



ELSEVIER

Contents lists available at ScienceDirect

Information Sciences

journal homepage: [www.elsevier.com/locate/ins](http://www.elsevier.com/locate/ins)

# Adaptive multi-channel contrastive graph convolutional network with graph and feature fusion

Luying Zhong<sup>a,b</sup>, Jielong Lu<sup>a,b</sup>, Zhaoliang Chen<sup>a,b</sup>, Na Song<sup>c</sup>, Shiping Wang<sup>a,d,\*</sup>

<sup>a</sup> College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China

<sup>b</sup> Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China

<sup>c</sup> School of Mechanical, Electrical and Information Engineering, Putian University, Putian 351100, China

<sup>d</sup> Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen 518172, China

## ARTICLE INFO

### Keywords:

Multi-view learning  
Graph convolutional network  
Semi-supervised classification  
Deep interactive information  
Graph fusion  
Contrastive learning

## ABSTRACT

The domain of multi-view semi-supervised classification is an appealing topic in real-world applications. Due to the powerful capability of gathering information from neighbors, Graph Convolutional Network (GCN) has become a hotspot in the classification task. However, most of multi-view classification works based on GCN only assign weights for feature fusion, and directly consider the weighted sum of the adjacency matrices, ignoring the interaction and correlation among features. These may be problematic since aggregating the matrices from less relevant views may destroy the original topology space, leading to undesired performance. To tackle the aforementioned challenges, this paper presents an Adaptive Multi-Channel Graph Convolutional Network (AMC-GCN). To extract the interactive information, AMC-GCN designs a deep interactive feature integration network to incorporate consensus and complementary information. To fuse the graph structures, AMC-GCN exploits the relevance between views and imposes an adjacency matrix fusion network on constructing multiple GCN channels, thereby delivering discriminative information on graphs. To enhance the homogeneity of the framework, AMC-GCN applies a contrastive loss to joint learning during the optimization for classification. With these considerations, AMC-GCN exploits relevant and interactive information between views to promote graph and feature fusion. Substantial experimental results on real-world datasets verify the superiority of AMC-GCN.

## 1. Introduction

In various real-world situations, multi-view data arises for the description of a large amount of data from diverse perspectives. Existing multi-view learning algorithms have achieved significant success across a wide spectrum of applications, including pattern recognition [1], [2], [3], computer vision [4], [5], [6], genetics [7], [8], [9] and data mining [10], [11], [12] etc. Numerous

\* Corresponding author at: College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China.

E-mail addresses: [luyingzhongfzu@163.com](mailto:luyingzhongfzu@163.com) (L. Zhong), [jielonglu2022@163.com](mailto:jielonglu2022@163.com) (J. Lu), [chenzl@outlook.com](mailto:chenzl@outlook.com) (Z. Chen), [ptcsn@126.com](mailto:ptcsn@126.com) (N. Song), [shipingwangphd@163.com](mailto:shipingwangphd@163.com) (S. Wang).

<https://doi.org/10.1016/j.ins.2023.120012>

Received 28 November 2022; Received in revised form 27 September 2023; Accepted 15 December 2023

Available online 21 December 2023

0020-0255/© 2023 Elsevier Inc. All rights reserved.

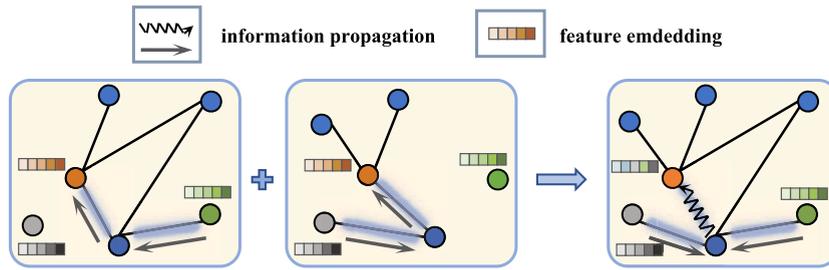


Fig. 1. Description of the importance of topological homogeneity. The gray node and green node are a pair of heterogeneous nodes, indicating that the feature information is dissimilar. For each node (orange), a direct weighted sum of topology graphs may aggregate green node and gray node information after graph fusion, leading to inconsistent feature information aggregated by the orange node, thereby obtaining undesirable embedding representation.

graph-oriented methods have been widely explored in multi-view semi-supervised classification learning. Especially, multi-view classification is broadly applied to intelligent recognition systems, such as scene interpretation, objective recognition, and visual search. Many researchers have investigated the benefits of multi-view features by exploring the consistent information across different views [13], [14], [15], which demonstrated superior performance compared to using single-view features.

Besides, considering that graph convolutional network (GCN) [16] has the powerful ability to integrate feature attributes from neighbors, GCN has been widely explored in a multitude of practical applications. GCN successfully extends the convolutional neural network from Euclidean to non-Euclidean domains. It has been applied to model graph-structured data such as affinity matrices [17], [18] and citation networks [19], [20], [21]. Due to the powerful representation capabilities, prior works [16], [22] introduced GCN into multi-view learning, which validates that GCN can adaptively learn the spectral information from other complementary views. GCN naturally integrates feature information from topology graphs of each convolutional layer, thereby promoting the propagation of valid messages between neighbors in different views.

Although prior works [23], [22] have been successfully applied GCN to the multi-view classification task, limited study has concentrated on the relevance of different views, leading to suboptimal fusion of graphs and features for multi-view data. Moreover, [23], [22] only applied a direct weighted combination of the adjacency matrices via GCN to downstream tasks. However, as illustrated in Fig. 1, the topology graphs from different views represent distinct node relationships. A linear weighted sum of topology graphs without considering the relevance between views, may compromise the homogeneity of adjacency matrices, thereby causing the extraction of undesirable topology structures. In addition, these GCN-based methods propagated the original feature views separately via topology structures, omitting the potential interactive feature information. This is problematic because only propagating view-specific feature information and ignoring the interactive view information may be not enough for nodes to learn intact feature representations, thereby leading to undesired classification performance. Based on these, it is natural to consider exploring the relations of views and extracting potential interactive information between views to integrate similar connections for learning consistent node embeddings.

To cope with these issues, this paper proposes an Adaptive Multi-channel Contrastive Graph Convolutional Network (AMC-GCN) to jointly explore the relevance of graphs and interactive information between views for semi-supervised classification learning. As outlined in Fig. 2, AMC-GCN includes three primary modules. First, a deep interactive feature integrating network (DIF-network) is utilized to extract both view-specific information and deep interactive information from various views. It applies several neural networks to capture specific information from different views, and then obtains potential interactive information on the view-specific attribute associations. Second, to integrate the relevant connection relationships, AMC-GCN explores the relations of views by designing an adaptive division strategy and introduces an adjacency matrix fusion network to integrate the relevant adjacency matrices of multiple graph channels. Finally, a multi-channel GCN network with semi-supervised contrastive loss (SC-network) is used to enhance the consistency among distinct topology graphs and perform the classification task. Overall, the main contributions of this study are outlined as follows:

- Propose a multi-channel graph convolutional network for capturing homogeneous node relationships and an intact feature representation. Specifically, the semi-supervised contrastive loss aims to enhance the graph consistency.
- Design an adjacency matrix fusion network with an adaptive division strategy to construct multiple graph channels by exploring the relevance between views and integrating topology graphs.
- Extract deep interactive information via a feature integration network, which enriches the underlying representations between views.
- Extensive experimental results on standard benchmark datasets demonstrate that the proposed method surpasses the performance of other state-of-the-art graph-based methods in multi-view semi-supervised classification.

The rest of this paper is structured as follows. Section 2 reviews the related works. The proposed AMC-GCN framework is elaborated in Section 3. Section 4 presents substantial experiments to validate the effectiveness of AMC-GCN. Finally, Section 5 concludes the whole paper.

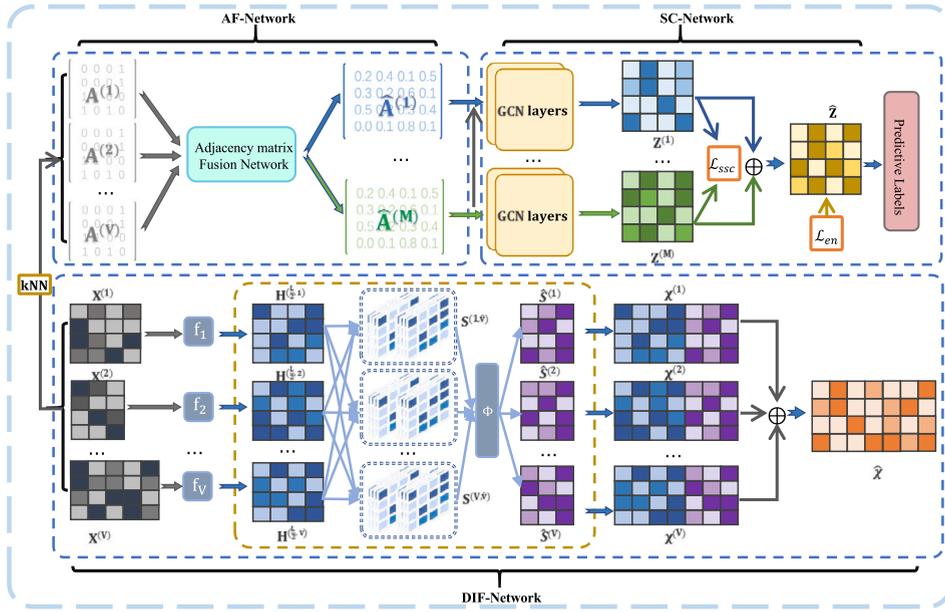


Fig. 2. Architecture of the proposed AMC-GCN consisting of three modules. DIF-Network aims to extract view-specific and interactive information. AF-Network is designed to integrate the topology information to obtain multiple graph channels. SC-Network maximizes the consistency of distinct graphs.

## 2. Related work

Due to the powerful capability of GCN and the wide application of multi-view data, GCN-based multi-view semi-supervised classification has become increasingly demanding. This section reviews some notable research works on GCNs and multi-view learning.

### 2.1. Graph convolutional network

In recent years, graph convolutional network has gained popularity in the field of machine learning. Spectral convolutional networks [24] often utilize the Fourier domain to process convolutional operations on a graph signal  $\mathbf{s} \in \mathbb{R}^n$  and then apply a spectral filter  $\mathbf{g}_\theta = \text{diag}(\theta)$  to its spectral components, that is

$$\mathbf{g}_\theta \cdot \mathbf{s} = \mathbf{U} \mathbf{g}_\theta \mathbf{U}^T \mathbf{s}, \quad (1)$$

where  $\cdot$  is the graph convolutional operation, and  $\mathbf{U}$  denotes the spectral decomposition of the normalized Laplacian operator. In order to simplify the feature decomposition complexity, Hammond et al. [25] employed Chebyshev polynomials to approximate the convolutional kernel of the spectral domain. To alleviate the problem of overfitting on the local graph structure, Kipf et al. [16] utilized truncated Chebyshev polynomials and performed the first-order approximation for graph convolutional networks, formulated as

$$\mathbf{Z} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \Theta), \quad (2)$$

where  $\sigma(\cdot)$  refers to an activation function,  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$  represents the adjacency matrix with self-loops,  $\tilde{\mathbf{D}}_{ii} = \sum_{j=1}^n \tilde{\mathbf{A}}_{ij}$  is the degree matrix from  $\tilde{\mathbf{A}}$ , and  $\Theta$  stands for the trainable parameter matrix.

As GCN has shown the superior performance, numerous GCN variants [26], [27], [28] have been developed. Xu et al. [29] proposed a novel answer-centric graph convolutional network with radial filters for visual processing tasks. Zhong et al. [30] introduced a graph-based contrastive framework that transformed the traditional instance-level conformity to the clustering tasks. Guo et al. [31] developed a GCN-based approach to distribute information through the affinity matrix for the incorporation of correlated objects in the scene graph generation tasks. AM-GCN [32] and CG<sup>3</sup> [33] established GCN-based methods that conveyed the valid message from neighbors to solve the limited supervision problem on node classification tasks. Besides, prior work Co-GCN [22] imposed a GCN-based model to cope with semi-supervised node classification on multi-view data. This approach utilized adaptive combined graph Laplacian matrices to explore graph information from views and optimized them through a co-training strategy. This paper also resorts to the GCN to cope with multi-view semi-supervised classification. In contrast to these GCN-based methods, AMC-GCN considers both the topology and feature fusion and the relevance of different views, while [22] only adopted a direct weighted sum of the topology matrices and then was applied to GCNs, thereby leading to undesirable connections and an incomplete representation with ignoring these problems.

**Table 1**

A summary of primary notations in this paper.

Notations	Explanations	Notations	Explanations
$\mathbf{X}^{(v)}, \mathbf{A}^{(v)}$	Feature and adjacency matrices of the $v$ -th view.	$\mathcal{X}^{(v)}$	View-specific data.
$\mathbf{\Omega}, \mathbf{Y}$	Labeled sample set and incomplete label matrix.	$\hat{\mathbf{A}}^{(m)}$	Fused adjacency matrix of the $m$ -th channel.
$f_v(\cdot)$	The $v$ -th fully connected network.	$\mathbf{C}$	Correlation matrix.
$\mathbf{W}_l^{(v)}, \mathbf{b}_l^{(v)}, \mathbf{W}_\Phi^{(v)}, \mathbf{b}_\Phi^{(v)}$	Weights and biases of fully connected networks.	$\alpha^{(v)}$	Weight of the $v$ -th adjacency matrix.
$\mathbf{S}^{(v, \hat{v})}$	Interactive matrix of the $v$ -th and $\hat{v}$ -th views.	$\pi^{(v)}$	Weight of specific feature data.
$\mathbf{H}^{\lfloor \frac{L}{2}, v \rfloor}$	Specific feature information in the $v$ -th view.	$\mathbf{Z}^{(m)}$	Output of the $m$ -th channel GCN.
$\mathbf{G}^{(v, \hat{v})}$	Interactive information the $v$ -th and $\hat{v}$ -th views.	$\mathcal{L}_{uc}, \mathcal{L}_{sc}$	Unsupervised and supervised losses.
$\mathbf{S}^v$	Interactive data in the $v$ -th view.	$\mathcal{L}_{en}$	Cross-entropy loss.

## 2.2. Multi-view learning

As data exhibits complex and diverse forms, multi-view learning [34], [35], [36] has gained significant attention in various machine learning tasks. Recently, there has been a surge of interest in developing multi-view classification methods [37], [38], [39]. Especially, [40] proposed a deep multi-view learning model that learned a discriminant and shared view-invariant representation between multiple views. Li et al. [41] designed a multi-view generative framework that simultaneously learned to fuse features from multiple views for classification. [42] introduced a neural network model, which is designed to learn sparse regularizers in a data-driven manner. Yang et al. [43] employed the non-negative matrix factorization algorithm to compress the distribution of multi-view data for preserving the geometric structure of each view. To effectively leverage both the consensus and complementary characteristics of multi-view data, Jia et al. [44] proposed a method that integrated independence and adversarial similarities to enhance the discriminability and reduce the redundancy in the learned representation. Huang et al. [45] reconstructed a linear regressing model and learned a shared indicator matrix to ensure the diversity, and consensus of semi-supervised classification tasks. Lin et al. [49] proposed a dynamic graph label propagation model which jointly learned a relationship between the multi-graphs and labels based on GCN networks for classification tasks.

These approaches have been proposed to enhance the performance of multi-view learning. Despite the existence of various graph-based semi-supervised classification methods, they ignored the topology relations between views, resulting in undesired performance. Based on this, we try to extract the topology graphs from different views and explores the correlations between them, thereby integrating the relevant graphs to ensure the homogeneity of graphs.

## 3. The proposed method

### 3.1. Overview and notation

For the purpose of exploring the interactive information between different views and learning graph fusion, we propose a multi-channel contrastive graph convolutional network that comprises three main modules: the generation process of deep interactive feature information, the adjacency matrix fusion network with adaptive division strategy and the learnable GCN with contrastive loss. The framework first aims to capture deep interactive information from different feature views to ensure the consistency of multiple views. Then the nearest neighbor graphs are integrated through a fusion network to preserve the structure of the topology space. Finally, graph convolutional networks with a contrastive loss are used to maximize the consistency between graphs.

This paper employs the notation  $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d_v}$  to represent the multi-view data matrix for the  $v$ -th view ( $v \in [V]$ ), where  $n$  represents the number of data samples, and  $d_v$  denotes the feature dimension for the  $v$ -th view.  $\mathbf{Y} \in \mathbb{R}^{|\Omega| \times c}$  denotes the incomplete label matrix generated from the labeled samples  $\mathbf{\Omega}$ , where  $|\Omega| \ll n$ , and  $c$  is the number of classes in the classification task. To ensure clear understanding of mathematical notations used in this paper, Table 1 provides explanations of the primary symbols.

### 3.2. Deep interactive feature integration network

To ensure the high coherence of multi-view data with varying dimensions, the proposed AMC-GCN first explores the underlying information from the multiple views to enhance complementarity before extracting the deep interactive information for the consensus of views. Therefore, we employ a set of autoencoders  $\{f_v\}_{v=1}^V$  to capture specific feature information for each view. These networks aim to reduce the dimensionality of the original features  $d_v$  to a common dimension  $d$ . Mathematically, the  $l$ -th layer output of the  $v$ -th autoencoder is denoted as  $\mathbf{H}^{(l,v)}$ ,

$$\mathbf{H}^{(l,v)} = \sigma \left( \mathbf{H}^{(l-1,v)} \mathbf{W}_l^{(v)} + \mathbf{b}_l^{(v)} \right), \quad (3)$$

where  $\mathbf{H}^{(l,v)}$  is the  $l$ -th layer of  $f_v(\cdot)$  with  $\mathbf{H}^{(0,v)} = \mathbf{X}^{(v)}$ ,  $\sigma(\cdot)$  denotes the activation function, and  $\mathbf{W}_l^{(v)} \in \mathbb{R}^{m_{l-1} \times m_l}$  and  $\mathbf{b}_l^{(v)} \in \mathbb{R}^{m_l}$  are layer-specific weight and bias, respectively. The loss function for the  $v$ -th view's autoencoder is expressed as follows,

$$\mathcal{L}_f^v = \|\mathbf{H}^{(L,v)} - \mathbf{X}^{(v)}\|_2^2. \quad (4)$$

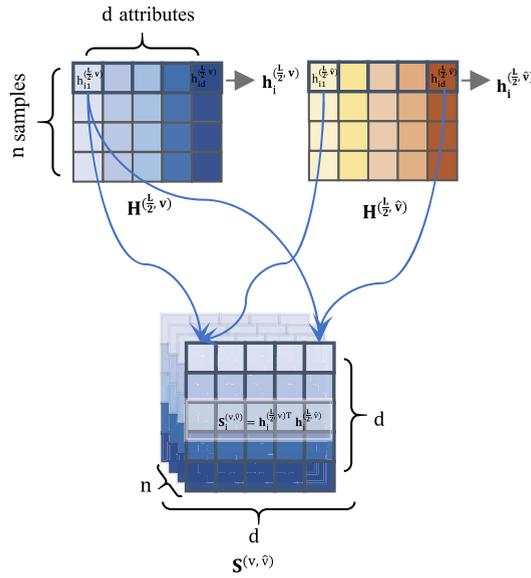


Fig. 3. Construction of the interactive matrix. The interactive matrix  $\mathbf{S}^{(v, \hat{v})}$  is generated by the attribute interactive matrices in  $\mathbf{H}^{(\frac{l}{2}, v)}$  and  $\mathbf{H}^{(\frac{l}{2}, \hat{v})}$ .

To explore the consistency between different views, we extract the deep interactive data from view-specific features. The deep interactive information is produced through the inter-relationships of attributes between two views, where each view interacts with the other  $(V - 1)$  views to generate such information. As for the  $v$ -th view, we first calculate the interactive matrix  $\mathbf{S}^{(v, \hat{v})}$  between  $\mathbf{H}^{(\frac{l}{2}, v)}$  and  $\mathbf{H}^{(\frac{l}{2}, \hat{v})}$ , where  $\mathbf{H}^{(\frac{l}{2}, v)}$  is the output of the  $\frac{l}{2}$ -th layer in  $f_v$ , and  $\hat{v} \in [V] \setminus v$ . Specifically,

$$\mathbf{s}_i^{(v, \hat{v})} = \mathbf{h}_i^{(\frac{l}{2}, v)T} \cdot \mathbf{h}_i^{(\frac{l}{2}, \hat{v})}, i \in [n], \quad (5)$$

where  $\mathbf{s}_i^{(v, \hat{v})} \in \mathbb{R}^{d \times d}$  represents the correlation between the feature attributes of the  $i$ -th sample in  $\mathbf{S}^{(v, \hat{v})}$ , and  $\mathbf{h}_i^{(\frac{l}{2}, v)} \in \mathbb{R}^{1 \times d}$  denotes the  $i$ -th row of  $\mathbf{H}^{(\frac{l}{2}, v)}$ . The attribute association of  $n$  nodes consists of the interactive matrix  $\mathbf{S}^{(v, \hat{v})} \in \mathbb{R}^{n \times d \times d}$  shown in Fig. 3.

Then a deep neural network  $\Phi$  is employed in  $\mathbf{S}^{(v, \hat{v})}$  to merge the attribute correlations between  $\mathbf{H}^{(\frac{l}{2}, v)}$  and  $\mathbf{H}^{(\frac{l}{2}, \hat{v})}$ , where  $\hat{v}$  denotes other view except  $v$ . Thus the interactive information between the  $v$ -th view and the  $\hat{v}$ -th view meets  $\mathbf{G}^{(v, \hat{v})}$ ,

$$\mathbf{G}^{(v, \hat{v})} = \Phi(\text{vec}(\mathbf{S}^{(v, \hat{v})})) = \sigma(\text{vec}(\mathbf{S}^{(v, \hat{v})})\mathbf{W}_\Phi^{(v)} + \mathbf{b}_\Phi^{(v)}), \quad (6)$$

where  $\text{vec}(\cdot)$  presents the vectorization of the matrix, with transforming  $\mathbf{S}^{(v, \hat{v})} \in \mathbb{R}^{n \times d \times d}$  to  $\text{vec}(\mathbf{S}^{(v, \hat{v})}) \in \mathbb{R}^{n \times d^2}$ . Here,  $\mathbf{W}_\Phi^{(v)} \in \mathbb{R}^{d^2 \times d}$  and  $\mathbf{b}_\Phi^{(v)} \in \mathbb{R}^{n \times d}$  denote the weight and bias, respectively. Then we combine the  $i$ -th vector of  $(V - 1)$  interactive information to obtain the final interactive information  $\hat{\mathbf{S}}^{(v)}$  of the  $v$ -th view,

$$\hat{\mathbf{S}}^{(v)} = [\mathbf{G}^{(v, 1)} \parallel \dots \parallel \mathbf{G}^{(v, v-1)} \parallel \mathbf{G}^{(v, v+1)} \parallel \dots \parallel \mathbf{G}^{(v, V)}], \quad (7)$$

where  $\hat{\mathbf{S}}^{(v)} \in \mathbb{R}^{n \times ((V-1) \times d)}$ , and  $\parallel$  is the horizontal splicing of the matrix. Finally, by integrating both the specific information and the final interactive information in the  $v$ -th view, we obtain the consistent interaction matrix set  $\{\mathcal{X}^{(v)}\}$  with  $\mathcal{X}^{(v)} \in \mathbb{R}^{n \times (V \times d)}$ ,

$$\mathcal{X}^{(v)} = \left[ \mathbf{H}^{(\frac{l}{2}, v)}, \hat{\mathbf{S}}^{(v)} \right], v \in [V]. \quad (8)$$

### 3.3. Adjacency matrix fusion network with division strategy

For the purpose of exploring the potential topology information of the feature space, the adjacency matrices  $\{\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(V)}\}$  are initialized by the  $k$ -nearest neighbors from the  $V$  original feature matrices. As for the complementarity of perspectives, each view has extra information that the others may not have. Based on this, we consider that the correlation between the feature information of each perspective is different, leading to the dissimilarity of the topology structures. Aggregating less correlated graphs may lead to topology homogeneity disruption. In addition, a simple linear combination of adjacency matrices may not be effective for multi-view graph learning, because it may generate spurious links by fusing poorly correlated topology information, leading to suboptimal performance.

To obtain more reliable topology structures, a natural idea is to aggregate the topology graphs as relevant as possible. Therefore, inspired by the division of the  $k$ -means method, we design an adaptive division strategy to divide multiple graph channels. To account for variations across different datasets, a relevance threshold  $\theta$  is established for evaluation purposes. It is worth noting that the adaptive division strategy does not fix the number of clusters and uses cluster centroids to calculate the cluster-to-cluster correlation. It first defines the view-specific feature information as  $V$  feature sets  $set_v = \{\mathbf{H}^{(\frac{L}{2}, v)}\}$ , with their centroids  $cent_v = \hat{\mathbf{H}}^{(v)}$  initialized as  $\mathbf{H}^{(\frac{L}{2}, v)}$ . Then the correlation matrix in  $t$ -th iteration  $\mathbf{C}^{(t)} \in \mathbb{R}^{m^{(t)} \times m^{(t)}}$  is constructed by these centroids, where  $t$  is the iteration step,  $m^{(t)}$  denotes centroid numbers, and  $c_{ij}^{(t)}$  represents the relevance of the  $i$ -th and the  $j$ -th centroids. Specifically, we initialize the  $c_{ij}^{(t)}$  in  $\mathbf{C}^{(t)}$  as

$$c_{ij}^{(1)} = \text{vec}(\hat{\mathbf{H}}^{(i)}) \cdot \text{vec}(\hat{\mathbf{H}}^{(j)})^T, \quad (9)$$

where  $\text{vec}(\hat{\mathbf{H}}^{(i)}) \in \mathbb{R}^{1 \times (n \times d)}$ . Then we select the highest correlation value  $c_{ij}^{(t)}$ . If  $c_{ij}^{(t)} \geq \theta$ , we merge the  $i$ -th and the  $j$ -th sets, and then the corresponding centroid becomes the average of the two centroids. The numbers of feature sets and centroids are reduced to one. We repeat the above steps until the maximum value  $c_{ij}^{(t)} < \theta$  or the maximum iteration  $\Gamma_1$  is reached. We obtain  $M$  graph channels according to the  $M$  merged feature sets with each channel containing the adjacency matrices. The detailed steps are presented in Algorithm 1. To fully integrate the topology graphs, a fusion network is proposed to combine the adjacency matrices for each graph

---

**Algorithm 1** The adaptive division strategy.

---

**Input:** View-specific features  $\{\mathbf{H}^{(\frac{L}{2}, 1)}, \dots, \mathbf{H}^{(\frac{L}{2}, V)}\}$ , threshold  $\theta$ , maximum interaction  $\Gamma_1$ .

**Output:**  $M$  merged feature sets  $\{ch_g\}_{g=1}^M$ .

- 1: Initialize  $set_v = \{\mathbf{H}^{(\frac{L}{2}, v)}\}$  and  $cent_v = \hat{\mathbf{H}}^{(v)}$  as  $\mathbf{H}^{(\frac{L}{2}, v)}$ ;
  - 2: **for**  $t = 1 \rightarrow \Gamma_1$  **do**
  - 3:   Construct correlation matrix  $\mathbf{C}^{(t)}$  using Eq. (9);
  - 4:   Select maximum value  $c_{ij}^{(t)}$  from  $\mathbf{C}^{(t)}$ ;
  - 5:   **if**  $c_{ij}^{(t)} \geq \theta$  **then**
  - 6:     Merge  $set_i \leftarrow set_i \cup set_j$ ;
  - 7:     Update  $cent_i \leftarrow (cent_i + cent_j)/2$ ;
  - 8:     Delete  $set_j, cent_j$ , update  $set_i, cent_i$ ;
  - 9:   **else** [ $c_{ij}^{(t)} < \theta$ ]
  - 10:    Break;
  - 11:   **end if**
  - 12: **end for**
  - 13: **Return**  $M$  merged feature sets  $\{ch_g\}_{g=1}^M$ .
- 

channel. Considering that each node has a different influence on their combinations, and there are still various correlations between different views after division, we employ a network to learn their importance and fuse them. Therefore, we focus on an adjacency matrix  $\mathbf{A}^{(m_j)}$  in a graph channel  $ch_g = \{\mathbf{A}^{(m_1)}, \mathbf{A}^{(m_2)}, \dots, \mathbf{A}^{(m_I)}\}$  with  $j \in [I]$  and  $g \in [M]$ . Based on this, a neural network is adopted to obtain the weight value of the  $i$ -th node in  $\mathbf{A}^{(m_j)}$ ,

$$w_i^{(m_j)} = \tanh(\mathbf{a}_i^{(m_j)} \mathbf{W}_{f_1} + \mathbf{b}_{f_1}) \mathbf{W}_{f_2} + \mathbf{b}_{f_2}, \quad (10)$$

where  $\mathbf{a}_i^{(m_j)} \in \mathbb{R}^{1 \times n}$  is the  $i$ -th row of  $\mathbf{A}^{(m_j)}$ ,  $\mathbf{W}_{f_1} \in \mathbb{R}^{n \times n'}$ ,  $\mathbf{W}_{f_2} \in \mathbb{R}^{n' \times 1}$ ,  $\mathbf{b}_{f_1} \in \mathbb{R}^{1 \times n'}$  and  $\mathbf{b}_{f_2} \in \mathbb{R}^{1 \times 1}$  denote weights and biases of the first and second layers, respectively. Similarly, we acquire the weight values for the  $i$ -th node of the other adjacency matrices in the same graph channel. Then a softmax function is employed to normalize the weight values to obtain the final weight of the  $i$ -th node,

$$\alpha_i^{(m_j)} = \frac{\exp(w_i^{(m_j)})}{\sum_{t \in [I]} \exp(w_t^{(m_j)})}, \quad (11)$$

where the value of  $\alpha_i^{(m_j)}$  is positively correlated with the importance of the  $i$ -th node of  $\mathbf{A}^{(m_j)}$ . For all  $n$  nodes in  $\mathbf{A}^{(m_j)}$ , we obtain the learnable weights  $\boldsymbol{\alpha}^{(m_j)} = [\alpha_1^{(m_j)}, \dots, \alpha_n^{(m_j)}]$  with  $\boldsymbol{\alpha}^{(m_j)} \in \mathbb{R}^{1 \times n}$ . We combine all the matrices in the graph channel  $ch_g$  to obtain the final matrix  $\hat{\mathbf{A}}^{(g)}$ ,

$$\hat{\mathbf{A}}^{(g)} = \sum_{t \in [I]} \text{diag}(\boldsymbol{\alpha}^{(m_t)}) \cdot \mathbf{A}^{(m_t)}. \quad (12)$$

Thus we obtain  $M$  merged adjacency matrices  $\{\hat{\mathbf{A}}^{(1)}, \dots, \hat{\mathbf{A}}^{(M)}\}$  and apply them to GCNs.

### 3.4. Learnable GCN with semi-supervised contrastive loss

In this subsection, we introduce the learnable GCN designed to automatically obtain potential feature representation with consistency and complementarity of multiple views. The adaptive weighted sum of view-specific feature matrices with consistent information can be defined as

$$\hat{\boldsymbol{\chi}} = \sum_{v=1}^V \pi^{(v)} \boldsymbol{\chi}^{(v)}, \quad (13)$$

where  $\pi^{(v)}$  is a learnable weight coefficient meeting  $\sum_{v=1}^V \pi^{(v)} = 1$ . Therefore, we employ a softmax function for each epoch,

$$\pi^{(v)} = \frac{\exp(\pi^{(v)})}{\sum_{v=1}^V \exp(\pi^{(v)})}. \quad (14)$$

Exploring relationships between related nodes through a shared feature matrix can obtain global relationships. Based on this, we employ a 2-layer GCNs  $\{\Psi_m\}_{m=1}^M$  on  $M$  graph channels with the shared feature representation  $\hat{\boldsymbol{\chi}}$ ,

$$\mathbf{Z}^{(m)} = \text{softmax} \left( \hat{\mathbf{A}}^{(m)} \sigma \left( \hat{\mathbf{A}}^{(m)} \hat{\boldsymbol{\chi}} \mathbf{W}_g^{(1)} \right) \mathbf{W}_g^{(2)} \right), \quad (15)$$

where  $\mathbf{Z}^{(m)}$  is the probability matrix of the  $m$ -th channel for any  $m \in [M]$ . Trainable weights  $\mathbf{W}_g^{(1)}$  and  $\mathbf{W}_g^{(2)}$  are shared in distinct topology graphs to extract the shared information.

We employ a semi-supervised contrastive loss which is a combination of both supervised and unsupervised contrastive losses to enhance the consistency between distinct topology graphs. As for the probability matrices  $\mathbf{Z}^{(\Phi_1)}$  and  $\mathbf{Z}^{(\Phi_2)}$ , the unsupervised contrastive loss is defined as

$$\mathcal{L}_{uc} = \frac{1}{2n} \sum_{i=1}^n \left( \mathcal{L}_{uc}^{\Phi_1}(\hat{\mathbf{x}}_i) + \mathcal{L}_{uc}^{\Phi_2}(\hat{\mathbf{x}}_i) \right), \quad (16)$$

where  $\Phi_1, \Phi_2$  represent different graph channels,  $\hat{\mathbf{x}}_i$  is the  $i$ -th node,  $\mathcal{L}_{uc}^{\Phi_1}(\cdot)$  and  $\mathcal{L}_{uc}^{\Phi_2}(\cdot)$  denote the unsupervised contrastive losses of the  $\Phi_1$ -th and  $\Phi_2$ -th channels, respectively. Specifically,  $\mathcal{L}_{uc}^{\Phi_1}(\hat{\mathbf{x}}_i)$  can be measured by:

$$\mathcal{L}_{uc}^{\Phi_1}(\hat{\mathbf{x}}_i) = -\log \frac{\exp(\mathbf{z}_i^{\Phi_1} \cdot \mathbf{z}_i^{\Phi_2 T})}{\sum_{j=1}^n \exp(\mathbf{z}_i^{\Phi_1} \cdot \mathbf{z}_j^{\Phi_2 T})}, \quad (17)$$

where  $\mathbf{z}_k^{\Phi_i} \in \mathbb{R}^{1 \times c}$  is the  $k$ -th row vector of  $\mathbf{Z}^{\Phi_i}$ . Similarly,  $\mathcal{L}_{uc}^{\Phi_2}(\hat{\mathbf{x}}_i)$  can be calculated as:

$$\mathcal{L}_{uc}^{\Phi_2}(\hat{\mathbf{x}}_i) = -\log \frac{\exp(\mathbf{z}_i^{\Phi_2} \cdot \mathbf{z}_i^{\Phi_1 T})}{\sum_{j=1}^n \exp(\mathbf{z}_j^{\Phi_2} \cdot \mathbf{z}_i^{\Phi_1 T})}. \quad (18)$$

Furthermore, to integrate the limited yet trustworthy labels, the supervised contrastive loss of  $\hat{\mathbf{X}}_i \in \Omega$  is defined as:

$$\mathcal{L}_{sc} = \frac{1}{2|\Omega|} \sum_{i=1}^{|\Omega|} \left( \mathcal{L}_{sc}^{\Phi_1}(\hat{\mathbf{x}}_i) + \mathcal{L}_{sc}^{\Phi_2}(\hat{\mathbf{x}}_i) \right). \quad (19)$$

In general, the supervised contrastive losses for  $\hat{\mathbf{x}}_i$  belonging to  $l$  labeled samples are computed by:

$$\mathcal{L}_{sc}^{\Phi_1}(\hat{\mathbf{x}}_i) = -\log \frac{\sum_{k=1}^{|\Omega|} \mathbb{1}_{[r_i^{\Phi_1}=r_k^{\Phi_2}]} \exp(\mathbf{z}_i^{\Phi_1} \cdot \mathbf{z}_k^{\Phi_2 T})}{\sum_{j=1}^{|\Omega|} \exp(\mathbf{z}_i^{\Phi_1} \cdot \mathbf{z}_j^{\Phi_2 T})}, \quad (20)$$

$$\mathcal{L}_{sc}^{\Phi_2}(\hat{\mathbf{x}}_i) = -\log \frac{\sum_{k=1}^{|\Omega|} \mathbb{1}_{[r_i^{\Phi_2}=r_k^{\Phi_1}]} \exp(\mathbf{z}_i^{\Phi_2} \cdot \mathbf{z}_k^{\Phi_1 T})}{\sum_{j=1}^{|\Omega|} \exp(\mathbf{z}_i^{\Phi_2} \cdot \mathbf{z}_j^{\Phi_1 T})}, \quad (21)$$

where  $r_i^{\Phi_1}$  represents the predictive class of  $\mathbf{z}_i^{\Phi_1}$  and  $\mathbb{1}_{[l, \cdot]}$  is equal to 1 if  $r_i^{\Phi_1} = r_k^{\Phi_2}$ , and 0 otherwise. By integrating both the unsupervised and the supervised contrastive losses, we obtain the semi-supervised contrastive loss,

$$\mathcal{L}_{ssc} = \mathcal{L}_{uc} + \lambda \mathcal{L}_{sc}, \quad (22)$$

where  $\lambda$  is a hyperparameter. Therefore, the semi-supervised contrastive loss for the  $M$  graph channels is,

$$\hat{\mathcal{L}}_{ssc} = \sum_{i=1}^{M-1} \sum_{j=i+1}^M \mathcal{L}_{ssc}. \quad (23)$$

### 3.5. Model training

For acquiring an overall output  $\hat{\mathbf{Z}}$ , we aggregate the embeddings  $\{\mathbf{Z}^{(m)}\}_{m=1}^M$  for capturing information of feature topology space as,

$$\hat{\mathbf{Z}} = \sum_{m=1}^M \beta^{(m)} \mathbf{Z}^{(m)}, \quad (24)$$

where  $\beta^{(m)}$  is an adjustable hyperparameter set with  $\sum_{m=1}^M \beta^{(m)} = 1$ . Besides, to measure the dissimilarity between the output  $\hat{\mathbf{Z}}$  and the ground truth  $\mathbf{Y}$ , we employ the cross-entropy loss as follows,

$$\mathcal{L}_{en} = - \sum_{i \in \Omega} \sum_{j=1}^c \mathbf{Y}_{ij} \ln \hat{\mathbf{Z}}_{ij}. \quad (25)$$

The overall loss function of the proposed method is given by

$$\mathcal{L} = \mathcal{L}_{en} + \gamma \hat{\mathcal{L}}_{ssc}, \quad (26)$$

where hyperparameter  $\gamma$  is used to control the importance of  $\mathcal{L}_{ssc}$  and  $\mathcal{L}_{en}$ . To sum up, the procedure for AMC-GCN is elaborated in Algorithm 2.

---

#### Algorithm 2 Adaptive multi-channel contrastive graph convolutional networks (AMC-GCN).

---

**Input:** Multi-view data  $\{\mathbf{X}^{(v)}\}_{v=1}^V$ , label matrix  $\mathbf{Y} \in \mathbb{R}^{I \times c}$ , maximum iterations  $\Gamma_1, \Gamma_2$  and threshold  $\theta$ .

**Output:** Output  $\hat{\mathbf{Z}}$ .

- 1: Initialize learnable weights  $\left\{ \pi^{(v)} = \frac{1}{V} \right\}_{v=1}^V$ ;
- 2: # Obtain view-specific information;
- 3: Initialize adjacency matrices  $\{\mathbf{A}^{(v)}\}_{v=1}^V$  via  $k$ NN;
- 4: **while** not convergent **do**
- 5:   **for**  $v = 1 \rightarrow V$  **do**
- 6:     Compute  $\mathbf{H}^{(\frac{L}{2}, v)}$  and  $\mathbf{H}^{(L, v)}$  using Eq. (3);
- 7:     Calculate the loss function by (4);
- 8:   **end for**
- 9: **end while**
- 10: # Integrate graph channels;
- 11: Initialize  $set_v = \mathbf{H}^{(\frac{L}{2}, v)}$  and  $cent_v = \mathbf{H}^{(L, v)}$  for  $v \in [V]$ ;
- 12: **for**  $t = 1 \rightarrow \Gamma_1$  **do**
- 13:   Update  $\mathbf{C}^{(t)}$  and select maximum value  $c_{ij}^{(t)}$  in Eq. (9);
- 14:   **if**  $c_{ij}^{(t)} \geq \theta$  **then**
- 15:     Merge  $set_i \leftarrow set_i \cup set_j$ ;
- 16:     Construct  $cent_{\phi} \leftarrow (cent_i + cent_j)/2$ ;
- 17:     Delete  $set_i, set_j, cent_i, cent_j$ , update  $set_i, cent_i$ ;
- 18:   **else** [ $c_{ij}^{(t)} < \theta$ ]
- 19:     break;
- 20:   **end if**
- 21: **end for**
- 22: Obtain  $M$  merged feature sets;
- 23: # Compute predictive representations;
- 24: **while** do not convergence or reach  $\Gamma_2$  **do**
- 25:   **for**  $v = 1 \rightarrow V$  **do**
- 26:     Acquire  $\{\mathcal{X}^{(v)}\}$  by Eq. (8);
- 27:   **end for**
- 28:   Compute consistent feature matrix  $\hat{\mathcal{X}}$  by Eq. (13);
- 29:   **for**  $m = 1 \rightarrow M$  **do**
- 30:      $\mathbf{Z}^{(m)} = \Psi_m(\hat{\mathbf{A}}_{(m)}, \hat{\mathcal{X}})$  by Eq. (15);
- 31:   **end for**
- 32:   Calculate the loss function by Eq. (26);
- 33: **end while**
- 34: **Return** Output  $\hat{\mathbf{Z}}$ .

---

We also analyze the computational complexity of AMC-GCN. Generally, the proposed AMC-GCN can be divided into three components: DIF-Network, AF-Network and SC-Network. Specifically, the computational complexity of DIF-Network costs  $\sum_{v=1}^V \mathcal{O}(ndd_v + nd^2)$ , AF-Network processing consumes  $\mathcal{O}(ndV^2)$ , and SC-Network takes  $\mathcal{O}(M(n^2d + ndc))$ . Owing to  $d_v \approx d$ ,  $c \ll d$  and  $M \leq V$ , the total computational complexity of AMC-GCN requires  $\mathcal{O}(Vnd(d + V))$ .

**Table 2**  
A brief description of all test multi-view datasets.

Datasets	# Samples	# Views	# Features	# Classes	Data Types
BBCsports	544	2	3, 183/3, 203	5	Textual documents
BBCnews	685	4	4, 659/4, 633/4, 665/4, 684	4	Textual documents
ALOI	1,079	4	64/64/77/13	10	Object images
Caltech7	1,474	6	48/40/254/1,984/512/928	7	Object images
HW	2,000	6	153/596/301/481/157/27	10	Digit images
MNIST7K	7,000	3	30/9/30	7	Digit images
Hdigit	10,000	2	784/256	10	Digit images
NoisyMNIST	15,000	2	784/784	10	Digit images

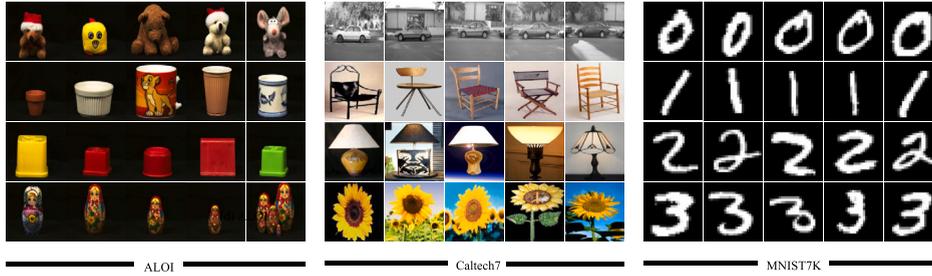


Fig. 4. Samples from three image datasets.

## 4. Experiments

### 4.1. Experimental setup

1). *Datasets*. In the subsection, we evaluate the performance of AMC-GCN on eight multi-view datasets for the semi-supervised classification task. These datasets cover three different applications, including generic news article categorization, generic object detection, and digit classification, which are shown in Table 2. Fig. 4 illustrates the several samples from the selected image datasets.

- **BBCsports** comprises 544 documents sourced from the BBC sport website. Each document has 2 types of features, with dimensions of 3, 183 and 3, 203.
- **BBCnews** is a dataset which contains 685 documents. 4 feature views are presented with 4, 659-D politics, 4, 633-D entertainments, 4, 665-D business sports, and 4, 684-D technology fields.
- **ALOI** is a colorful image dataset that contains 1,079 objects for capturing the variations in object recordings. The dataset has 4 multi-view features including 64-D RGB color histograms, 64-D HSV color histograms, 77-D color similarities, and 13-D Haralick features.
- **Caltech7** is a widely used image dataset consisting of 7 classes. There are 6 representations, namely 48-D Garbor features, 40-D Wavelet Moments, 254-D CENTRIST features, 1,984-D HOG features, 512-D GIST features, and 928-D LBP features.
- **HW** is a handwriting image dataset with 2,000 samples and 10 categories. Feature views contain 153-D Profile-correlation features, 596-D Fourier-coefficients, 301-D Karhunen-Loeve-coefficients, 481-D intensity-avarages, 157-D Zernike Moments, and 27-D Morphological features.
- **MNIST7K** is an extended digit image dataset of MNIST, which has 7,000 data points and 7 classes, with multi-view features as 30-D, 9-D, and 30-D, respectively.
- **Hdigit** has 10,000 image samples with 10 categories, which were crawled from the UCI repository. Its feature dimensions are presented with 784-D and 256-D, respectively.
- **NoisyMNIST** consisting of 15,000 data points employs the original images as a feature view, and randomly selects within-class images with white Gaussian noise as another view, with their dimensions being 784-D and 784-D, respectively.

2). *Baselines*. To validate the efficacy of AMC-GCN, we conduct a comparative study with other state-of-the-art approaches for semi-supervised classification in the multi-view setting. A succinct overview of each method is provided below.

- **MVAR** [46] employs  $\ell_{2,1}$  matrix norm to regularize the loss function and balances the contributions of each view using adaptive weights to enhance the classification performance.
- **MLAN** [13] is a unified model which performs unsupervised multi-view clustering and semi-supervised classification tasks concurrently without requiring explicit weights and penalty parameters.
- **MvNNcor** [47] is a multi-view modeling framework which leverages the specific features and interactive information of views through a well-designed multi-view loss function to accurately predict the categories of the data.

**Table 3**

Classification performance of all compared semi-supervised classification methods with 10% labeled samples as supervision, where the best performance is highlighted in bold and the second best result is underlined.

Metrics	Datasets/Methods	MVAR	MLAN	MvNNcor	Co-GCN	ERL-MVSC	LGCN	IMv-GCN	AMC-GCN
ACC	BBCsports	73.03	64.01	88.33	86.31	88.08	<u>97.00</u>	95.11	<b>97.56</b>
	BBCnews	75.71	72.74	83.33	86.41	85.21	88.00	<u>90.35</u>	<b>90.44</b>
	ALOI	66.36	84.42	94.38	96.53	85.61	<u>96.90</u>	72.55	<b>97.21</b>
	Caltech7	84.99	59.23	<u>94.19</u>	92.57	93.54	85.82	72.55	<b>94.19</b>
	HW	82.25	<u>96.05</u>	96.04	92.44	89.93	92.83	95.61	<b>97.78</b>
	MNIST7K	90.06	89.31	94.68	93.43	<u>95.35</u>	93.86	93.75	<b>97.10</b>
	Hdigit	81.08	98.28	<u>98.29</u>	92.36	95.47	72.70	98.12	<b>98.98</b>
	NoisyMNIST	71.01	10.52	24.20	83.52	82.52	51.90	<u>87.10</u>	<b>88.05</b>
F1	BBCsports	69.15	63.33	87.10	86.24	62.89	<u>96.96</u>	94.93	<b>97.60</b>
	BBCnews	71.08	72.81	79.10	85.92	60.14	87.89	<u>89.35</u>	<b>89.97</b>
	ALOI	62.01	84.65	93.60	96.51	71.50	<u>96.91</u>	69.28	<b>97.23</b>
	Caltech7	48.51	58.77	<u>76.60</u>	0.92	59.44	37.28	69.28	<b>76.64</b>
	HW	42.22	<u>96.04</u>	95.70	92.44	74.94	92.92	95.62	<b>97.78</b>
	MNIST7K	89.80	88.43	<u>94.30</u>	93.36	74.11	93.75	93.63	<b>97.04</b>
	Hdigit	76.59	98.28	97.70	92.36	79.56	68.91	<u>98.12</u>	<b>98.98</b>
	NoisyMNIST	70.89	10.68	21.50	83.41	82.01	45.24	<u>86.53</u>	<b>87.67</b>

- **Co-GCN** [22] is a new graph convolutional network method that integrates both co-training and spectral information into the framework. Co-GCN adaptively exploited graph information from different views by combining Laplacian matrices.
- **ERL-MVSC** [45] utilizes a linear regression model and a shared indicator matrix regularized by  $\ell_{2,1}$ -norm to achieve diverse, sparse, and consistent representation learning for classification tasks.
- **LGCN** [48] is a learnable neural network framework which solves the multi-view learning problem by incorporating several sparse autoencoders and a fully-connected network to integrate features and views.
- **IMv-GCN** [23] is a new multi-view classification neural network framework that utilizes a graph filter and an orthogonal normalization to enhance the interpretability of GCN-based classification models.

Among the above methods, Co-GCN, LGCN and IMv-GCN are GCN-oriented methods, and the rest compared approaches are graph-oriented algorithms. It is noted that there are still few GCN-based methods involved in the existing researches, which indicates that limited work has focused on GCN contributing to the semi-supervised classification task on multi-view data.

3). *Parameter Setting.* In the experiment, we adopt all parameter settings recommended by the corresponding papers for all baselines. To comprehensively assess the proposed model, we set the labeled node rate to 10% per class in the training set, and the remaining nodes are designated as the test set. As for AMC-GCN model, the view-specific feature information is constructed by the autoencoders  $\{f_v\}_{v=1}^V$  with the number of latent representations being (400, 200). To validate the output value within certain limits, we set  $\sigma(\cdot)$  as the ReLU function. To maximize the consistency of feature attributes of  $\mathbf{G}^{(v,\theta)}$ , we employ one fully-connected layer with (400, 200). The initial adjacency matrices  $\{A^v\}_{v=1}^V$  are constructed by  $k$ NN method with  $k = 10$ . We adopt a 2-layer GCN with the weight decay as  $5 \times 10^{-4}$ , and the first layer's neuron activation is computed by ReLU activation function. We utilize the Adam optimizer to optimize the learnable parameters with a fixed learning rate of  $1 \times 10^{-2}$ , and we set the hyperparameter  $\lambda$  to 1. Because of the variation of different datasets, for BBCnews, BBCsports, HW, and ALOI, we choose correlation threshold  $\theta$  as 10, 10, 30, 40, and the rest are 0.

We utilize two well-known metrics including Accuracy (ACC) and macro F1-score (F1) for performance evaluation. To investigate the classification performance of different algorithms, we compare all the methods with 10% labeled samples for training. We conduct experiments with fixed epochs of 2,000 and randomly selected labeled data for each dataset, and report the classification accuracy (ACC) and F1-score (F1). In this paper, the proposed framework is implemented using PyTorch and executed on a computer equipped with an AMD R9-5900X CPU, Nvidia RTX 3060 GPU, and 48 GB RAM.

#### 4.2. Experimental results

**Classification results.** To visually compare the classification capability of AMC-GCN with other state-of-the-art methods, we exhibit the experimental results on eight graph datasets in Table 3. The experimental results demonstrate that the proposed AMC-GCN achieves remarkable classification performance on all test datasets. Notably, AMC-GCN outperforms these graph-oriented methods, especially on BBCnews and BBCsports. This discovery suggests that AMC-GCN has a more robust ability to propagate node attributes and leverage structural information from diverse perspectives. Compared with GCN-based methods, AMC-GCN shows significant performance improvements on the Caltech7, HW, MNIST7K, and Hdigit datasets. This indicates that the proposed AMC-GCN can learn more discriminative node relationships and graph information to extract feature representations on both small and large datasets, thereby verifying the efficacy of the proposed graph and feature fusion approaches.

Fig. 5 depicts the performance of all compared methods at different label ratios, varying from 1% to 10%. The results indicate that AMC-GCN achieves satisfactory performance with relatively small amounts of supervision, outperforming other methods that

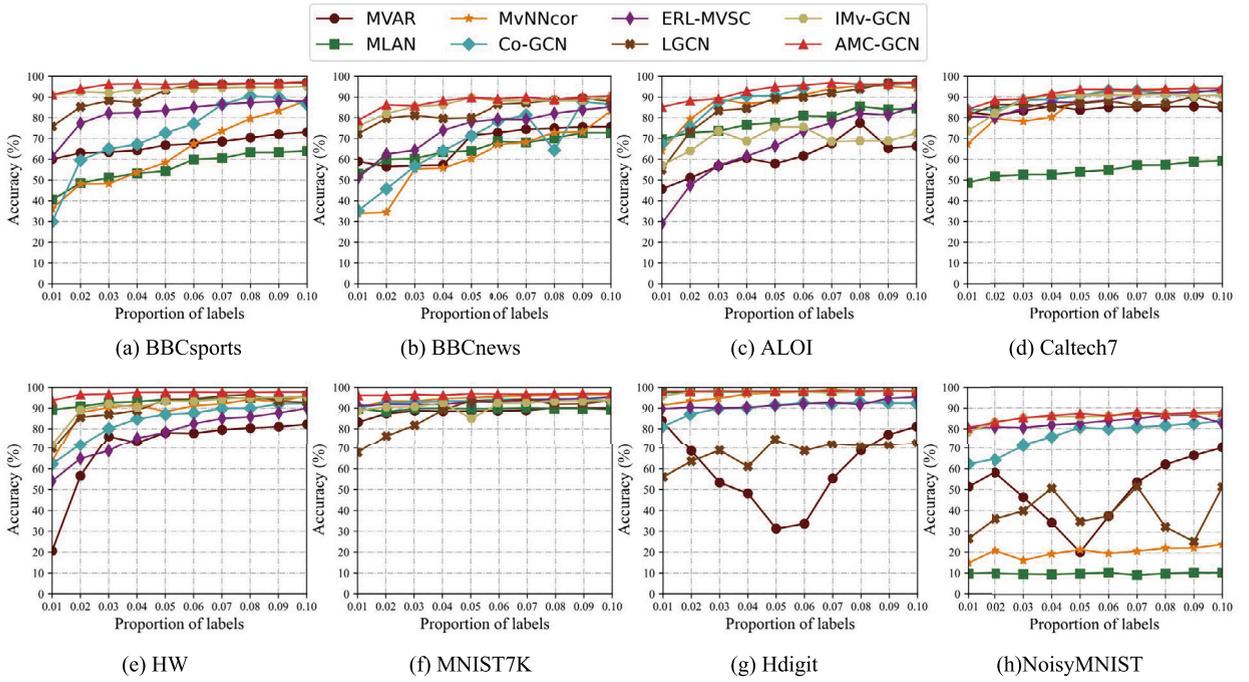


Fig. 5. Performance comparison (ACC%) of various methods as the ratio of labeled data ranges in  $\{0.01, 0.02, \dots, 0.1\}$ .

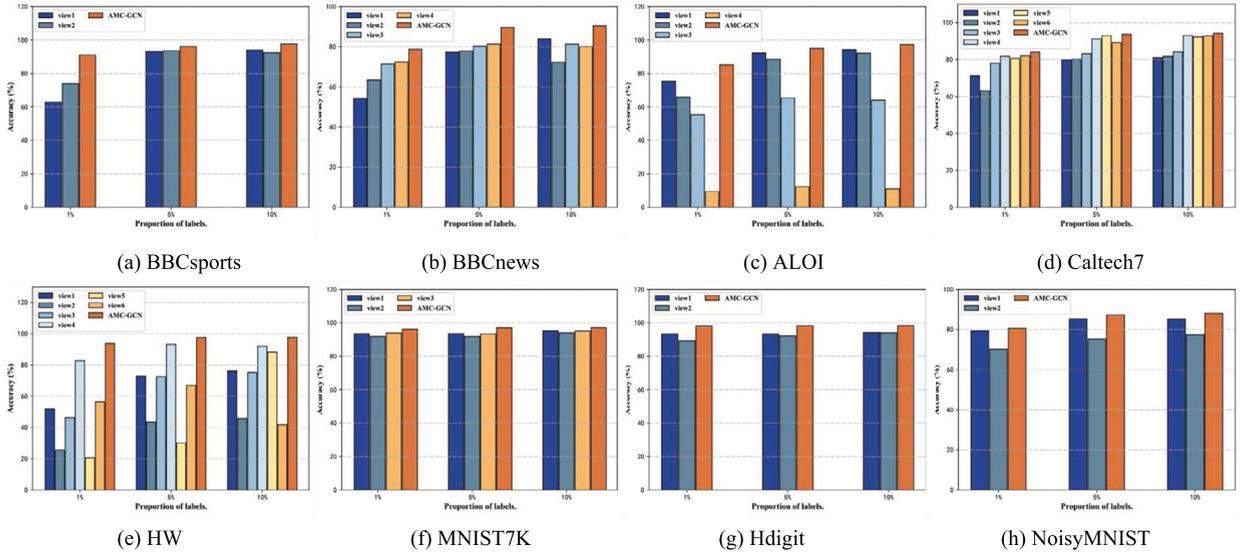


Fig. 6. Comparison (ACC%) of AMC-GCN with fused multi-view and single-view features on all datasets.

need more supervision to attain comparable classification capacity. This highlights that AMC-GCN has a superior capacity for label utilization in the case of scarce label information and can effectively utilize the supervision information and propagate them to unlabeled data. The obtained results reinforce the competitive performance of AMC-GCN compared to other state-of-the-art methods, which aligns with the purpose of semi-supervised classification. This emphasizes the importance of a well-designed contrastive framework that integrates both feature and topology information to enhance the capacity for extracting robust and generalized representations.

**Single-view baseline comparison.** Fig. 6 presents the classification ability of AMC-GCN with fused multi-view and single-view information on all test datasets, respectively. From observation, we can find that AMC-GCN with fused multi-view information consistently outperforms the single-view-based methods in terms of classification accuracy. The findings further suggest that the proposed method can effectively integrate multiple views to leverage the consensus and complementary information for improving learning performance.

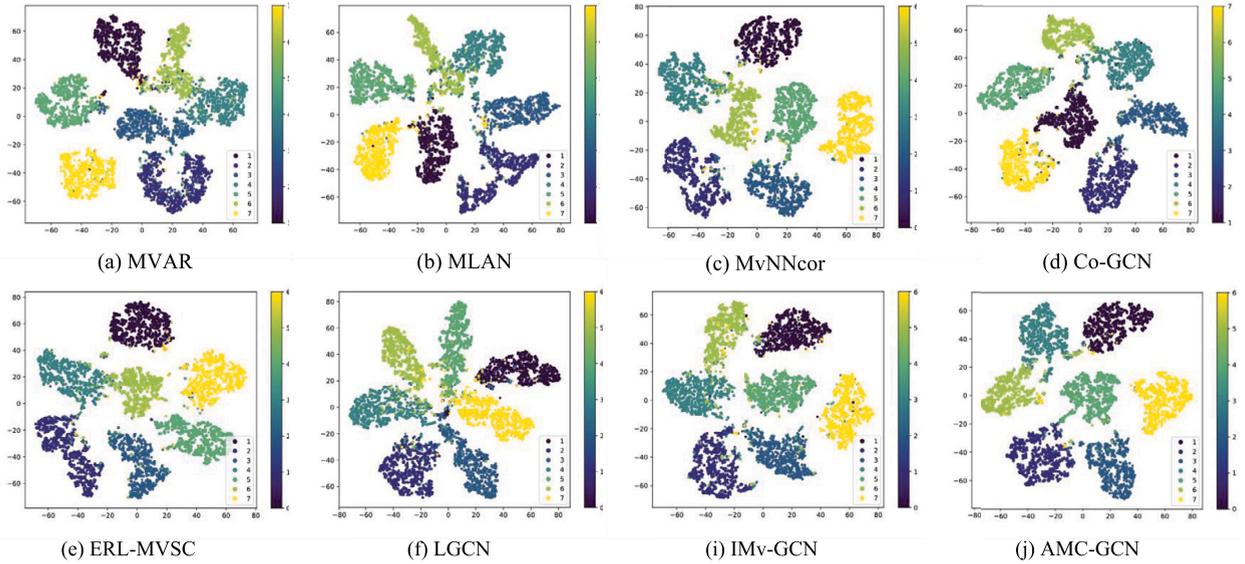


Fig. 7. T-SNE visualization of data distributions with different compared methods on MNIST7K with 10% ratio labeled samples.

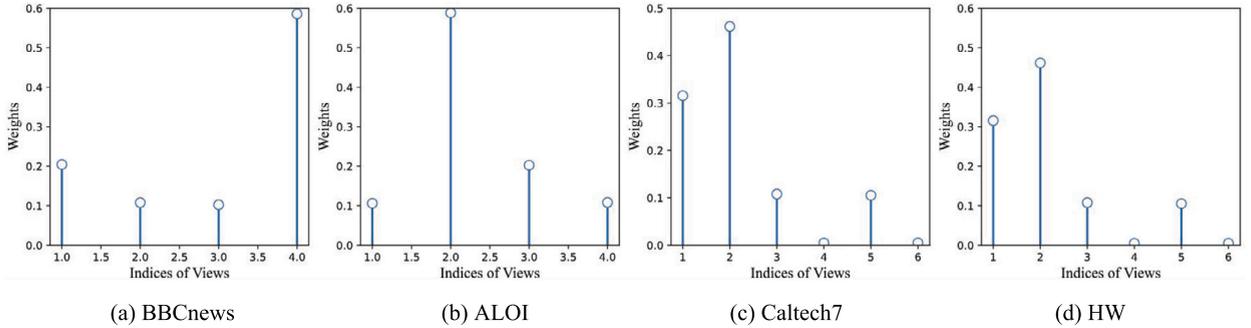


Fig. 8. Visualization of learned weights with distinct views by AMC-GCN on BBCnews, ALOI, Caltech7, HW datasets.

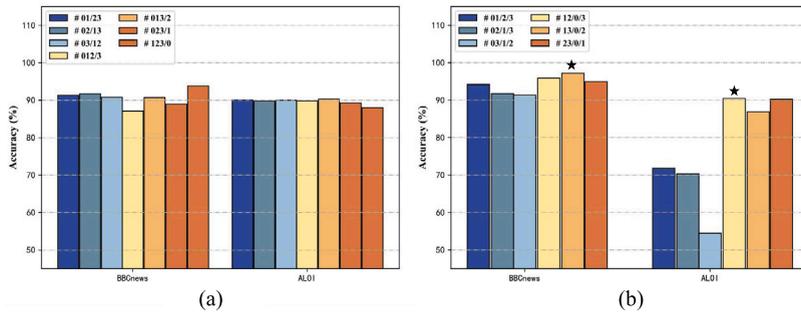


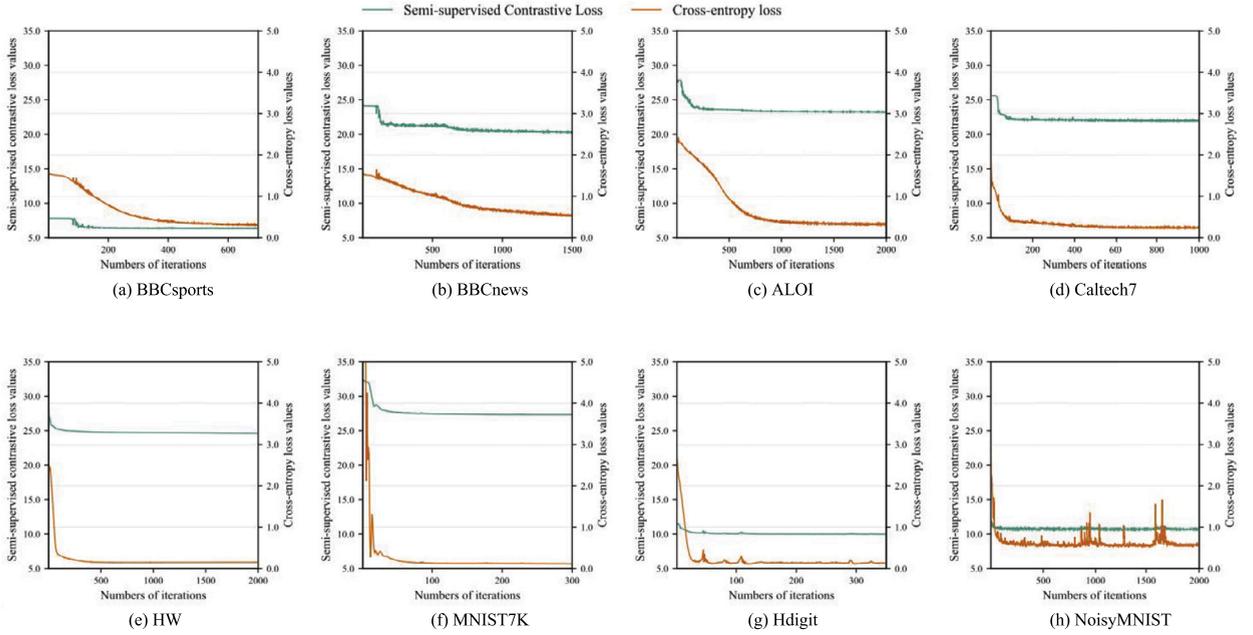
Fig. 9. Comparison of AMC-GCN with different division of graph channels (of which #01/23 denotes that AMC-GCN divides two graph channels containing the information of view 0 and view 1, view 2 and view 3, respectively. Especially,  $\star$  is the graph channels divided by the adaptive division strategy in AMC-GCN).

**Visualization of classification performance.** To intuitively depict the classification performance of different algorithms, Fig. 7 exhibits scatter plots with 10% label ratio on MNIST7K. The figures indicate that our AMC-GCN method achieves more consistent class assignments compared to the ground truth, thus confirming the superiority of AMC-GCN. Besides, Fig. 8 presents the learned weights of different feature views on AMC-GCN on BBCnews, ALOI, Caltech7, HW datasets. The learned weight of each view indicates that each view of datasets provides the proportion of valuable feature information and the degree of contribution for classification tasks. The reason may be that different views are described from different perspectives on the same data, leading to provision of diverse valid feature representations. To validate the effectiveness of the adaptive division strategy, the proposed AMC-GCN also compares the performance with other graph channel division strategy. In addition, Fig. 9 exhibits that the different division of graph

**Table 4**

Ablation study of the proposed AMC-GCN on all test datasets, where ACC% and F1% are recorded.

Datasets / Methods	GCN-fusion	IA-GCN	IMC-GCN	AMC-GCN
BBCsports	82.04 (83.80)	93.08 (92.83)	95.72 (95.58)	<b>97.56 (97.60)</b>
BBCnews	74.88 (73.26)	88.01 (88.04)	88.66 (88.06)	<b>90.44 (89.97)</b>
ALOI	84.00 (83.90)	84.24 (84.36)	95.05 (95.05)	<b>97.21 (97.23)</b>
Caltech7	91.63 (70.89)	91.68 (71.55)	93.74 (76.14)	<b>94.19 (76.64)</b>
HW	95.67 (95.66)	95.67 (95.69)	95.69 (95.69)	<b>97.78 (97.78)</b>
MNIST7K	86.10 (87.41)	95.95 (95.88)	96.30 (96.24)	<b>97.10 (97.04)</b>
Hdigit	86.79 (88.16)	94.09 (94.08)	98.16 (98.16)	<b>98.98 (98.98)</b>
NoisyMNIST	84.41 (84.17)	85.79 (85.87)	87.92 (87.69)	<b>88.05 (87.67)</b>



**Fig. 10.** Convergence curves of semi-supervised contrastive loss values and cross-entropy loss values on all datasets.

channels, from which we observe that the accuracy divided by the adaptive division strategy in AMC-GCN achieves the optimal performance, validates the fusion of topology graphs using the adaptive division strategy.

#### 4.3. Model analysis

**Ablation study.** In order to assess the contribution of the proposed models, we conduct experiments to evaluate the classification ability of AMC-GCN with its variants on all datasets. Specifically, we test the accuracy of GCN-fusion [16] as a baseline, since the original model cannot directly apply to multi-view classification task, we achieve this by calculating the average adjacency matrix during graph convolutions. We also compare the performance of IA-GCN based on GCN-fusion, which learns the deep interactive information between views and then combines the average adjacency matrix to a graph convolutional network. IMC-GCN is constructed on IA-GCN, which employs an adjacency matrix fusion network to integrate relevant topology structures instead of averaging topology structures, then applies multi-channel graph convolutional networks to the classification tasks. Moreover, the AMC-GCN framework further adds the semi-supervised contrastive loss on the basis of IMC-GCN. Table 4 presents the results of ablation study. The results demonstrate that AMC-GCN significantly enhances the performance of the framework, indicating that each constituent of AMC-GCN plays a important role in the semi-supervised node classification task. This may account for the reason that AMC-GCN completes the intact representation of different views, and maximizes the homogeneity of views, which reduces the impact of noise generated by the complementarity of different views. Moreover, the performance evaluation also confirms the efficacy of the semi-supervised loss in enhancing the classification accuracy.

**Convergence validation.** Fig. 10 presents the convergence of AMC-GCN on eight real-world datasets with 10% labeled samples. To enhance the analysis, we visualize the values of both the semi-supervised contrastive loss and the cross-entropy loss within a single figure. Several valuable insights can be drawn from these plots. First, the value of  $\mathcal{L}_{ssc}$  drops rapidly within 200 iterations

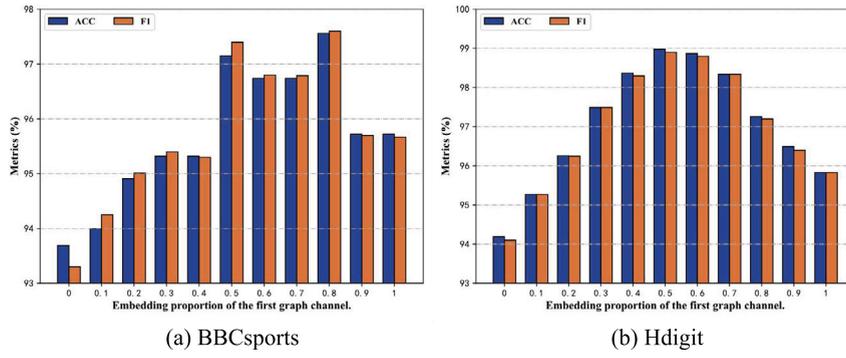


Fig. 11. Parameter sensitivity of AMC-GCN w.r.t.  $\beta^{(m)}$ ,  $m = 1$  on BBCsports and Hdigit.

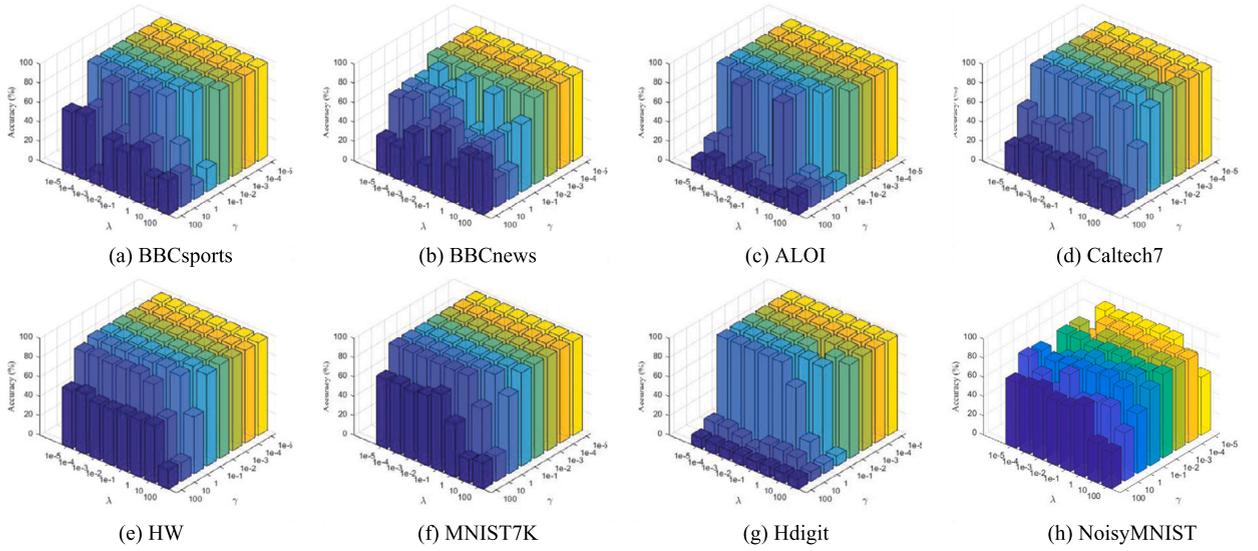


Fig. 12. Parameter sensitivity (ACC %) of AMC-GCN w.r.t.  $\lambda$  and  $\gamma$  on all datasets.

and then stabilizes during subsequent iterations. This can be explained by the fact that the semi-supervised contrastive loss promotes consistency between different topology graphs by sharing feature representations composed of consensus and interactive information. Second, the loss value of  $\mathcal{L}_{en}$  decreases smoothly during the training process on most datasets except MNIST7K, Hdigit and NoisyMNIST, attributed to the fact of the instability of the model in the initial training iterations. Then the proposed AMC-GCN tends to be convergent in the late iterations. Besides, the loss value of  $\mathcal{L}_{en}$  in NoisyMNIST jitters dramatically, but it converges generally. This may be that it is inherently noisy, resulting in a relatively large magnitude of loss during training. Generally, the analysis of  $\mathcal{L}_{SSC}$  and  $\mathcal{L}_{en}$  emphasizes the effectiveness and stable convergence of AMC-GCN.

**Parameter sensitivity.** We conduct parameter sensitivity analysis to examine the performance changes of the proposed method under different parameter settings.  $\beta^{(m)}$  is a hyperparameter ranging in  $[0, 1]$ . The value of  $\beta^{(m)}$  is positively correlated to the importance of for the  $m$ -th graph channel. From Fig. 11, we can observe that satisfactory performance emerges for  $\beta^{(1)}$  values within the range of  $[0.5, 0.8]$ . This further underscores the effectiveness of the proposed adaptive division strategy. In addition, Fig. 12 presents the parameter sensitivity of  $\lambda$  and  $\gamma$  in  $[10^{-5}, 10^2]$  with 10% labeled samples on eight real-world datasets. From the figures, some observations are shown below. First, the optimal values for AMC-GCN are achieved when  $\lambda = 1$  across all datasets. This suggests that both the supervised and unsupervised contrastive losses exert similar influences on AMC-GCN. Second, it is observed that the classification performance of AMC-GCN remains relatively stable when  $\gamma$  falls in the range of  $[10^{-1}, 10^{-4}]$ . However, when  $\gamma$  ranges in  $[1, 100]$ , the classification performance fluctuates dramatically. This finding emphasizes the significance of both the semi-supervised contrastive loss and the cross-entropy loss in AMC-GCN. Therefore, it is better to take a smaller value for parameter  $\gamma$ .

Besides, we investigate the impact of the number of GCN layers on the performance of the proposed model, as illustrated in Fig. 13. It can be seen that the model consistently attains optimal performance when utilizing a two-layer model in comparison to other layer numbers. This may be because that a shallow network with a small number of parameters may not provide sufficient feature propagation. When the layer number increases, it often leads to undesired performance due to over-smoothing.

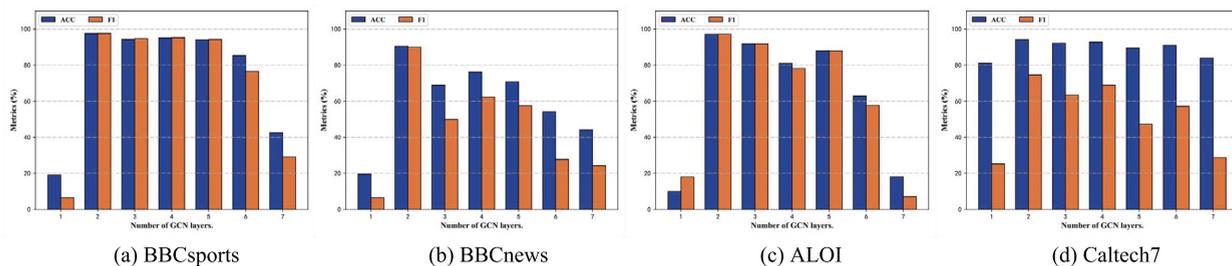


Fig. 13. Influence of AMC-GCN with the varying numbers of layers on four datasets.

## 5. Conclusion

In this paper, we proposed an adaptive multi-channel contrastive graph neural network framework with deep interactive information named AMC-GCN which considered the correlation between node relationships and the relevance of views. For better incorporating consensus and complementary information, we applied a feature integrating network to extracting deep interactive information between views. To fuse the topology graph with minimizing the destruction of topology structures, we explored the relevance of views, imposed an adjacency matrix fusion network and designed multi-graph channels for delivering valuable information. In addition, AMC-GCN used a semi-supervised contrastive loss with parameter sharing on GCNs to enhance the homogeneity of graph channels. The experimental results on graph datasets demonstrated the superiority of AMC-GCN compared with other popular baselines and classification methods.

There are several potential research directions that can be explored in the field of multi-view learning. In our proposed model, we focus on undirected graphs. However, in real-world scenarios, node relationships are more likely to be directed graphs. Therefore, it is important to investigate more GCN-based models on directed graphs. In future work, we plan to extend AMC-GCN and develop a multi-view learning framework that can handle directed graphs.

### CRedit authorship contribution statement

**Luying Zhong:** Conceptualization, Methodology, Software, Writing – original draft. **Jielong Lu:** Software, Validation, Visualization, Writing – review & editing. **Zhaoliang Chen:** Validation, Writing – review & editing. **Na Song:** Investigation, Validation. **Shiping Wang:** Conceptualization, Methodology, Supervision, Validation.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgements

This work is in part supported by the National Natural Science Foundation of China under Grants U21A20472 and 62276065, the National Key Research and Development Plan of China under Grant 2021YFB3600503, and the Open Research Fund from the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen, under Grant B10120210117-OF10.

### References

- [1] S. Bouhenni, S. Yahiaoui, N. Nouali-Taboudjemat, H. Kheddouci, Distributed graph pattern matching via bounded dual simulation, *Inf. Sci.* 610 (2022) 549–570.
- [2] M.A. Castán-Lascorz, P. Jiménez-Herrera, A.T. Lora, G. Asencio-Cortés, A new hybrid method for predicting univariate and multivariate time series based on pattern forecasting, *Inf. Sci.* 586 (2022) 611–627.
- [3] M. Ahmad, S.-W. Lee, Human action recognition using shape and clg-motion flow from multi-view image sequences, *Pattern Recognit.* 41 (2008) 2237–2252.
- [4] H. Gao, F. Nie, X. Li, H. Huang, Multi-view subspace clustering, in: *Proceedings of the 15th IEEE/CVF International Conference on Computer Vision*, 2015, pp. 4238–4246.
- [5] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3d shape recognition, in: *Proceedings of the 15th IEEE/CVF International Conference on Computer Vision*, 2015, pp. 945–953.
- [6] Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, L. Quan, Blendedmvs: a large-scale dataset for generalized multi-view stereo networks, in: *Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1790–1799.
- [7] N. Arya, S. Saha, Generative incomplete multi-view prognosis predictor for breast cancer, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19 (2022) 2252–2263.

- [8] K. Yang, Y. Zheng, K. Lu, K. Chang, N. Wang, Z. Shu, J. Yu, B. Liu, Z. Gao, X. Zhou, Pdgnnet: predicting disease genes using a deep neural network with multi-view features, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19 (2022) 575–584.
- [9] M. Lan, M. Meng, J. Yu, J. Wu, Generalized multi-view collaborative subspace clustering, *IEEE Trans. Circuits Syst. Video Technol.* 32 (2022) 3561–3574.
- [10] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, W. Gao, Late fusion incomplete multi-view clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2018) 2410–2423.
- [11] Y.-M. Xu, C.-D. Wang, J.-H. Lai, Weighted multi-view clustering with feature selection, *Pattern Recognit.* 53 (2016) 25–35.
- [12] L. Cai, H. Wang, F. Jiang, Y. Zhang, Y. Peng, A new clustering mining algorithm for multi-source imbalanced location data, *Inf. Sci.* 584 (2022) 50–64.
- [13] F. Nie, G. Cai, X. Li, Multi-view clustering and semi-supervised classification with adaptive neighbours, in: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017, pp. 2408–2414.
- [14] C. Zhang, Y. Liu, H. Fu, Ae2-nets: autoencoder in autoencoder networks, in: *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2577–2585.
- [15] F. Wu, X. Jing, P. Wei, C. Lan, Y. Ji, G. Jiang, Q. Huang, Semi-supervised multi-view graph convolutional networks with application to webpage classification, *Inf. Sci.* 591 (2022) 142–154.
- [16] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: *Proceedings of the 5th International Conference on Learning Representations*, 2017, pp. 1–14.
- [17] J. Park, M. Lee, H.J. Chang, K. Lee, J.Y. Choi, Symmetric graph convolutional autoencoder for unsupervised graph representation learning, in: *Proceedings of the 17th IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6518–6527.
- [18] X. Zhou, F. Shen, L. Liu, W. Liu, L. Nie, Y. Yang, H.T. Shen, Graph convolutional network hashing, *IEEE Trans. Cybern.* 50 (2018) 1460–1472.
- [19] H. Gao, Z. Wang, S. Ji, Large-scale learnable graph convolutional networks, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018, pp. 1416–1424.
- [20] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, Y. Zhao, Graphsleepnet: adaptive spatial-temporal graph convolutional networks for sleep stage classification, in: *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, 2020, pp. 1324–1330.
- [21] J. Kim, T. Kim, S. Kim, C.D. Yoo, Edge-labeling graph neural network for few-shot learning, in: *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11–20.
- [22] S. Li, W.-T. Li, W. Wang, Co-gcn for multi-view semi-supervised learning, in: *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020, pp. 4691–4698.
- [23] Z. Wu, X. Lin, Z. Lin, Z. Chen, Y. Bai, S. Wang, Interpretable graph convolutional network for multi-view semi-supervised learning, *IEEE Trans. Multimed.* (2023) 1–14, <https://doi.org/10.1109/TMM.2023.3260649>.
- [24] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, in: *Proceedings of the 2nd International Conference on Learning Representations*, 2014, pp. 1–14.
- [25] D.K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, *Appl. Comput. Harmon. Anal.* 30 (2011) 129–150.
- [26] S. Fu, S. Wang, W. Liu, B. Liu, B. Zhou, X. You, Q. Peng, X. Jing, Adaptive graph convolutional collaboration networks for semi-supervised classification, *Inf. Sci.* 611 (2022) 262–276.
- [27] Z. Huang, W. Zhang, D. Wang, Y. Yin, A GAN framework-based dynamic multi-graph convolutional network for origin-destination-based ride-hailing demand prediction, *Inf. Sci.* 601 (2022) 129–146.
- [28] H.T. Phan, N.T. Nguyen, D. Hwang, Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis, *Inf. Sci.* 589 (2022) 416–439.
- [29] X. Xu, T. Wang, Y. Yang, A. Hanjalic, H.T. Shen, Radial graph convolutional network for visual question generation, *IEEE Trans. Neural Netw. Learn. Syst.* 32 (2021) 1654–1667.
- [30] H. Zhong, J. Wu, C. Chen, J. Huang, M. Deng, L. Nie, Z. Lin, X. Hua, Graph contrastive clustering, in: *Proceedings of the 18th IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9204–9213.
- [31] Y. Guo, L. Gao, J. Song, P. Wang, N. Sebe, H.T. Shen, X. Li, Relation regularized scene graph generation, *IEEE Trans. Cybern.* 52 (2022) 5961–5972.
- [32] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, J. Pei, Am-gcn: adaptive multi-channel graph convolutional networks, in: *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2020, pp. 1243–1253.
- [33] S. Wan, S. Pan, J. Yang, C. Gong, Contrastive and generative graph convolutional networks for graph-based semi-supervised learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 10049–10057.
- [34] C. Tang, X. Zheng, W. Zhang, X. Liu, X. Zhu, E. Zhu, Unsupervised feature selection via multiple graph fusion and feature weight learning, *Sci. China Inf. Sci.* (2022) 1–17, <https://doi.org/10.1007/s11432-022-3579-1>.
- [35] C. Tang, Z. Li, J. Wang, X. Liu, W. Zhang, E. Zhu, Unified one-step multi-view spectral clustering, *IEEE Trans. Knowl. Data Eng.* (2022) 1–12, <https://doi.org/10.1109/TKDE.2022.3172687>.
- [36] C. Tang, X. Liu, X. Zhu, J. Xiong, M. Li, J. Xia, X. Wang, L. Wang, Feature selective projection with low-rank embedding and dual Laplacian regularization, *IEEE Trans. Knowl. Data Eng.* 32 (2019) 1747–1760.
- [37] F. Nie, G. Cai, J. Li, X. Li, Auto-weighted multi-view learning for image clustering and semi-supervised classification, *IEEE Trans. Image Process.* 27 (2017) 1501–1511.
- [38] S. Wang, Z. Wang, W. Guo, Accelerated manifold embedding for multi-view semi-supervised classification, *Inf. Sci.* 562 (2021) 438–451.
- [39] S. Ke, L. Zhouchen, Z. Zhanxing, Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes, in: *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020, pp. 5892–5899.
- [40] M. Kan, S. Shan, X. Chen, Multi-view deep network for cross-view classification, in: *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4847–4855.
- [41] J. Li, B. Zhang, G. Lu, D. Zhang, Generative multi-view and multi-feature learning for classification, *Inf. Fusion* 45 (2019) 215–226.
- [42] S. Wang, Z. Chen, S. Du, Z. Lin, Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (2022) 5042–5055.
- [43] Z. Yang, N. Liang, W. Yan, Z. Li, S. Xie, Uniform distribution non-negative matrix factorization for multiview clustering, *IEEE Trans. Cybern.* 51 (2021) 3249–3262.
- [44] X. Jia, X. Jing, X. Zhu, S. Chen, B. Du, Z. Cai, Z. He, D. Yue, Semi-supervised multi-view deep discriminant representation learning, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (2021) 2496–2509.
- [45] A. Huang, Z. Wang, Y. Zheng, T. Zhao, C.-W. Lin, Embedding regularizer learning for multi-view semi-supervised classification, *IEEE Trans. Image Process.* 30 (2021) 6997–7011.
- [46] H. Tao, C. Hou, F. Nie, J. Zhu, D. Yi, Scalable multi-view semi-supervised classification via adaptive regression, *IEEE Trans. Image Process.* 26 (2017) 4283–4296.
- [47] J. Xu, W. Li, X. Liu, D. Zhang, J. Liu, J. Han, Deep embedded complementary and interactive information for multi-view classification, in: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2020, pp. 6494–6501.
- [48] Z. Chen, L. Fu, J. Yao, W. Guo, C. Plant, S. Wang, Learnable graph convolutional network and feature fusion for multi-view learning, *Inf. Fusion* 95 (2023) 109–119.
- [49] G. Lin, K. Liao, B. Sun, Y. Chen, F. Zhao, Dynamic graph fusion label propagation for semi-supervised multi-modality classification, *Pattern Recognit.* 68 (2017) 14–23.