the multi-view data and aligning the parts of the data that are inconsistent across views.

This approach includes several constraints to ensure the validity of the graph-based multi-view clustering process. Specifically, the consensus graph matrix U must be non-negative, with each row summing to 1. The common spectral representation matrix F is required to be orthonormal, meaning $F^T F = I$, where I is the identity matrix. Additionally, the consistent graph matrix $A_v$ for each view must be non-negative and less than or equal to the graph matrix $S_v$ of the respective view.

Together, these terms and constraints aim to learn a consensus graph that captures the most consistent structure across the views, while also considering the individual inconsistencies in each view. The model balances between capturing the global structure (via the consensus graph) and maintaining the consistency within each view (via the consistent graphs). By jointly learning the consensus graph, spectral representation, and the consistent parts of the graphs, this approach enables more accurate and robust multi-view clustering.

### 6.3 Constrained multi-view spectral clustering via integrating nonnegative embedding and spectral embedding (CNESE)

In El Hajjar et al. (2021), the authors address the limitations of spectral clustering-based methods, which usually require an additional clustering step after performing a non-linear projection of the data, which can lead to poorer clustering results due to factors such as initialization or outliers. To overcome these challenges, they propose a constrained version of a method called "Constrained Multi-view spectral clustering via integrating Nonnegative Embedding and Spectral Embedding" (CNESE) by integrating nonnegative embedding and spectral embedding. The model retains the advantages of the original method while incorporating two important constraints:

- Ensuring consistent smoothing of the nonnegative embedding across all views and,
- Imposing an orthogonality constraint on the columns of the nonnegative embedding.

This approach provides the clustering result directly and avoids the need to post-process the clustering and use additional parameters by finding the non-negative embedding and the spectral embedding matrix simultaneously. The proposed objective function aims to jointly estimate the soft cluster assignment matrix H and the individual spectral projection matrices $P_v$. This function is given by:

$$\min_{P_v, H} \sum_{v=1}^{V} ||S_v - H P_v^T||_2 + \lambda \sum_{v=1}^{V} \sqrt{Tr\left(H^T L_v H\right)} + \alpha Tr\left((H^T H - I)^T (H^T H - I)\right) \quad (3)$$
$$s.t. \ H \geq 0, \ P_v^T P_v = I.$$

where $\lambda$ is a regularization parameter, and $\alpha$ is a large positive value ensuring the orthogonality of the matrix H (last term).

The objective function consists of three main terms:

- Factorization term, $\sum_{v=1}^{V} ||S_v - H P_v^T||_2$, aims to factorize the input graphs $S_v$ of each view into the product of the cluster indicator matrix H and the view-specific spectral

projection matrix $P_v$. This term ensures that the clustering information contained in each view is well represented by the cluster assignment and projection matrices.

- Smoothing term, $\sum_{v=1}^{V} \sqrt{Tr(H^T L_v H)}$, introduces a smoothing effect on the cluster indicator matrix H. This term promotes consistency of cluster assignments across views by promoting similarity within clusters across views, where $L_v$ is a graph Laplacian matrix representing the local structure of the data.

- Orthogonality constraint, $\alpha Tr((H^T H - I)^T (H^T H - I))$, ensures that the matrix H is orthogonal. The orthogonality constraint is controlled by the parameter $\alpha$, which is a large positive value. This term helps maintain the distinctness of the clusters by enforcing that the cluster indicator matrix H forms an orthonormal matrix, meaning that each cluster is represented by a unique vector in the space.

Finally, the optimization problem is subject to the following constraints:

1. $H \geq 0$, ensuring that the cluster assignment matrix has non-negative entries.
2. $P_v^T P_v = I$, ensuring that the projection matrices for each view are orthonormal. This objective function jointly optimizes the cluster indicator matrix H and the projection matrices $P_v$ across all views, taking into account both the consistency within the individual views and the orthogonality of the cluster representations.

## 6.4 High-order multi-view clustering (HMvC)

In Pan and Kang (2023), the authors propose a novel approach to multi-view clustering, called High-Order Multi-view Clustering (HMvC), which addresses several challenges in graph-based clustering methods. The key points of this approach can be summarized as follows:

- graph-based clustering with high-order information: HMvC incorporates higher-order neighborhood information to capture complex interactions within the data that are often overlooked by conventional lower-order methods.
- Graph filtering for structure encoding: The approach uses graph filtering to encode structural information and enables unified processing of attributed graph data and non-graph data in a single framework.
- Exploring long-distance intrinsic relationships: By utilizing intrinsic relationships up to infinite order, HMvC enriches the learned graph, captures distant connections between data points, and improves the representation of underlying structures.
- Adaptive graph fusion mechanism: To integrate the consistent and complementary information from different views, the authors propose an adaptive graph fusion mechanism. This mechanism generates a consensus graph that effectively combines the relevant information from all views.
- Superior performance: Experimental results show that HMvC outperforms several state-of-the-art multi-view clustering methods, including some deep learning-based techniques, on both non-graph and attributed graph datasets.

This approach highlights the importance of high-order relationships and adaptive fusion mechanisms in achieving robust and effective multi-view clustering. The objective function is given by:

$$\min_{S_v,\, S,\, \gamma} \sum_{v=1}^{V} \gamma_v \left( ||X_v^T - X_v^T S_v||_2^2 + \alpha\, ||S_v - f_v(W_v)||_2^2 \right) + \beta\, ||S - \sum_{v=1}^{V} \gamma_v\, S_v||_2^2 + \mu\, ||S||^2$$

$$\text{s.t.} \ \sum_{v=1}^{V} \gamma_v = 1, \gamma_v \geq 0 \tag{4}$$

Here the data matrix $X_v$ corresponds to a filtered version with $k$-order filtering based on the adjacency matrix of view $v$, $W_v$ is a normalized similarity matrix based on the cosine similarity, and $f(W_v) = W_v + W_v^2 + ... + W_v^n$.

The objective function consists of four main terms:

- Self-representation term, $\sum_{v=1}^{V} \gamma_v ||X_v^T - X_v^T S_v||_2^2$, ensures that the graph of each view $S_v$ can accurately reconstruct the filtered data matrix $X_v^T$. The data matrix $X_v$ is derived by filtering the order $k$ based on the adjacency matrix of the view $v$, which captures the structural information within the view.

- Higher- order consistency term, $\sum_{v=1}^{V} \gamma_v \alpha ||S_v - f_v(W_v)||_2^2$, includes the higher order neighborhood information in the graph construction. Here, $f_v(W_v) = W_v + W_v^2 + \cdots + W_v^n$ aggregates the information from multi-order relationships based on the normalized similarity matrix $W_v$. This term ensures that the learned graph $S_v$ matches the higher order structural information within each view.

- Consensus graph alignment term, $\beta ||S - \sum_{v=1}^{V} \gamma_v S_v||_2^2$, encourages the consensus graph S to be a weighted combination of the view-specific graphs $S_v$. The weights $\gamma_v$ are adaptively learned to reflect the importance of each view and to ensure consistency between views.

- Regularization term, $\mu ||S||^2$, serves as a regularization of the consensus graph S to promote smoothness and avoid overfitting.

The constraints $\sum_{v=1}^{V} \gamma_v = 1$ and $\gamma_v \geq 0$ ensure that the weights $\gamma_v$ form a valid convex combination that maintains the balance between the contributions of the different views

This formulation effectively integrates self-representation, higher-order relations, and adaptive graph fusion, making it robust for multiview clustering tasks across different datasets.

## 6.5  Multi-view structured graph learning (MSGL)

In Kang et al. (2021), the authors propose a scalable graph learning framework tailored for subspace clustering that addresses three key challenges of existing graph-based methods: high computational cost, the inability to explicitly detect clusters, and the lack of generalizability to unseen data. The main contributions of their work are summarized as follows:

- Bipartite graph construction: Instead of constructing a full graph for $n$ samples, the proposed framework builds a bipartite graph to model the relationships between data

samples and anchor points. This approach significantly reduces computational complexity and improves scalability, making the method suitable for large datasets.

- Cluster interpretability: The method contains a connectivity constraint that ensures that the connected components in the bipartite graph correspond directly to the clusters. This eliminates the need for an additional clustering step and the cluster memberships are explicitly displayed.
- Link to K-means clustering: The authors provide a theoretical link between their learning approach for bipartite graphs and the K-means clustering algorithm, which provides insights into its clustering mechanism and further justifies its effectiveness.
- Extension to multi-view data: The framework is extended to handle multi-view datasets. This multi-view model achieves linear scalability with respect to $n$, maintaining computational efficiency while effectively integrating information from multiple views.

The work in Kang et al. (2021) represents a significant advancement in graph-based clustering, particularly for large-scale and multi-view datasets, by combining efficiency, explicit cluster discovery, and the ability to generalize to new data points. The objective function in Kang et al. (2021) is designed to tackle the challenges of scalability, explicit cluster identification, and multi-view data integration in subspace clustering. The function is expressed as:

$$
\min_{\alpha, Z, P} \sum_{v=1}^{V} \alpha_v \, ||X_v - A_v \, Z^T||_2^2 + \lambda_1 ||Z||_2^2 + \lambda_2 \, Tr\left(P^T \, L \, P\right) + \sum_{v=1}^{V} \alpha_v^{\gamma} \tag{5}
$$
$$
s.t. \ \ P^T \, P = I, \ \ Z \geq 0, \ \ Z^T \, 1 = 1, \gamma < 0
$$

where $A_v$ is the matrix of anchors in view $v$, Z is the consensus graph matrix between data and their anchors.

The terms and constraints in this objective function are described below:

- Reconstruction term, $\sum_{v=1}^{V} \alpha_v \, ||X_v - A_v \, Z^T||_2^2$, models the reconstruction error between the data matrix $X_v$ of view $v$ and the product of the anchor matrix $A_v$ and the consensus graph matrix $Z^T$. This term ensures that the learned bipartite graph accurately represents the relationship between the data points and their anchors, weighted by the view-specific coefficient $\alpha_v$.

- Consensus graph regularization, $\lambda_1 ||Z||_2^2$, imposes an $\ell_2$-norm penalty on the consensus graph matrix Z, encouraging sparsity and regularizing the learned graph.
- Smoothness term, $\lambda_2 \, Tr\left(P^T \, L \, P\right)$, promotes the smoothness of the clustering solution by incorporating the Laplacian matrix L. Here, P is the projection matrix encoding the cluster memberships, and this term ensures that points connected in the graph are assigned similar cluster memberships.
- Weight regularization term, $\sum_{v=1}^{V} \alpha_v^{\gamma}$, regularizes the view-specific weights $\alpha_v$, controlling their influence on the clustering process. The hyperparameter $\gamma < 0$ promotes a balanced contribution from different views.

The constraints in the objective function serve to ensure the validity and interpretability of the learned matrices. The orthogonality constraint $P^T \, P = I$ guarantees that the projection

matrix P encodes unique cluster memberships, where the clusters are represented as distinct, non-overlapping components.

The non-negativity condition $Z \geq 0$ enforces that the consensus graph matrix Z contains only non-negative entries, which is crucial for interpretability as it represents the relationships between data points and anchor points in a meaningful way.

Finally, the normalization constraint $Z^T 1 = 1$ ensures that each row of Z sums to one, effectively treating it as a valid probability distribution. This combination of constraints ensures a robust and meaningful clustering solution. This objective function integrates graph learning and clustering in a unified framework, and ensures scalability, interpretability and robustness in both single-view and multi-view subspace clustering scenarios.

## 6.6 Fast parameter-free multi-view subspace clustering with consensus anchor guidance (FPMVS-CAG)

In Wang et al. (2022), the authors propose a new method called Fast Parameter-free Multi-view Subspace Clustering with Consensus Anchor Guidance (FPMVS-CAG). This approach addresses the main limitations of existing multi-view subspace clustering techniques, especially their cubic time complexity and dependence on heuristic anchor sampling strategies. The main contributions of this method can be summarized as follows:

- Unified optimization framework: FPMVS-CAG integrates anchor selection and subspace graph construction into a single optimization problem. This joint formulation allows the two processes to interact and reinforce each other, improving the quality of clustering.
- Parameter-free learning: Unlike conventional methods that require manual tuning of hyperparameters, FPMVS-CAG automatically learns an optimal anchor subspace graph without introducing additional parameters. This eliminates the time-consuming selection of parameters, which improves the applicability of the method.
- Scalability: The method achieves linear time complexity with respect to the number of samples. This makes it particularly suitable for large-scale applications and eliminates the inefficiency of previous methods with cubic complexity.

The combination of these properties ensures that FPMVS-CAG is well suited for large-scale multi-view clustering tasks while maintaining high accuracy and robustness

To achieve this, the authors proposes to estimate a graph matrix that relate the data to their anchors, optimizing the following objective function:

$$\min_{\alpha, W_{v=1,\ldots,V}, A, Z} \sum_{v=1}^{V} \alpha_v^2 \, ||X_v - W_v \, A \, Z||_2^2 \tag{6}$$
$$s.t. \; W_v^T \, W_v = I, A^T \, A = I, \; Z \geq 0, \; Z^T \, 1 = 1, \; \alpha^T \, 1 = 1$$

where $X_v \in \mathbb{R}^{d_v \times n}$ is the data matrix in view $v$, the respective projection matrices $W_v, v = 1, \ldots, V$ target the consensus anchor guidance, $A \in \mathbb{R}^{d \times l}$ is the latent consensus anchor matrix (a learnable dictionary matrix). The number and size of the anchors, $d$ and $l$ are specified beforehand).

The objective function aims to minimize the discrepancy between the data matrices and their reconstruction across all views. This term enforces that the data from each view is well represented by the anchor and projection matrices, ensuring consistency between multiple views.

The constraints ensure meaningful representations: The projection matrices $W_v$ and the anchor matrix A are orthonormal. The consensus graph matrix Z must be non-negative ($Z \geq 0$) and its rows are normalized such that $Z^T 1 = 1$, where 1 is a vector of ones. In addition, the weights $\alpha_v$ are normalized, where $\alpha^T 1 = 1$.

Once the consensus graph matrix Z is obtained, its singular vectors are used as the consensus spectral embedding. The final clustering is achieved by applying the $k$-mean algorithm to these embeddings. This framework integrates the reconstruction of data from multiple views, anchor learning and spectral clustering into a unified model.

## 6.7 Efficient orthogonal multi-view subspace clustering (OMSC)

In Chen et al. (2022), the authors propose a method titled Efficient Orthogonal Multi-view Subspace Clustering (OMSC), which addresses the challenges of clustering multi-view data in large-scale scenarios. The main contributions can be outlined as follows:

- Integrated approach: OMSC combines anchor selection, graph construction and clustering in a cohesive framework. This integration ensures that these components reinforce each other, resulting in more robust and adaptive anchor representations and cluster assignments.
- Learning orthogonal bases: In contrast to traditional approaches that predefine anchors using $k$-means or uniform sampling, OMSC introduces a mechanism to learn high-quality orthogonal bases within the unified model. This approach increases the algebraic structure and improves the accuracy of clustering.
- High scalability: The model achieves near linear complexity with respect to the size of the dataset, making it suitable for clustering tasks with extremely large datasets. This efficiency results from the joint modeling process and a carefully developed alternative optimization strategy.

By addressing these critical aspects, OMSC provides a scalable and effective solution for clustering multi-view data in real-world, large-scale applications (Chen et al. 2022). The objective function is given by:

$$\min_{\alpha, W_v, A, Z, G, F} \sum_{v=1}^{V} \alpha_v^2 \, ||X_v - W_v \, A \, Z||_2^2 + \lambda \, ||Z - G \, F||_2^2$$

$$s.t. \ \ W_v^T \, W_v = I, A^T \, A = I, \ \ Z \geq 0, \ \ Z^T \, 1 = 1, \ \ \alpha^T 1 = 1, \tag{7}$$

$$G^T \, G = I, F_{ij} \in \{0,1\}, \sum_{i=1}^{k} F_{ij} = 1$$

where $A \in \mathbb{R}^{d \times l}$ is the latent consensus anchor matrix, Z is the consensus graph matrix between data and their anchors, $G \in \mathbb{R}^{l \times k}$ is the centroid matrix ($k$ being the number of

clusters) and $F \in \mathbb{R}^{k \times n}$ is an indicator matrix providing the data partition, $F_{ij} = 1$ if the i-th instance is assigned to the $k$-th cluster and 0 otherwise.

- The first term, $\sum_{v=1}^{V} \alpha_v^2 ||X_v - W_v A Z||_2^2$, minimizes the reconstruction error for each view $v$. Where $X_v$ is the data matrix for the view $v$, $W_v$ is the projection matrix, A is the latent consensus anchor matrix and Z is the consensus graph matrix. The weights $\alpha_v^2$ balance the contribution of the individual views.

- The second term, $\lambda ||Z - G F||_2^2$, ensures that the consensus graph matrix Z matches the cluster representation formed by the centroid matrix G and the cluster indicator matrix F. The parameter $\lambda$ controls the strength of the regularization.

### 6.8 Anchor-based multi-view subspace clustering with graph learning (AMVSCGL)

The article (Chao et al. 2023) introduces a novel method to address the challenges in Multi-view Subspace Clustering (MVSC), a key problem in pattern recognition and data mining. This method generates a joint coefficient matrix rather than constructing large, view-specific graphs, which enhances scalability and reduces memory consumption. Additionally, it incorporates a graph learning term that combines both global and local information from multiple views, making it more efficient and adaptable for large datasets. In contrast, Chen et al. (2022) relies on predefined orthogonal anchors and constructs affinity graphs for each view, which can be computationally expensive and less scalable for large datasets. Furthermore, Chao et al. (2023) is parameter-free, making it easier to implement, while Chen et al. (2022) requires the manual tuning of hyperparameters, such as the number of anchors.

The objective function is as follows.

$$\min_{\alpha, A_v, Z} \sum_{v=1}^{V} \left( \alpha_v ||X_v - A_v Z^T||_2^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{k} ||X_v(i,:) - A_v(:,j)||_2^2 + \alpha_v^\gamma \right) \tag{8}$$
$$s.t. \ A_v^T A_v = I_k, \ Z \geq 0, \ Z^T 1 = 1, \gamma < 0$$

where $k$ is the number of clusters, $A_v$ is the matrix of anchors in view $v$ (a learnable matrix), Z is the consensus graph matrix between data and their anchors.

- First term: $\alpha_v ||X_v - A_v Z^T||_2^2$. This term measures the reconstruction error between the data in view $v$ ($X_v$) and the approximation of the data using the anchor matrix $A_v$ and the consensus graph matrix Z. The weight $\alpha_v$ controls the contribution of each view to the overall objective.
- Second term: $\lambda \sum_{i=1}^{n} \sum_{j=1}^{k} ||X_v(i,:) - A_v(:,j)||_2^2$. This term ensures that the data points $X_v(i,:)$ are well approximated by the anchors $A_v(:,j)$ from the corresponding view. The regularization parameter $\lambda$ controls the strength of this term and helps to maintain the consistency of the anchors across the different views.

- Third term: $\alpha_v^\gamma$. This term introduces a penalty based on the value of the weight $\alpha_v$ for each view. The exponent $\gamma$ is a negative constant, which usually leads to a regularization effect that favors smaller values of $\alpha_v$. This helps to control the influence of the individual views on the overall model so that the method can handle different contributions

of the views more flexibly.

## 6.9 Self-taught multi-view spectral clustering (SMSC)

The work (Zhong and Pun 2023) focuses on the relax-and-discretize strategy, which uses predefined similarity graphs and learns a consensual Laplacian embedding for clustering. The main contributions of this work are as follows:

- Problem addressed: The work addresses the problem of information loss in MVC methods that arises from the independent processing of similarity graphs and the subsequent graph partitioning steps. This approach is often inefficient and reduces the quality of clustering.
- Proposed framework: The authors present the *Self-taught Multi-view Spectral Clustering (SMSC)* framework. This method considers both the manifold structure induced by Laplacian embeddings and the cluster information embedded in the discrete indicator matrix, which enables learning an optimal consensus similarity graph.
- Graph fusion schemes: Two graph fusion schemes are presented:

  – Convex combination scheme: This approach combines the similarity graphs from different views by a convex combination.
  – Centroid graph fusion scheme: In this method, the similarity graphs are merged taking into account the centroid of each view's data representation of the individual views.

- Self-taught mechanism: The self-taught mechanism integrates manifold structure and clustering information to learn an optimal consensus similarity graph for graph partitioning and thus improve clustering results.

These contributions make SMSC a powerful and efficient approach by overcoming the challenges of information loss and optimizing the learning process for clustering from multiple views. The objective function of this approach is given by:

$$\min_{\alpha, S, P, Y} || \sum_{v=1}^{V} \alpha_v \, S_v - S ||_2^2 + \lambda_1 \, Tr \left( P^T \, L \, P \right) + \lambda_2 \, ||P - Y \left( Y^T Y \right)^{-1}||_2^2$$

$$s.t. \ P^T \, P = I, \ S \geq 0, S^T \, 1 = 1, \ \alpha \, 1 = 1, \ Y_{ij} \in \{0, 1\}, \sum_{i=1}^{k} Y_{ij} = 1 \tag{9}$$

where S denotes the consensus graph matrix, P denotes the consensus spectral representation matrix, where Y denotes the indicator matrix. The terms that constitute the objective function, as given by Eq. 9, are described as follows:

- The first term, $|| \sum_{v=1}^{V} \alpha_v \, S_v - S ||_2^2$, penalizes the difference between a weighted sum of view-specific matrices $S_v$ and the consensus matrix S. The weights for each view are labeled $\alpha_v$, and the target encourages the consensus matrix S to be close to the weighted sum of view-specific matrices.

- The second Term, $\lambda_1 \, Tr \, (\mathrm{P}^T \, \mathrm{L} \, \mathrm{P})$, regularizes the projection matrix P by encouraging it to align with the Laplacian of the graph L, where $\lambda_1$ is a regularization parameter. This ensures that the data points are mapped in such a way that the structure defined by the Laplacian matrix is preserved.
- The third term, $\lambda_2 \, ||\mathrm{P} - \mathrm{Y} \, (\mathrm{Y}^T \mathrm{Y})^{-1}||_2^2$, enforces consistency between the projection matrix P and a matrix derived from Y, where Y is an indicator matrix that provides the cluster assignments. The regularization parameter $\lambda_2$ controls the weighting of this term and promotes the alignment between P and the cluster structure represented by Y.

The authors also propose another learning model that releases the estimation of the blending weights $\alpha_v$. They propose to minimize the following objective function with the concept of automatic view weighting:

$$
\min_{\alpha, \mathrm{S}, \mathrm{P}, \mathrm{Y}} \sum_{v=1}^{V} ||\mathrm{S}_v - \mathrm{S}||_2^2 + \lambda_1 \, Tr \, (\mathrm{P}^T \, \mathrm{L} \, \mathrm{P}) + \lambda_2 \, ||\mathrm{P} - \mathrm{Y} \, (\mathrm{Y}^T \mathrm{Y})^{-1}||_2^2
$$

$$
s.t. \;\; \mathrm{P}^T \, \mathrm{P} = \mathrm{I}, \;\; \mathrm{S} \geq 0, \;\; \mathrm{S}^T \, 1 = 1, \;\; Y_{ij} \in \{0, 1\}, \sum_{i=1}^{k} Y_{ij} = 1
$$

(10)

## 6.10 Multi-view clustering using unified graph learning and spectral representation (MCGLSR)

Most existing multi-view clustering methods follow a sequential process in three steps: estimating individual or consistent similarity matrices, performing spectral embedding, and then partitioning into clusters. These methods often reach their limits when integrating all the necessary components. To address this issue, the authors in Dornaika and El Hajjar (2023) present a novel approach to multi-view clustering which integrates multiple components into a single framework, overcoming the weaknesses of previous methods. It jointly solves the consistent similarity matrix for all views, the spectral representation, the soft cluster assignments and the weights of the views within a single criterion The key innovations are as follows:

- The method eliminates the need for an additional clustering step as it resolves the clustering assignments directly.
- The soft cluster assignments are directly linked to the representation of the views, which improves consistency across all views.

This approach is characterized by its ability to jointly optimize multiple components of the clustering process without the need for an additional partitioning step. This provides a more efficient and integrated solution compared to conventional methods.

The proposed objective function is given by:

$$\min_{S^v, S^*, P^*, H} \sum_{v=1}^{V} \left\{ Tr\left(K^v - 2K^v S^v + S^{vT} K^v S^v\right) + ||S^v||_2^2 + \lambda_1 ||S^* - S^v||_2^2 \right\} \tag{11}$$
$$+ \lambda_2 ||S^* - H P^{*T}||_2^2 + \lambda_3 Tr\left(H^T L^* H\right) + \lambda_4 Tr\left(P^{*T} L^* P^*\right),$$

where $\lambda_1$, $\lambda_2$ and $\lambda_3$ and $\lambda_4$ are regularization parameters. $K_v \in \mathbb{R}^{n \times n}$ is the kernel data matrix in view $v$, $S_v \in \mathbb{R}^{n \times n}$ is the graph matrix in view $v$, $S^* \in \mathbb{R}^{n \times n}$ the consensus graph matrix, $H \in \mathbb{R}^{n \times k}$ is the matrix of soft cluster assignments, and $P^* \in \mathbb{R}^{n \times n}$ is the consensus spectral representation. The terms that constitute the objective function, as given by Eq. 11, are described as follows:

- First term (for each view $v$): $Tr\left(K^v - 2K^v S^v + S^{vT} K^v S^v\right)$ This term calculates the trace of the difference between the kernel matrix $K^v$ and the graph matrix $S^v$, which measures the alignment between the data and the graph for each view.
- Second term (graph matrix regularization): $||S^v||_2^2$ This regularization term applies the $\ell_2$ norm to the graph matrix $S^v$, which promotes sparsity and reduces overfitting.
- Third term (consistency between graphs): $\lambda_1 ||S^* - S^v||_2^2$ This term ensures that the consensus graph matrix $S^*$ is similar to each individual graph matrix $S^v$, enforcing consistency between the different views.
- Fourth term (regularization of the consensus representation): $\lambda_2 ||S^* - H P^{*T}||_2^2$ This term forces the consensus graph matrix $S^*$ to be close to the product of the soft cluster assignments $H$ and the consensus spectral representation $P^{*T}$, which strengthens the relationship between the graph and the clustering.
- Fifth term (smoothness in the consensus representation): $\lambda_3 Tr\left(H^T L^* H\right)$ This term regulates the cluster assignment matrix $H$ by ensuring a smoothing of the cluster labels via the Laplacian matrix $L^*$.
- Sixth term (regularization of the spectral representation): $\lambda_4 Tr\left(P^{*T} L^* P^*\right)$ This term regulates the consensus spectral representation $P^*$ by ensuring smoothing over the Laplacian matrix $L^*$ and preserving the structure of the data across the views.

## 7 Dataset exploration

In this section, we provide a comprehensive overview of the key datasets that are important for the evaluation of multi-view clustering methods. We provide relevant information such as the number of samples, the number of views and the number of clusters. Table 3 gives an overview of the common datasets including image-based and feature-based datasets.

Indeed, features are usually extracted either by deep neural networks or manually created features. The choice of feature extraction method plays a crucial role in the effectiveness of the clustering process.

Deep neural networks have proven to be powerful tools for feature extraction in various domains due to their ability to automatically learn hierarchical representations from raw data. Notable architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been extensively used for extracting discriminative features from images in multi-view clustering scenarios ( LeCun et al. (2015); Goodfellow et al. (2016)). These networks show that they are able to capture complicated patterns and rela-

tionships within the data and thereby improve the quality of the extracted features for subsequent clustering tasks.

In contrast, handcrafted features (Moujahid and Dornaika 2020, 2019; Rushi et al. 2015; Tan and Triggs 2010) involve a careful design process that utilizes expertise to develop features that contain relevant information for clustering purposes. The handcrafted feature selection process requires a sophisticated understanding of the characteristics of the dataset, and the features chosen should encapsulate the inherent structures of the data to enable effective clustering.

It is important to note that the number of views within a given dataset is a variable parameter that varies in different studies and applications. Determining the optimal number of views is a crucial consideration that affects the overall performance of the multi-view clustering algorithm. Researchers have applied various strategies to solve this problem. They range from empirical selection based on expert knowledge to automated methods guided by statistical criteria (Lin et al. 2023; Chao et al. 2021).

## 8 Conclusion

In this comprehensive survey we systematically explored the landscape of multi-view clustering, providing a structured and insightful journey through various aspects of the field. We started with an introduction where we highlighted the increasing importance of multi-view clustering in modern machine learning and data analytics and explained both the challenges and motivations. The following sections covered the basics of multi-view clustering, classical methods, kernel-based approaches, subspace-based methods and depth-based techniques, providing a differentiated understanding of each method.

We also looked at graph-based multi-view clustering and introduced novel algorithms that can overcome specific challenges of multi-view clustering such as missing data, noise suppression and computational efficiency. The section on multi-view clustering with missing or incomplete data looked at recent developments in data analysis, focusing on the challenges posed by data incompleteness in clustering scenarios.

A careful examination of typical approaches was conducted in a formal mathematical sense, shedding light on the mathematical framework underlying these methods. This rigorous analysis contributes to a deeper understanding of the methods used in multi-view clustering and focuses on the formal aspects of their mathematical foundations.

In addition, a comprehensive overview of the key datasets that are critical to the evaluation of multi-view clustering methods is provided. This includes a concise summary of essential details covering common feature extraction methods and a critical examination of how to determine the optimal number of views.

To conclude, this review not only summarizes existing knowledge but also lays the foundation for future advances in multiview clustering research. By systematically organizing and presenting the various facets of the field, it serves as a valuable resource for researchers seeking to foster a deeper understanding of the complexity and potential of the multiview clustering landscape. By the end of this overview, it is clear that multi-view clustering is a dynamic and evolving field with a variety of perspectives.

**Table 3** Exploring variability in multi-view clustering datasets

| Dataset | Description | Samples | Views | Clusters |
|---|---|---|---|---|
| Caltech101 (Fei-Fei et al. 2004)[a] | The Caltech101 dataset contains images of objects belonging to 101 categories. Here are two of the most frequently used subcategories | | | |
| | - Caltech101-7 is a subset with 7 object categories | 1474 | 6 | 7 |
| | - Caltech101-20 is a subset with 20 object categories | 2386 | 6 | 20 |
| SUN-RGBD (Song et al. 2015) | Large-scale indoor scene understanding dataset with RGB-D images | 10335 | 45 | 2 |
| Animal (Lampert Christoph et al. 2013) | Subset of the Animals with Attributes dataset (AwA) dataset containing images of animals with attribute annotations | 11673 | 20 | 4 |
| AWA (Lampert Christoph et al. 2013) | The Animals with Attributes dataset (AwA) containing images of animals with attribute annotations | 30475 | 50 | 6 |
| NUS-WIDE (Chua et al. 2009) | Dataset with web images and associated object labels | 30000 | 31 | 5 |
| YoutubeFace[b] | Dataset focused on face recognition with video frames and face labels | 101499 | 31 | 5 |
| BBC-Sport[c] | Collection of articles from the sports section of the BBC website relating to sports news in five subject areas, including athletics, cricket, football, rugby, and tennis. Each article has two views (Li et al. 2024). The first view consists of 3183 features, and the second view consists of 3203 features | 544 | 2 | 5 |
| COIL-20[d] | Contains 1,440 grayscale images, divided into 20 classes, each class consisting of 72 images. Each image was taken from three different (Dornaika and El Hajjar 2023): intensity feature (1024), LBP feature (3304), and Gabor feature (6750) | 1440 | 3 | 20 |
| UCI-digits[e] | Comprises 2,000 data points divided into 10 digit classes, each class containing 200 handwritten digits. The dataset includes 3 views (Lu et al. 2015): profile correlation (216), Fourier coefficient (76) and Karhunen-Loeve coefficient (64) | 2000 | 6 | 10 |
| ORL[f] | Consists of 40 individuals, with each individual represented by 10 different images. Each image is characterized by three views (Li et al. 2024): intensity feature (4096), LBP feature (3304) and Gabor feature (6750) | 400 | 3 | 40 |
| Outdoor Scene | Consist of 2688 images. These images are divided into 8 groups. Each image is characterized by 4 (Monadjemi et al. 2002): GIST features (512), color moment features (432), HOG features (256), and LBP features (48) | 2688 | 4 | 8 |

**Table 3** (continued)

| Dataset | Description | Samples | Views | Clusters |
|---|---|---|---|---|
| MSRCv1 | Consists of 210 instances from Microsoft Research in Cambridge divided into 7 groups. Each image is character-ized by 5 views (Winn and Jojic 2005): GIST features (512), color moment features (24), CENTRIST features (254), SIFT features (512), and LBP features (256) | 210 | 5 | 7 |
| COVIDx[g] | Consists of 13892 samples divided into 5458 instances corresponding to the COVID19 class, and 7966 instances corresponding to the normal class. Each image is char-acterized by 3 views: ResNet50 (2048), ResNet101 (2048), and DenseNet169 feature vector (1664) | 13892 | 3 | 2 |

[a] https://data.caltech.edu/records/mzrjq-6wc02

[b] https://www.cs.tau.ac.il/wolf/ytfaces/

[c] http://mlg.ucd.ie/datasets/bbc.html

[d] https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

[e] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

[f] http://cam-orl.co.uk/facedatabase.html/

[g] https://github.com/lindawangg/COVID-Net/blob/master/docs/COVIDx.md

# 9 Future directions

Multi-view clustering has achieved significant progress in overcoming various challenges related to the integration of different data views. These advances have laid the foundation for robust methods and various applications, but much remains to be done to tackle unsolved problems and exploit new opportunities. The following key areas represent both milestones achieved and exciting prospects for future research:

- Integration with deep learning: Multi-view clustering has already benefited significantly from the integration of deep learning techniques. Methods such as autoencoders and graph neural networks have enabled more flexible and accurate learning of representations (Zhu et al. 2024). However, challenges such as the need for effective unsupervised architectures and scalable models remain. Future research should focus on developing architectures that exploit correlations between multiple views, taking into account advances in unsupervised deep learning and self-supervised paradigms.

- Adaptive and self-learning mechanisms: Recent developments have introduced adaptive models that are able to dynamically weight views based on their contributions, improving robustness to noisy or incomplete views (Wang et al. 2022). However, these methods often rely on hand-crafted assumptions. Future work could focus on self-learning frameworks that are able to automatically discover optimal representations and weighting strategies, and leverage advances in meta-learning and reinforcement learning to refine their adaptability.

- Scalability to large data sets: Progress has already been made in scaling multi-view clustering algorithms using techniques such as anchor-based methods and distributed computing systems (Wang et al. 2022; Chen et al. 2022). However, the exponential growth in data volume and complexity requires further innovation. Research needs to focus on lightweight algorithms with linear or sublinear complexity that are capable of processing streaming or real-time multi-view data in distributed environments.

- Heterogeneity of views: Existing multi-view clustering methods have addressed heterogeneity by using techniques such as kernel-based approaches and joint latent space learning (Li et al. 2022). Despite these efforts, dealing with different distributions, dimensions, and noise levels between views remains a challenge. Future research should explore robust frameworks that can dynamically adapt to heterogeneous data features while ensuring interpretability and effectiveness.

- Interpretability and explainability: Multi-view clustering models increasingly use latent representation learning to improve performance, but this often reduces transparency. To improve interpretability, frameworks such as Self-taught Multi-view Spectral Clustering (SMSC) (Zhong and Pun 2023) integrate manifold structure with clustering information to create a consensus similarity graph that matches the cluster representations with the input graphs. Similarly, Multi-view Clustering using Unified Graph Learning and Spectral Representation (MCGLSR) (Dornaika and El Hajjar 2023) ensures consistency between views while resolving cluster mappings directly. Despite these advances, it remains a challenge to develop fully interpretable multi-view clustering models. Future directions should focus on developing models that not only provide accurate clustering results, but also provide actionable insights into the relationships between the views and the clustered data.

Significant progress has been made in overcoming the main challenges of multi-view clustering, such as data integration and scalability. However, further progress in the areas of interpretability, heterogeneity and adaptive learning are crucial for further development. The integration of emerging trends such as deep learning, explainable AI and domain-specific adaptation will play a crucial role in improving the relevance and robustness of multi-view clustering methods for tackling complex real-world problems.

In recent years, domain-specific knowledge, especially in areas such as bioinformatics and medical diagnostics, has been successfully applied to improve clustering performance (Benton et al. 2019; Feng et al. 2022). Nevertheless, frameworks for the seamless integration of domain-specific priorities into clustering algorithms have not yet been sufficiently explored. Future research could focus on developing automated approaches to encode domain knowledge using tools such as knowledge graphs and ontology-based constraints to improve the accuracy and applicability of clustering.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

Aiping Huang, Weiling Chen, Tiesong Zhao, Wen Chen Chang (2021) Joint learning of latent similarity and local embedding for multi-view clustering. IEEE Trans Image Process 30:6772–6784

Al-kuhali HA, Shan M, Hael MA et al (2022) Multiview clustering of multi-omics data integration by using a penalty model. BMC Bioinform 23(1):288

Ao Li, Cong Feng, Yuan Cheng, Yingtao Zhang, Hailu Yang (2024) Incomplete multiview subspace clustering based on multiple kernel low-redundant representation learning. Inform Fusion 103:102086

Aparajita Khan, Pradipta Maji (2021) Multi-manifold optimization for multi-view subspace clustering. IEEE Trans Neural Netw Learn Syst 33:3895

Ben Yang, Xuetao Zhang, Zhiping Lin, Feiping Nie, Chen BD, Fei Wang (2022) Efficient and robust multiview clustering with anchor graph regularization. IEEE Trans Circuits Syst Video Technol 32:6200–6213

Benton Adrian, Khayrallah Huda, Gujral Biman, Reisinger Dee Ann, Zhang Sheng, Arora Raman (2019) Deep generalized canonical correlation analysis. In Isabelle Augenstein, Spandana Gella, Sebastian Ruder, Katharina Kann, Burcu Can, Johannes Welbl, Alexis Conneau, Xiang Ren, and Marek Rei (eds) Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019), Florence, Italy. Association for Computational Linguistics, pp 1–6

Blum Avrim, Mitchell Tom (1998) Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory (COLT'98). ACM, pp 92–100

Bruno Eric, Marchand-Maillet Stéphane (2009) Multiview clustering: a late fusion approach using latent models. pp 736–737 Cited by: 86

Cai Hao, Liu Bo, Xiao Yanshan, Lin LuYue (2020) Semi-supervised multi-view clustering based on ortho-normality-constrained nonnegative matrix factorization. Inf Sci 536:171–184

Cai Xianhua, Nie Feiping, Huang Heng (2013) Heterogeneous image feature integration via multi-modal multi-kernel learning. In Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML'13), vol 28, p 1309–1317

Cai Xiaojun, Nie Feiping, Huang Heng (2011) Heterogeneous image feature integration via multi-modal spectral clustering. In Proceedings of the 24th International Conference on Neural Information Processing Systems (NeurIPS), p 350–358

Chao Su, Haoliang Yuan, Lei Lai Loi, Qiang Yang (2023) Anchor-based multi-view subspace clustering with graph learning. Neurocomputing 547:126320

Chao Gao, Sun Shiliang, Bi Jinbo (2021) A survey on multi-view clustering. IEEE Trans Artif Intell 2(2):146–168

Chaudhuri Kamalika, Kakade Sham M, Livescu Karen, Sridharan Karthik (2009) Multi-view clustering via canonical correlation analysis. In Proceedings of the 26th Annual International Conference on Machine Learning, pp 129–136

Chen Jin, Huafu Xu, Xue Jingjing, Gao Quanxue, Deng Cheng, Lv Ziyu (2025) Incomplete multi-view clustering based on hypergraph. Inform Fusion 117:102804

Chen Man-Sheng, Huang Ling, Wang Chang-Dong, Huang Dong (2020) Multi-view clustering in latent embedding space. Proc AAAI Conf Artif Intell 34(04):3513–3520

Chen H, Liu X (2022) An improved multi-view spectral clustering based on tissue-like P systems. Sci Rep 12:18616

Chen Yongyong, Xiao Xiaolin, Zhou Yicong (2019) Jointly learning kernel representation tensor and affinity matrix for multi-view clustering. IEEE Trans Multimedia 22(8):1985–1997

Chen Man-Sheng, Wang Chang-Dong, Huang Dong, Lai Jian-Huang, Yu Philip S (2022) Efficient orthogonal multi-view subspace clustering. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22. Association for Computing Machinery, pp 127–135

Chen Peng, Liu Liang, Ma Zhengrui, Kang Zhao (2021) Smoothed multi-view subspace clustering. In International Conference on Neural Computing for Advanced Applications. Springer, pp 128–140

Chen Yongyong, Xiao Xiaolin, Peng Chong, Lu Guangming, Zhou Yicong (2021) Low-rank tensor graph learning for multi-view subspace clustering. IEEE Transactions on Circuits and Systems for Video Technology

Chen Kai, Huang Yujie, Wang Ziyuan (2024) A multiview approach to tracking people in crowded scenes using fusion feature correlation. Commun Comput Inform Sci 1911 CCIS:204–217

Chua Tat-Seng, Tang Jinhui, Hong Richang, Li Haojie, Luo Zhiping, Zheng Yantao (2009) Nus-wide: a real-world web image database from national university of singapore. In Proceedings of the ACM international conference on image and video retrieval, pp 1–9

Diallo Bassoma Hu, Jie Li Tianrui, Ahmad Khan Ghufran, Xinyan Liang, Hongjun Wang (2023) Auto-attention mechanism for multi-view deep embedding clustering. Pattern Recogn 143:109764

Dong Xiaowen, Frossard Pascal, Vandergheynst Pierre, Nefedov Nikolai (2013) Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. In 2013 IEEE Global Conference on Signal and Information Processing, pp 993–996

Dornaika F, El Hajjar S (2023) Single phase multi-view clustering using unified graph learning and spectral representation. Inform Sci 645:119366

Dornaika F, El Hajjar S (2023) Direct multi-view spectral clustering with consistent kernelized graph and convoled nonnegative representation. Artif Intell Rev 56(10):10987–11015

Dou Zengfa, Peng Nian, Hou Weiming, Xie Xianghua, Ma Xiaoke (2025) Learning multi-level topology representation for multi-view clustering with deep non-negative matrix factorization. Neural Netw 182:106856

El Hajjar Sally, Dornaika Fadi, Abdallah Fahed, Omrani Hichem (2022) Multi-view spectral clustering via integrating label and data graph learning. In International Conference on Image Analysis and Processing. Springer, pp 109–120

Fang Uno, Li Man, Li Jianxin, Gao Longxiang, Jia Tao, Zhang Yanchun (2023) A comprehensive survey on multi-view clustering. IEEE Trans Knowl Data Eng 35(12):12350–12368

Fang Z, Yang X, Han L, Liu X (2019) A sequentially truncated higher order singular value decomposition-based algorithm for tensor completion. IEEE Trans Cyber 49(5):1956–1967

Fei-Fei Li, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In 2004 Conference on Computer Vision and Pattern Recognition Workshop, pp 178–178

Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Bing Xu, Warde-Farley David, Ozair Sherjil, Courville Aaron, Bengio Yoshua (2020) Generative adversarial networks. Commun ACM 63(11):139–144

Goodfellow Ian, Bengio Yoshua, Courville Aaron, Bengio Yoshua (2016) Deep learning, vol 1. MIT press Cambridge

Greenacre M, Groenen PJF, Hastie T et al (2022) Principal component analysis. Nat Rev Methods Primers 2:100

Guojun Du, Zhou Lei, Yang Yuchao et al (2021) Deep multiple auto-encoder-based multi-view clustering. Data Sci Eng 6:323–338

Guoqing Chao, Sun Jiangwen Lu, Jin Wang An-Li, Langleben Daniel D, Chiang-Shan Li, Jinbo Bi (2019) Multi-view cluster analysis with incomplete data to understand treatment effects. Inf Sci 494:278–293

Gupta Pooja, Goel Anurag, Majumdar Angshul, Chouzenoux Emilie, Chierchia Giovanni (2024) Deconfcluster: Deep convolutional transform learning based multiview clustering fusion framework. Signal Process 224:109597

El Hajjar S, Dornaika F, Abdallah F (2021) Multi-view spectral clustering via constrained nonnegative embedding. Information Fusion 78:209

Horie Mitsuhiko, Kasai Hiroyuki (2021) Consistency-aware and inconsistency-aware graph-based multi-view clustering. In 2020 28th European Signal Processing Conference (EUSIPCO). IEEE, pp 1472–1476

Hu Ghufran Ahmad Khan Jie, Li Tianrui, Diallo Bassoma, Shengdong Du (2023) Multi-view subspace clustering for learning joint representation via low-rank sparse representation. Appl Intell 53:22511

James IM (1976) The topology of stiefel manifolds. Cambridge books online, Cambridge University Press

Jiang Meng (2023) Transfer learning across graph convolutional networks: methods, theory, and applications. ACM Trans Knowl Discov Data 18(1):1–23

Jiayu Li, Ding Chris HQ, Licheng Tang (2018) Multi-view clustering via joint non-negative matrix factorization. Knowl-Based Syst 144:159–171

Jie Chen, Zhu Wang, Hua Mao, Xi Peng (2023) Low-rank tensor learning for incomplete multiview clustering. IEEE Trans Knowl Data Eng 35(11):11556–11569

Jinhong Yu, Kun Sun, Kunqian Li, Chuan Tang, Ruyi Feng (2023) Towards accurate image matching by exploring redundancy between multiple descriptors. IEEE Geosci Remote Sens Lett 20:1–5

Kang Zhao, Lin Zhiping, Zhu Xiaofeng, Wenbo Xu (2021) Structured graph learning for scalable subspace clustering: from single view to multiview. IEEE Trans Cybern 52:8976–8986

Khan Ghufran Ahmad, Hu Jie, Li Tianrui, Diallo Bassoma, Wang Hongjun (2022) Multi-view data clustering via non-negative matrix factorization with manifold regularization. International Journal of Machine Learning and Cybernetics, pp 1–13

Kim So Yeon, Sohn Kyung-Ah (2017) Multi-view network-based social-tagged landmark image clustering. In 2017 IEEE International Conference on Image Processing (ICIP), pp 3680–3684

Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Rev 51(3):455–500

Kumar Mahesh P, Udupa Raghavendra (2011) A co-regularized approach to multi-view spectral clustering. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11), vol 1, p 33

Lampert Christoph H, Hannes Nickisch, Stefan Harmeling (2013) Attribute-based classification for zero-shot visual object categorization. IEEE Trans Pattern Anal Mach Intell 36(3):453–465

Lan Chao, Yang Yuhao, Li Xiaoli, Luo Bo, Huan Jun (2017) Learning social circles in ego-networks based on multi-view network structure. IEEE Trans Knowl Data Eng 29(8):1784–1797

LeCun Yann, Bengio Yoshua, Hinton Geoffrey (2015) Deep learning. Nature 521(7553):436–444

Lele Fu, Lin Pengfei, Vasilakos Athanasios V, Wang Shiping (2020) An overview of recent multi-view clustering. Neurocomputing 402:148–161

Li Jingwen Wu, Wei Zhang Dan, Dayong Fan, Jiang Jianwu Lu, Yanling Gao Ertao, Tao Yue (2023) Multi-pedestrian tracking based on KC-Yolo detection and identity validity discrimination module. Appl Sci 13(22):12228

Li Kai, Liu Hongfu, Zhang Yulun, Li Kunpeng, Yun Fu (2022) Self-guided deep multiview subspace clustering via consensus affinity regularization. IEEE Trans Cybern 52(12):12734–12744

Li Miaomiao, Liu Xinwang, Zhang Yi, Liang Weixuan (2024) Late fusion multiview clustering via min–max optimization. IEEE Trans Neural Netw Learn Syst 35(7):9417–9427

Li Wang, Ren-Cang Li, Wen-Wei Lin (2023) Multiview orthonormalized partial least squares: regularizations and deep extensions. IEEE Trans Neural Netw Learn Syst 34(8):4371–4385

Li Jing, Wang Qianqian, Yang Ming, Gao Quanxue, Gao Xinbo (2024) Efficient anchor graph factorization for multi-view clustering. IEEE Trans Multimedia 26:5834–5845

Li Miaomiao, Liu Xinwang, Wang Lei, Dou Yong, Yin Jianping, En Zhu (2016) Multiple kernel clustering with local kernel alignment maximization. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). IJCAI/AAAI

Li Lin, Zhou Xiaojun, Lu Zhiqiang, Li Dongxiao, Xu Qinxu, Song Li (2022) Joint learning of latent representation and global similarity for multi-view image clustering. In Proc. SPIE 12083, Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021), p 120830H

Liang Weixuan, Zhou Sihang, Xiong Jian, Liu Xinwang, Wang Siwei, Zhu En, Cai Zhiping, Xin Xu (2022) Multi-view spectral clustering with high-order optimal neighborhood Laplacian matrix. IEEE Trans Knowl Data Eng 34(7):3418–3430

Liang Youwei, Huang Dong, Wang Chang-Dong (2019) Consistency meets inconsistency: A unified graph learning framework for multi-view clustering. In *2019 IEEE International Conference on Data Mining (ICDM)*, pp 1204–1209

Lijuan Wang, Lin Zhang, Ming Yin, Zhifeng Hao, Ruichu Cai, Wen Wen (2022) Double embedding-transfer-based multi-view spectral clustering. Exp Syst Appl 210:118374

Lin Renjie, Shide Du, Wang Shiping, Guo Wenzhong (2023) Multi-view clustering via optimal transport algorithm. Knowl-Based Syst 279:110954

Liu X et al (2020) Efficient and effective regularized incomplete multi-view clustering. IEEE Trans Pattern Anal Mach Intell 43(8):2634–2646

Liu Bing, Anzhu Yu, Xuchu Yu, Wang Ruirui, Gao Kuiliang, Guo Wenyue (2021) Deep multiview learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 59(9):7758–7772

Liu Jun, Ji Shuiwang, Ye Jie (2013) Sleec: Sparse learning with efficient euclidean constraints for low-resource multiview learning. In Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML'13), vol 28, p 1491–1499

Liu Jun, Ji Shuiwang, Ye Jie (2013) Sleec: Sparse learning with efficient euclidean constraints for low-resource multiview learning. In Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML'13), vol 28, pp 1491–1499

Liu Z, Chen C, Yang X, Zhou J, Li X, Song L (2018) Heterogeneous graph neural networks for malicious account detection. In CIKM, pp 2077–2085

Lu Canyi, Yan Shuicheng, Lin Zhouchen (2015) Convex sparse spectral clustering: single-view to multi-view. CoRR, abs/1511.06860

Luo Shirui, Cao Xiaochun (2022) Multiview subspace dual clustering. IEEE Trans Neural Netw Learn Syst 33(12):7425–7437

Lv Z, Gao Q, Zhang X, Li Q, Yang M (2022) View-consistency learning for incomplete multiview clustering. IEEE Trans Image Process 31:4790–4802

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

Manuel Graña, Igone Morais-Quilez (2023) A review of graph neural networks for electroencephalography data analysis. Neurocomputing 562:126901

Monadjemi Amir, Thomas BT, Mirmehdi Majid (2002) Experiments on high resolution images towards outdoor scene classification. Technical Rep

Moujahid Abdelmalik, Dornaika Fadi (2019) A pyramid multi-level face descriptor: application to kinship verification. Multimedia Tools Appl 78:9335–9354

Moujahid Abderrahim, Dornaika Fadi (2020) Multi-scale multi-block covariance descriptor with feature selection. Neural Comput Appl 32:6283–6294

Muhammad Haris, Yusliza Yusoff, Mohd Zain Azlan, Saeed Khattak Abid, Fawad Hussain Syed (2024) Breaking down multi-view clustering: a comprehensive review of multi-view approaches for complex data structures. Eng Appl Artif Intell 132:107857

Niu Guoli, Yang Youlong, Sun Liqin (2021) One-step multi-view subspace clustering with incomplete views. Neurocomputing 438:290–301

Pan Erlin, Kang Zhao (2023) High-order multi-view clustering for generic data. Inform Fusion 100:101947

Pfeifer Bastian, Bloice Marcus D, Schimek Michael G (2023) PAREA: Multi-view ensemble clustering for cancer subtype discovery. J Biomed Inform 143:104406

Pierre De Handschutter, Nicolas Gillis, Xavier Siebert (2021) A survey on deep matrix factorizations. Comput Sci Rev 42:100423

Pucci A, Gori M, Hagenbuchner M, Scarselli F, Tsoi A (2006) Investigation into the application of graph neural networks to large-scale recommender systems. Syst Sci 32(4):17–26

Qin Yalan, Zhang Xinpeng, Shui Yu, Feng Guorui (2025) A survey on representation learning for multi-view data. Neural Netw 181:106842

Qiyuan Ou, Wang Siwei, Zhou Sihang, Li Miaomiao, Guo Xifeng, Zhu En (2020) Anchor-based multiview subspace clustering with diversity regularization. IEEE Multimedia 27(4):91–101

Rebafka Tabea (2024) Model-based clustering of multiple networks with a hierarchical algorithm. Stat Comput 34(1):32

Ren Zhenwen, Li Haoran, Yang Chao, Sun Quansen (2020) Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning. Knowl-Based Syst 188:105040

Rushi Lan, Yicong Zhou, Yan Tang Yuan (2015) Quaternionic local ranking binary pattern: a local descriptor of color images. IEEE Trans Image Process 25(2):566–579

Serra Angela, Galdi Paola, Tagliaferri Roberto (2019) Chapter 13 - multiview learning in biomedical applications. In Robert Kozma, Cesare Alippi, Yoonsuck Choe, and Francesco Carlo Morabito, editors, Artificial Intelligence in the Age of Neural Networks and Brain Computing. Academic Press, pp 265–280

Shao-Yuan Li, Yuan Jiang, Zhi-Hua Zhou (2014) Partial multi-view clustering. Proc AAAI Conf Artif Intell. https://doi.org/10.1609/aaai.v28i1.8973

Shao Weixiang, He Lifang, Lu Chun-ta, Philip S Yu (2016) Online multi-view clustering with incomplete views. In 2016 IEEE International conference on big data (Big Data). IEEE, pp 1012–1017

Shao Weixiang, He Lifang, Yu Philip S (2015) Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization. In Joint European conference on machine learning and knowledge discovery in databases. Springer, pp 318–334

Shudong Huang, Zhao Kang, Tsang Ivor W, Zenglin Xu (2019) Auto-weighted multi-view clustering via kernelized graph learning. Pattern Recogn 88:174–184

Sindhwani Vikas, Bhattacharya, Abhishek, Rakshit, Soumen: Information theoretic co-clustering. In Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI'08), pp 504–511 (2008)

Song Shuran, Lichtenberg Samuel P, Xiao Jianxiong (2015) Sun RGB-d: A RGB-d scene understanding benchmark suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 567–576

Sun Baochen, Feng Jiashi, Saenko Kate (2013) Return of frustratingly easy domain adaptation. In Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML'13), vol 28, pp 897–904

Sura Raya, Mariam Orabi, Imad Afyouni, Al Aghbari Zaher (2024) Multi-modal data clustering using deep learning: a systematic review. Neurocomputing 607:128348

Tan Junpeng, Shi Yukai, Yang Zhijing, Wen Caizhen, Lin Liang (2020) Unsupervised multi-view clustering by squeezing hybrid knowledge from cross view and each view. IEEE Trans Multimedia 23:2943–2956

Tan Xiaoyang, Triggs Bill (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650

Tang Kewei Xu, Zhixun Kaiqiang Su, Nan Zhang (2023) Multi-view subspace clustering via consistent and diverse deep latent representations. Inform Sci 651:119719

Tang Chang, Zhu Xinzhong, Liu Xinwang, Li Miaomiao, Wang Pichao, Zhang Changqing, Wang Lizhe (2018) Learning a joint affinity graph for multiview subspace clustering. IEEE Trans Multimed 21(7):1724–1736

Teepe Torben, Wolters Philipp, Gilg Johannes, Herzog Fabian, Rigoll Gerhard (2024) Lifting Multi-View Detection and Tracking to the Bird's Eye View. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 667–676, Los Alamitos, CA, USA, June 2024. IEEE Computer Society

Tucker LR (1966) Some mathematical notes on three-mode factor analysis. Psychometrika 31(3):279–311

Wang Hengkang, Han Lu, Sun Ju, Safo Sandra E (2024) Interpretable deep learning methods for multiview learning. BMC Bioinform 25(1):69

Wang Jiao Wu, Bin Ren Zhenwen, Yunhui Zhou (2024) Decomposed deep multi-view subspace clustering with self-labeling supervision. Information Sciences 653:119798

Wang Yang, Lin Wu, Lin Xuemin, Gao Junbin (2018) Multiview spectral clustering via structured low-rank matrix factorization. IEEE Trans Neural Netw Learn Syst 29(10):4833–4843

Wang Yang, Lin Xuemin, Lin Wu, Zhang Wenjie, Zhang Qing, Huang Xiaodi (2015) Robust subspace clustering for multi-view data by exploiting correlation consensus. IEEE Trans Image Process 24(11):3939–3949

Wang Siwei, Liu Xinwang, Zhu Xinzhong, Zhang Pei, Zhang Yi, Gao Feng, Zhu En (2022) Fast parameter-free multi-view subspace clustering with consensus anchor guidance. IEEE Trans Image Process 31:556–568

Wang Hao, Yang Yan, Liu Bing (2020) GMC: Graph-based multi-view clustering. IEEE Trans Knowl Data Eng 32(6):1116–1129

Wang Hao, Yang Yan, Liu Bing, Fujita Hamido (2019) A study of graph-based system for multi-view clustering. Knowl-Based Syst 163:1009–1019

Wang Hao, Zong Linlin, Liu Bing, Yang Yan, Zhou Wei (2019) Spectral perturbation meets incomplete multi-view data. In International Joint Conference on Artificial Intelligence, pp 3677–3683

Wang Siwei, Liu Xinwang, Liu Li, Zhou Sihang, En Zhu (2021) Late fusion multiple kernel clustering with proxy graph refinement. IEEE Transactions on Neural Networks and Learning Systems

Wang Xingchao, Zhao Weidi, Wu Jiashi (2018) Multi-view generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV'18)

Wen Jie, Yan Ke, Zhang Zheng, Yong Xu, Wang Junqian, Fei Lunke, Zhang Bob (2020) Adaptive graph completion based incomplete multi-view clustering. IEEE Trans Multimedia 23:2493–2504

Wen Jie, Zhang Zheng, Fei Lunke et al (2022) A survey on incomplete multiview clustering. IEEE Trans Syst Man Cyber Syst 53(2):1136–1149

Winn John, Jojic Nebojsa (2005) Locus: Learning object classes with unsupervised segmentation. In Tenth IEEE International Conference on Computer Vision (ICCV'05), vol 1. IEEE, pp 756–763

Wu S, Tang Y, Zhu Y, Wang L, Xie X, Tan T (2019) Session-based recommendation with graph neural networks. In AAAI 33:346–353

Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, Tianyi Joey Zhou (2019) COMIC: Multi-view clustering without parameter selection. Int Conf Mach Learn

Xia Wei, Gao Quanxue, Wang Qianqian, Gao Xinbo (2022) Tensor completion-based incomplete multiview clustering. IEEE Trans Cybern 52(12):13635–13644

Xia Rong, Pan Yun, Lai Hanjiang, Liu Congcong (2014) Multi-view spectral clustering with high-order graphs. IEEE Trans Cyber 44(9):1619–1631

Xiangyu Liu, Peng Song (2022) Incomplete multi-view clustering via virtual-label guided matrix factorization. Exp Syst Appl 210:118408

Xiao Yu, Liu Hui, Yan Wu, Zhang Caiming (2021) Fine-grained similarity fusion for multi-view spectral clustering. Inf Sci 568:350–368

Xie Deyan, Zhang Xiangdong, Gao Quanxue, Han Jiale, Xiao Song, Gao Xinbo (2020) Multiview clustering by joint latent representation and similarity learning. IEEE Trans Cybern 50(11):4848–4854

Xin Xu, Tian Yuan (2015) A survey of multi-view machine learning. Neural Comput Appl 26(3):665–681

Xing L, Chen B, Du S, Gu Y, Zheng N. (2019) Correntropy-based multiview subspace clustering. IEEE Trans Cybern, pp 1–14

Xingyu Xu, Zhao Jie, Wang Jianqiu, Wang Yebin (2019) Multi-view spectral embedding. Pattern Recogn 94:88–98

Xinyu Pu, Baicheng Pan, Hangjun Che (2023) Robust low-rank graph multi-view clustering via Cauchy norm minimization. Mathematics 11(13):2940

Xu Chang, Tao Dacheng, Xu Chao (2013) A survey on multi-view learning

Yan Yan, Zhang Zhen, Yang Jie (2016) Multi-view learning: theoretical foundations and empirical study. IEEE Trans Cyber 46(12):2725–2738

Yang Yan, Wang Hao (2018) Multi-view clustering: a survey. Big Data Min Anal 1(2):83–107

Yi Xu, Niu Guoqing (2024) Research on multi-view clustering algorithm based on sequential three-way decision. Appl Soft Comput 158:111590

Yin Jun, Cai Runcheng, Sun Shiliang (2022) Anchor-based incomplete multi-view spectral clustering. Neurocomputing 514:526–538

Yin Hongwei, Wenjun Hu, Li Fanzhang, Lou Jungang (2021) One-step multi-view spectral clustering by learning common and specific nonnegative embeddings. Int J Mach Learn Cybern 12(7):2121–2134

Yixue Feng, Mansu Kim, Xiaohui Yao, Kefei Liu, Qi Long, Li Shen (2022) Deep multiview learning to identify imaging-driven subtypes in mild cognitive impairment. BMC Bioinform 23:402

Yu-Ling Hsueh, Wen-Nung Lie, Guan-You Guo (2020) Human behavior recognition from multiview videos. Inf Sci 517:275–296

Yuan Changan, Zhong Zhi, Lei Cong, Zhu Xiaofeng, Rongyao Hu (2021) Adaptive reverse graph learning for robust subspace learning. Inform Process Manag 58(6):102733

Yuan Changan, Zhu Yonghua, Zhong Zhi, Zheng Wei, Zhu Xiaofeng (2022) Robust self-tuning multi-view clustering. World Wide Web 25(2):489–512

Zhan K, Zhang C, Guan J, Wang J (2018) Graph learning for multiview clustering. IEEE Trans Cybern 48(10):2887–2895

Zhan Qianyi, Hu Wei (2020) An epilepsy detection method using multiview clustering algorithm and deep features. Comput Math Methods Med 202:512

Zhang Han, Danyang Wu, Nie Feiping, Wang Rong, Li Xuelong (2021) Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection. Inform Fusion 70:129–140

Zhang Xiang, Wang Shuhui, Tao Dacheng (2017) Adaptive similarity metric fusion for multi-view clustering. IEEE Trans Cybern 47(7):1746–1759

Zhang Xiaowei, Zheng Zhiming, Gao Dongming et al (2023) Multi-view consistent generative adversarial networks for compositional 3D-aware image synthesis. Int J Comput Vision 131(11):2219–2242

Zhang Changqing, Hu Qinghua, Fu Huazhu, Zhu Pengfei, Cao Xiaochun (2017) Latent multi-view subspace clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4279–4287

Zhanxuan Hu, Nie Feiping, Chang Wei, Hao Shuzheng, Wang Rong, Li Xuelong (2020) Multi-view spectral clustering via sparse graph learning. Neurocomputing 384:1–10

Zhanxuan Hu, Nie Feiping, Wang Rong, Li Xuelong (2020) Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. Inform Fusion 55:251–259

Zhao Kang, Guoxin Shi, Shudong Huang, Chen Wenyu Pu, Xiaorong Zhou Joey, Zenglin Tianyi Xu (2020) Multi-graph fusion for multi-view spectral clustering. Knowl-Based Syst 189:105102

Zhao Xingwang, Wang Shujun, Liu Xiaolin, Liang Jiye (2024) Multi-view clustering via dynamic unified bipartite graph learning. Pattern Recogn 156:110715

Zhao Handong, Liu Hongfu, Fu Yun (2016) Incomplete multi-modal visual data grouping. In IJCAI, pp 2392–2398

Zheng Dacheng, Zhiwen Yu, Chen Wuxing, Zhang Weiwen, Feng Qiying, Shi Yifan, Yang Kaixiang (2024) Multiview ensemble clustering of hypergraph p-Laplacian regularization with weighting and denoising. Inf Sci 681:121187

Zhong Guo, Pun Chi-Man (2023) Self-taught multi-view spectral clustering. Pattern Recogn 138:109349

Zhou T, Changqing Zhang, Xi Peng, Bhaskar H, Yang J (2020) Dual shared-specific multiview subspace clustering. IEEE Trans Cybern 50:3517–3530

Zhou Peng, Liu Xinwang, Liang Du, Li Xuejun (2023) Self-paced adaptive bipartite graph learning for consensus clustering. ACM Trans Knowl Discov Data 17(5):1–35

Zhou Shibing, Wang Xi, Yang Mingrui, Song Wei (2025) Multi-view clustering with adaptive anchor and bipartite graph learning. Neurocomputing 611:128627

Zhou Jie, Zhang Runxin (2024) A weighted multi-view clustering via sparse graph learning. Clust Comput 27(10):13517–13530

Zhou Zhi-Hua, Wang Min-Ling (2007) Multi-instance multi-label learning with application to scene classification. In Advances in Neural Information Processing Systems (NeurIPS), pp 1609–1616

Zhu Pengfei, Xinjie Yao Yu, Wang Binyuan Hui, Dawei Du, Qinghua Hu (2024) Multiview deep subspace clustering networks. IEEE Trans Cybern 54(7):4280–4293

Zhu Xiaofeng, Zhang Shichao, He Wei, Rongyao Hu, Lei Cong, Zhu Pengfei (2018) One-step multi-view spectral clustering. IEEE Trans Knowl Data Eng 31(10):2022–2034

Zhu H, Klus S, Sahai T (2022) A dynamic mode decomposition approach for decentralized spectral clustering of graphs. In Proceedings of the IEEE Conference on Control Technology and Applications (CCTA), p 1202–1207

Zonghan Shi, Haitao Zhao (2023) Deep multi-view clustering based on reconstructed self-expressive matrix. Appl Sci (Switzerland) 13(15):8791

Zonghan Wu, Pan Shirui, Chen Fengwen, Long Guodong, Zhang Chengqi, Philip SYu (2021) A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 32(1):4–24