

# Supplementary Materials: Towards Multi-view Consistent Graph Diffusion

Anonymous Authors

## A DETAILS ABOUT THE DERIVATION OF THE ENERGY FUNCTION

$$\begin{aligned}\frac{\partial \tilde{E}(\{\mathbf{H}^{(v)}\}_{v=1}^V)}{\partial \mathbf{H}^{(i)}} &= (\mathbf{I} - \mathbf{S}^{(i)})\mathbf{H}^{(i)} + \frac{1}{2} \left[ - \sum_{m=1}^v \mathbf{P}_{mi}(\mathbf{H}^{(m)} - \mathbf{H}^{(i)}) \right] + \\ &\quad \frac{1}{2} \left[ \sum_{n=1}^v \mathbf{P}_{in}(\mathbf{H}^{(n)} - \mathbf{H}^{(i)}) \right] \\ &= (\mathbf{I} - \mathbf{S}^{(i)})\mathbf{H}^{(i)} + \mathbf{H}^{(i)} \sum_{m=1}^V \mathbf{P}_{mi} - \sum_{m=1}^V \mathbf{P}_{mi}\mathbf{H}^{(m)} \\ &= (2\mathbf{I} - \mathbf{S}^{(i)})\mathbf{H}^{(i)} - \sum_{m=1}^V \mathbf{P}_{im}\mathbf{H}^{(m)}\end{aligned}\quad (1)$$

$$\begin{aligned}\mathbf{H}^{(i,k+1)} &= \mathbf{H}^{(i,k)} - \gamma \frac{\partial \tilde{E}(\{\mathbf{H}^{(v,k)}\}_{v=1}^V)}{\partial \mathbf{H}^{(i,k)}} \\ &= \mathbf{H}^{(i,k)} - \gamma(2\mathbf{I} - \mathbf{S}^{(i)})\mathbf{H}^{(i)} + \gamma \sum_{m=1}^V \mathbf{P}_{im}^{(k)}\mathbf{H}^{(m,k)} \\ &= \left[ (1 - 2\gamma)\mathbf{I} + \gamma\mathbf{S}^{(i,k)} \right] \mathbf{H}^{(i,k)} + \gamma \sum_{m=1}^V \mathbf{P}_{im}^{(k)}\mathbf{H}^{(m,k)}.\end{aligned}\quad (2)$$

## B DETAILS ABOUT DATASETS

### B.1 Multi-view datasets

- BDGP: It contains 2,500 images of drosophila embryos, where 1,750-D visual features and 79-D textual features of each image are extracted.
- Flickr: It consists of 12,154 images covering 7 categories with 1,386 text tags downloaded from the social photography site Flickr, the feature processing reference.
- Caltech102: It is a dataset consisting of 9,144 pictures grouped into 102 categories, including 48-D Gabor features, 49-D WM features, 254-D GENTRIST features, 1,984-D HOG features, 512-D GIST features, and 928-D LBP features.
- GRAZ02: This widely used object dataset comprises images from four different classes and includes six commonly used representations: 512-D GIST features, 225-D WT features, 256-D LBP features, 500-D SIFT features, 500-D SURF features, and 680-D PHOG features.
- HW: It consists of 2,000 pictures categorized into 6 classes, with 153-D Profile-correlation features, 596-D Fourier-coefficient features, 301-D Karhunen-Loeve-coefficient features, 27-D Morphological features, 481-D intensity-averaged features, 157-D Zernike Moment features.
- OutScene: This image dataset contains 2,688 instances categorized into eight classes. It includes 512-D GIST features, 59-D LBP features, 864-D HOG features, and 254-D GENT features.

- Scene15: This scene image dataset comprises 4,485 images categorized into 15 different categories, with three perspectives captured for each image. The feature dimensions for each perspective are 1,800, 1,180, and 1,240, respectively.
- Youtube: This video dataset comprises 2,000 instances in 10 classes, with six views of both visual and audio features. The views include 2,000-D cuboids histogram, 1,024-D motion estimate histogram, 64-D HOG features, 512-D MFCC features, 64-D volume streams, and 647-D spectrogram streams.
- NoisyMNIST: It is comprised of randomly selected 30,000 samples from the MNIST image database in 10 classes. Therein, the given images come with white Gaussian noise of varied intensities.

### B.2 Heterogeneous datasets

- ACM dataset is a citation network comprising 3,025 nodes classified into three types: papers, authors, and topics. These nodes are utilized to build citation networks, study paper contents, and integrate other data. For our experiments, we employ the meta-path set PAP, PSP.
- DBLP dataset is extracted from the DBLP citation network website and contains 334 attributes per node. All nodes are categorized into four types: authors, papers, terms, and conferences. For our experiments, we utilize the meta-path set APA, APCPA, APTPA.
- IMDB dataset is a movie dataset comprising four types of nodes: movie, actor, director, and year. These nodes are categorized into three types based on the movie genre: comedy, light comedy, and drama. Movie features correspond to bag-of-words representation elements of the drama genre. For our experiments, we conduct experiments using the meta-path set MAM, MDM, MYM.
- YELP dataset is a subset of merchant review sites, comprising four types of nodes: business, user, service, and level. For our experiments, we generate the set of meta-paths BUB, BLB, BSB.

## C DETAILS ABOUT COMPARED METHODS

### C.1 Multi-view semi-supervised classification

- HLR-M<sup>2</sup>VS: The framework constructs a unified tensor space to jointly explore the relationships among multiple views using a local geometric structure. We select the weighted factors as  $\lambda_1 = 0.2$  and  $\lambda_2 = 0.4$ .
- ERL-MVSC: The framework integrates diversity, sparsity, and consensus to deftly handle multi-view data with limited labels. We set the smoothing factor  $\alpha = 2$ , the embedding parameter  $\beta = 1$ , the regularization parameter  $\gamma = 1$ , and the fitting coefficient  $\delta = 10$ .

- Co-GCN: The method introduces GCN into multi-view learning and obtains multi-view spectral information by adaptively combining Laplacian matrices. The settings for the graph convolutional layers are 2 and the number of neighbors is 10.
- DSRL: The framework uses a deep sparse regularizer learning model to adaptively learn data-driven sparse regularizers for multi-view clustering and semi-supervised classification. The number of layers is fixed at 10.
- LGCN-FF: The framework considers a joint neural network of both feature and graph fusion. The default setting for hyperparameters controlling the sparsity penalty degree is  $\beta = 1$ .
- IMvGCN: The framework introduces multi-view reconstruction errors paired with Laplace embeddings to capture independence and consistency. The default setting for hyperparameters  $\lambda = 0.5$  and  $\alpha = 1e^{-5}$ .
- PDMF: The framework learns relations and the auxiliary representation through pre-training to tune the mappings from the original data to the comprehensive representation.
- GEGCN: This framework operates by combining the extraction of topological consistency and complementarity with downstream tasks. The default parameters are set to  $\epsilon = 0.05$ .

## C.2 Heterogeneous graph semi-supervised classification

- GCN is a semi-supervised homogeneous graph convolutional network that obtains node embeddings by aggregating message from local neighborhood structures.
- SGC is a simplified version of GCN framework, which only employs the product of high-order adjacency matrices and attribute matrix, removing non-linear transformation for the semi-supervised classification tasks.
- HAN explores the node-level and semantic-level attention on multiplex networks to learn the importance of nodes and meta-paths, thereby generating node representations in a hierarchical manner.
- DGI is an unsupervised graph learning representation approach that maximizes mutual information between the graph-level summary embeddings and the local patches to obtain global graph structures.
- DMGI is an unsupervised attributed multiplex network that jointly integrates the node embeddings from multiple relations to learn high-quality representations through a consensus regularization framework and a universal discriminator for downstream tasks.
- IGNN is a graph learning framework which employs a fixed-point equilibrium equation and the Perron-Frobenius theory to iterate graph convolutional aggregation until converging for node classification tasks.
- SSDCM is a semi-supervised framework for representation learning which aims to maximize the mutual information between local and contextualized global graph summaries and employs the cross-layer links to impose the regularization of the node embeddings.

- MHGCN automatically learns the useful relation-aware topological structural signals by the multiplex relation aggregation and a multi-layer graph convolution for graph representation learning tasks.

## D DETAILED EXPERIMENTAL RESULTS

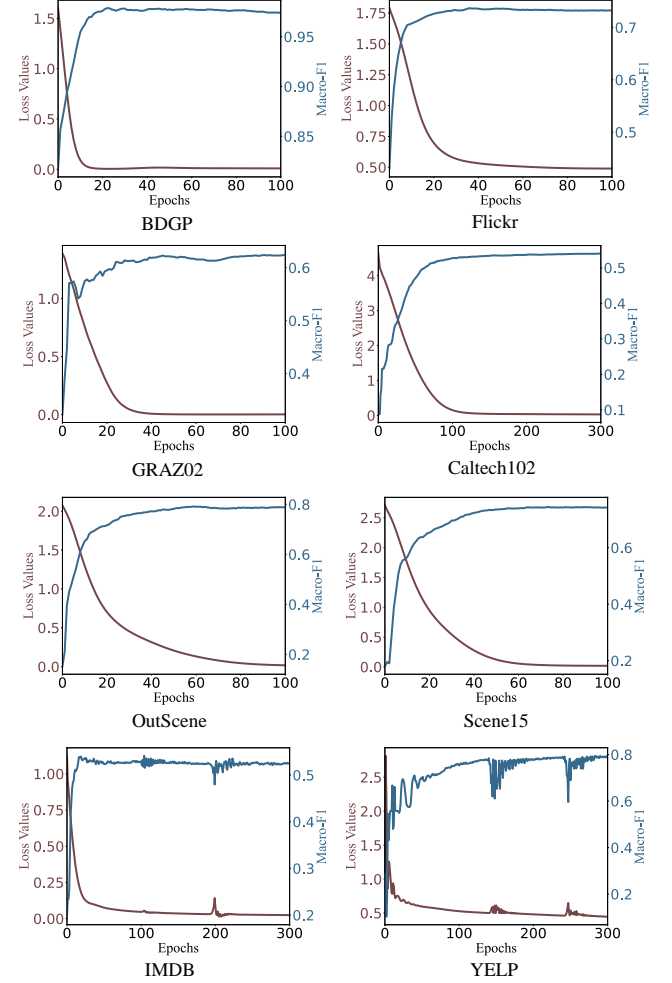


Figure 1: Loss and accuracy / macro-F1 curves of ECMGD.

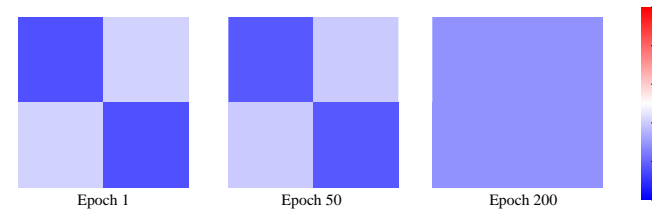


Figure 2: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the BDGP dataset.

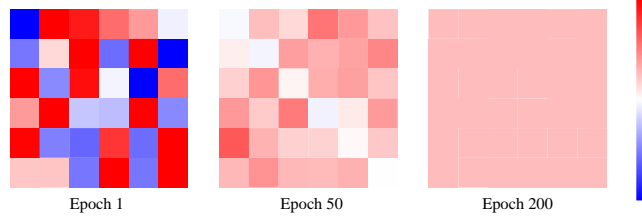


Figure 8: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the Youtube dataset.

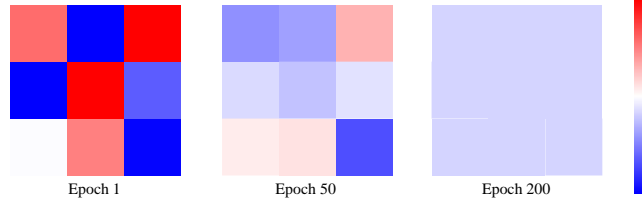


Figure 9: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the ACM dataset.

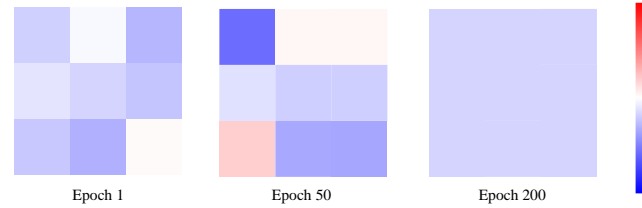


Figure 10: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the DBLP dataset.

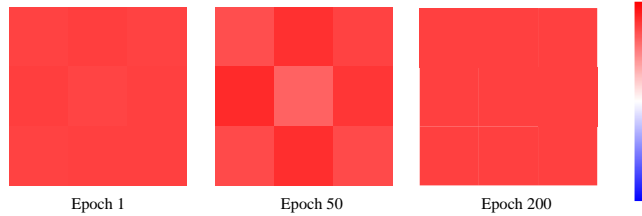


Figure 11: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the IMDB dataset.

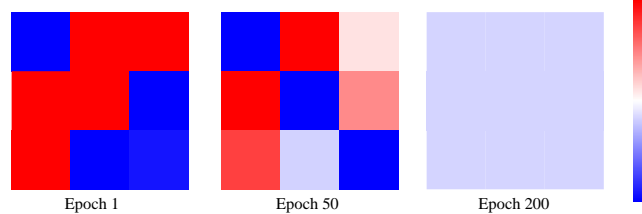


Figure 12: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the YELP dataset.

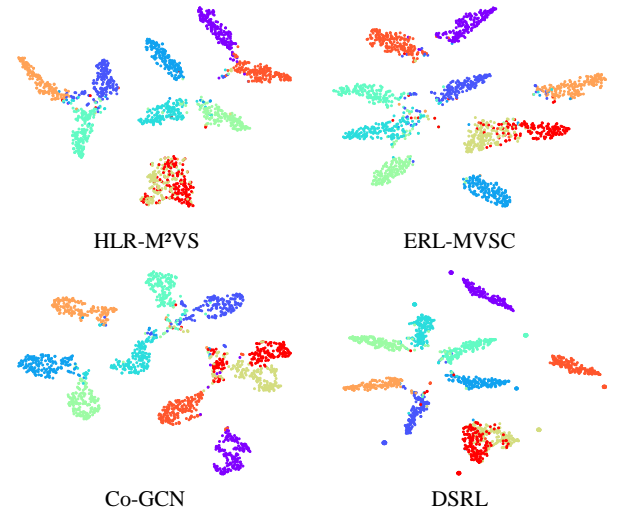


Figure 13: T-sne visualization of HLR-M<sup>2</sup>VS, ERL-MVSC, Co-GCN, and DSRL on dataset HW.

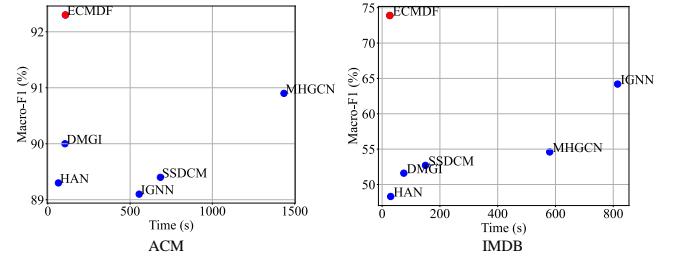


Figure 14: Running time (seconds) of compared HGNNs with 500 training epochs on dataset DBLP and YELP.

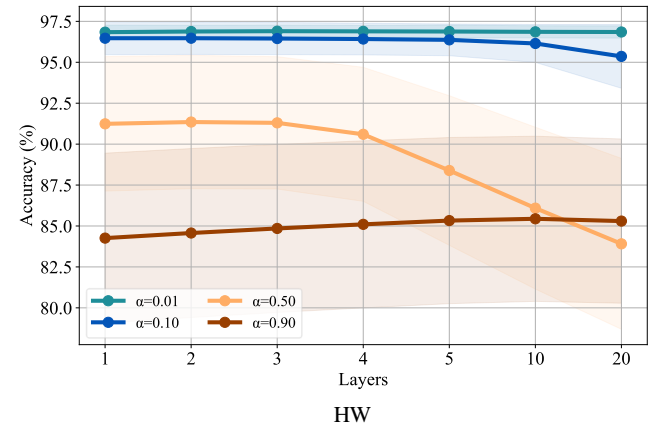


Figure 15: The classification accuracy of ECMGD w.r.t hyper-parameters  $\alpha$  and  $K$  on the HW dataset.

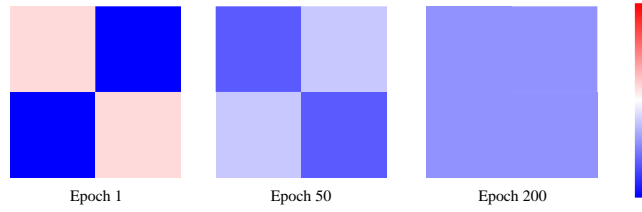


Figure 3: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the Flickr dataset.

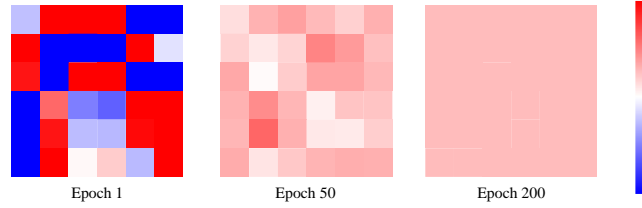


Figure 4: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the Caltech102 dataset.

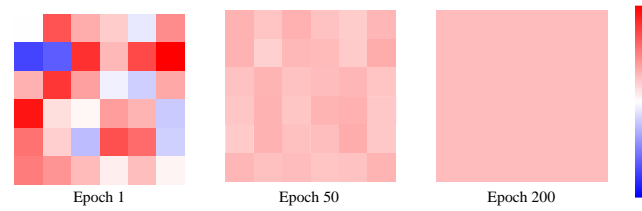


Figure 5: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the GRAZ02 dataset.

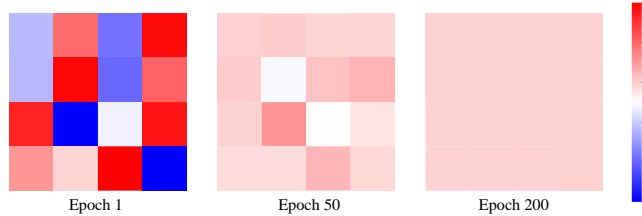


Figure 6: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the OutScene dataset.

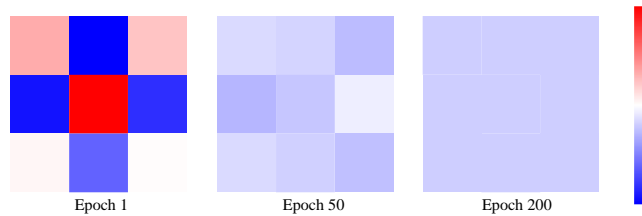


Figure 7: Visualization of the inter-view diffusion matrix  $P$  of ECMGD at various epochs on the Scene15 dataset.

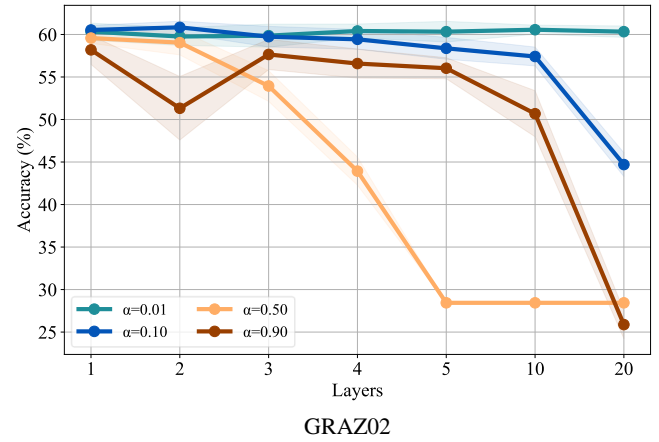


Figure 16: The classification accuracy of ECMGD w.r.t hyper-parameters  $\alpha$  and  $K$  on the GRAZ02 dataset.

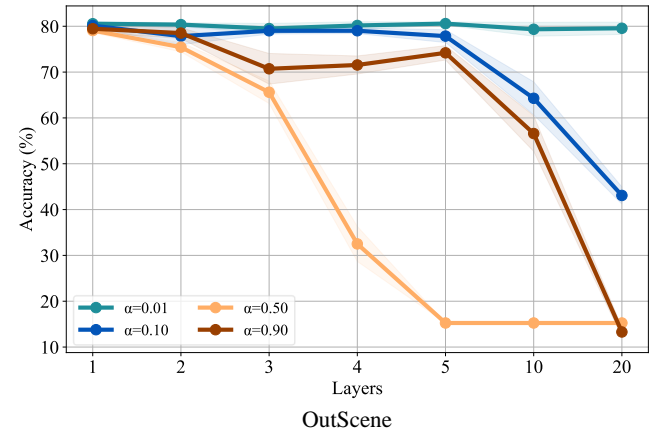


Figure 17: The classification accuracy of ECMGD w.r.t hyper-parameters  $\alpha$  and  $K$  on the OutScene dataset.

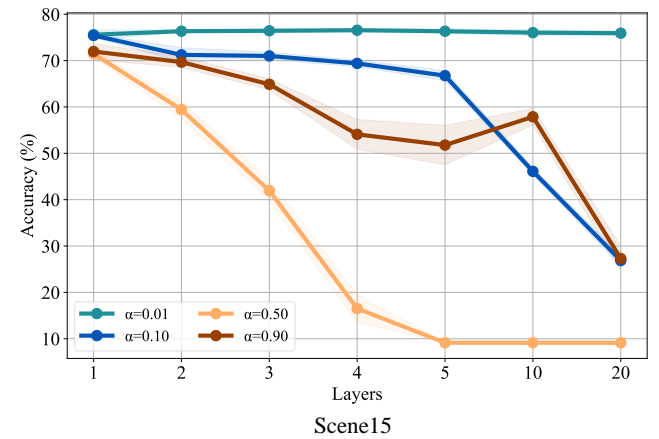


Figure 18: The classification accuracy of ECMGD w.r.t hyper-parameters  $\alpha$  and  $K$  on Scene15 dataset.

## REFERENCES