Appendix for Generalized Deep Multi-view Clustering via Causal Learning with Partially Aligned Cross-view Correspondence

Xihong Yang[1,3], Siwei Wang[2], Jiaqi Jin[1], Fangdi Wang[1], Tianrui Liu[1],
, Yueming Jin[3], Xinwang Liu[1,*], En Zhu[1,*], Kunlun He[4]
[1]College of Computer Science and Technology, National University of Defense Technology, Changsha, China [2]Intelligent Game and Decision Lab, Beijing, China
[3]National University of Singapore, Singapore [4]Chinese PLA General Hospital, Beijing, China
yangxihong@nudt.edu.cn

1. Preliminary

In this paper, we address the multi-view clustering task in scenarios with partial alignment. The raw features from different views represent various descriptions of the same sample. Therefore, in this setting, we consider the aligned data as the variant feature for the same sample cross different views. For a given dataset $\{x^{(v)}\}_{v=1}^{V}$, we randomly spit the dataset into two partitions with 50% ratio, i.e., the aligned data and unaligned data. Let x_{va} and x_{in} denote the variant and invariant features, respectively. Here, we consider the unaligned data as the shift phenomenon compared with the aligned scenario, which is denoted as x'_{va}. The invariant features x_{in} are obtained by the encoder network $\mathcal{F}_{\phi(\cdot)}$. Then, we obtain the extracted variant representations e_{va} and invariant representations e_{in} by the encoder networks $\mathcal{G}_{\theta_1}$ and $\mathcal{G}_{\theta_2}$. Moreover, we define the clustering results as r. The fundamental notations used in this paper are outlined in Tab. 1.

Notation	Meaning
x_{va}	Fully Aligned Variant Features
x_{in}	Invariant Features
x'_{va}	Partially Aligned Variant Features
e_{va}	Extracted Variant Representations
e_{in}	Extracted Invariant Representations
r	Clustering Results
$\mathcal{F}_{\phi}(\cdot)$	Variantional Auto-Encoder Network
$\mathcal{G}_{\theta}(\cdot)$	Post-Intervention Inference Network
V	The Number of Views
N	The Number of Samples
Z	The Similarity Matrix

Table 1. Basic notations used in the whole paper.

2. Related Work

2.1. Multi-view Clustering

Recently, Multi-view Clustering (MVC) has garnered significant attention [4, 10, 15, 18, 25, 30, 34]. Existing MVC methods can be broadly categorized into two groups based on cross-view correspondence: MVC with fully aligned data and MVC with partially aligned data. Fully aligned data implies predefined mapping relationships for every pair of cross-view data. There are several works under this assumption, which can be encompassed in five main categories: (1) Non-negative matrix factorization-based MVC [19] aims to identify a shared latent factor, which is used to process information from multi-view input. (2) Kernel learning-based MVC [11, 12] involves predefining a base kernels set for each views. After that, this method optimally fuse the weights of the kernels to improve clustering outcomes. (3) Subspace-based MVC [9] is based on the assumption that all views in the multi-view task share a low-dimensional latent space, with the final outcomes derived from learning this shared representation. (4) Benefiting from the successful of graph learning and its applications [13, 14, 21, 24, 26–29, 31, 32], Graph-based MVC [8] seeks to constructing a unified graph from multiple views, with clustering results derived from spectral decomposition. (5) Thanks to the robust representational capabilities of deep networks, deep neural network-based MVC [6, 25] has the capacity to extract more sophisticated representations. through neural networks. Despite achieving promising clustering performance, most of these methods heavily rely on the assumption that cross-view data are fully aligned.

To tackle this issue, many MVC algorithms have been proposed [5, 20, 22, 23]. PVC is designed to use a differentiable surrogate of the non-differentiable Hungarian algorithm to learn the correspondence of partially aligned

*Corresponding author

data. MVC-UM [33], based on non-negative matrix factorization, learns the correspondence by exploring cross-view relationships. SURE [23] uses available pairs as positives and randomly selects some cross-view samples as negatives. UPMGC-SM [20] leverages structural information from each view to refine cross-view correspondences. In contrast to the above methods, we approach partially aligned data from a causal perspective, aiming to improve the generalization of the model.

2.2. Causal Disentangled Representation Learning

Traditional approaches for disentangled representation learning focus on examining mutually independent latent factors through the use of encoder-decoder networks. In this approach, a standard normal distribution is utilized as the prior for the latent code. Moreover, the variational posterior $q(z|x)$ is employed to approximate the unknown posterior $p(z|x)$. β-VAE [3] introduces an adaptive framework to adjust the weight of the KL term. Factor VAE [1] designs a framework, which focuses solely on the independence of factors. After that, the exploration of causal graphs from observations has gained significant attention, leveraging either purely observational data or a combination of observational and interventional data. NOTEARs [35] incorporates a novel Directed Acyclic Graph (DAG) constraint for causal learning. LiNGAM [16] ensures the identifiability of the model based on the assumptions of linear relationships and non-Gaussianity. In cases where interventions are feasible, Heckerman et al. [2] demonstrate the causal structure learned from interventional data can be identified. More recently, there has been an increasing interest in combining causality and disentangled representation. Suter et al. [17] employs causality to explain disentangled latent representations, while Kocaoglu et al. [7] introduces CausalGAN, a method supporting ”do-operations” on images. Drawing inspiration from the success of causal learning, we apply causal modeling to multi-view clustering. To the best of our knowledge, our work represents the first attempt to leverage causal learning to improve model generalization with partially aligned data in the multi-view clustering task.

3. Detailed of CauMVC

3.1. Algorithm

Due to page limitations, we provide the algorithm table for CauMVC in this section.

3.2. Datasets

3.3. Hyper-parameter Settings

To ensure reproducibility, we provide a summary of the statistics and the hyper-parameter settings of our proposed method in Tab. 2.

Algorithm 1 Inference Pipeline of CauMVC with Partially Aligned Data

Input: The partially aligned data x'_{va}; the interation number I
Output: The clustering result r.

```
1: for i = 1 to I do
2:     Obtain the invariant features x_in by F_φ(·) with Eq. (10).
3:     Encoder the representations e'_va and e_in by G_θ1 and G_θ2.
4:     Obtain the post-intervention inference r with Eq. (11).
5:     Calculate the ELBO loss, contrastive loss, and
       reconstruction loss with Eq. (9), (13) and (14).
6:     Calculate the total loss L by Eq. (15).
7:     Update model by minimizing L with Adam optimizer.
8: end for
9: return r
```

4. Additional Experiments

4.1. Ablation Studies

In this section, we first present the ablation study on all datasets under the partially aligned scenario. “(w/o) Cau”, “(w/o) Con”, “(w/o) Cau&Con”, and “Ours” represent the ablated models where the causal module, the contrastive regularizer, and both modules combined, respectively, are individually removed. In the “(w/o) Cau&Con” configuration, we employ an autoencoder network as the backbone to derive representations for the downstream clustering task. Consistent with the conclusions drawn in the main text, the results are summarized as follows.

- We resort the multi-view clustering from the causal perspective. The model generalization is improved when the input is partially aligned data, thus achieving promising performance.
- The contrastive module could push the positive sample close, and pull the negative sample away, enhancing the model’s discriminative capacity. The model could achieve better clustering outcomes.

Besides, we perform ablation studies with fully aligned data to assess the effectiveness of our designed modules, namely the causal module and contrastive regularizer. Specifically, ”(w/o) Cau,” ”(w/o) Con,” ”(w/o) Cau&Con,” and ”Ours” denote reduced models created by individually omitting the causal module, the contrastive regularizer, and both modules together. In this paper, we employ an autoencoder network as the core architecture to derive representations for the subsequent clustering task, referred to as ”(w/o) Cau&Con,”. The outcomes are depicted in Fig. 2. These results clearly demonstrate that the exclusion of any of the designed modules leads to a significant decrease in clustering performance, underscoring the essential role each module plays in optimizing overall performance.

	Dataset	BBCSport	Movies	WebKB	Reuters	Caltech101-7	UCI-digit	SUNRGB-D	STL-10
Statistics	Samples	544	617	1051	1200	1400	2000	10335	13000
	Clusters	5	17	2	6	7	10	45	10
	Views	2	2	2	5	5	3	3	4
Hyper-parameters	α	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	β	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Learning Rate	0.003	0.005	0.003	0.003	0.003	0.003	0.003	0.003

Table 2. Statistics and hyper-parameter settings of eight benchmark datasets.

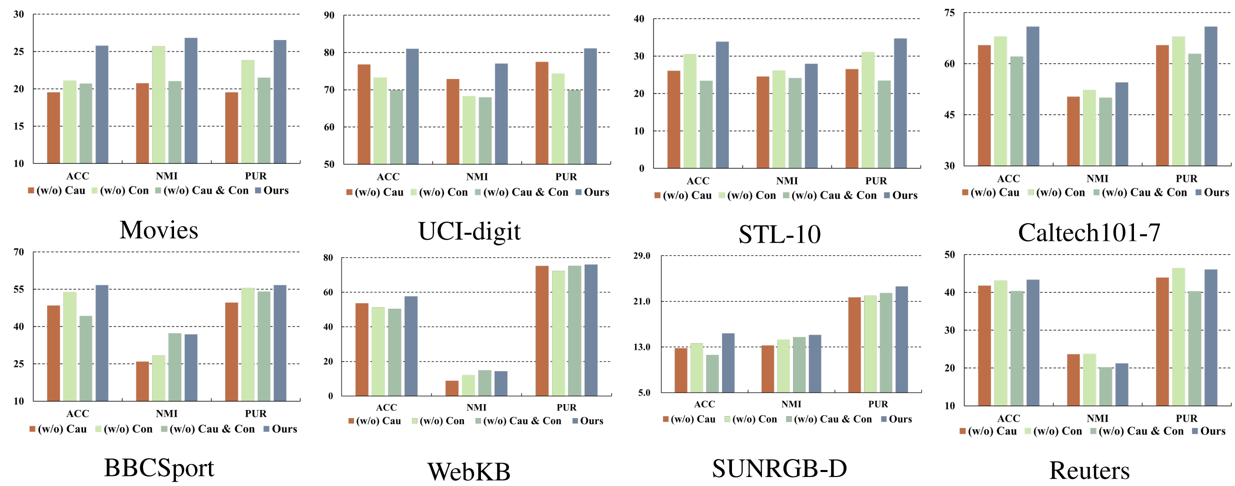

Figure 1. Ablation studies on eight datasets with partially aligned data. "(w/o) Cau", "(w/o) Con", "(w/o) Cau&Con", and "Ours" correspond to reduced models by individually removing the causal module, the contrastive regularizer, and all aforementioned modules combined, respectively.

4.2. Different Align Ratio

To evaluate the performance of CauMVC under different alignment ratios, we conduct experiments on eight datasets, with the results presented in Fig. 3, Fig. 4, and Fig. 5. The results clearly demonstrate that CauMVC outperforms other baseline models across various alignment ratios in most scenarios. This highlights its strong generalization capability in handling partially aligned data effectively.

4.3. Sensitivity Analysis of α and β

To further investigate the impact of the parameters α and β on our model, we conduct experiments on the BBCSport dataset, analyzing parameter values within the range of $\{0.01, 0.1, 1.0, 10, 100\}$. Due to space constraints, the experimental results for the BBCSport dataset are provided in the Appendix. Based on the results presented in Fig. 7, we draw the following observations:

- When α and β are assigned extreme values (0.1 or 100), the clustering performance tends to degrade. We hypothesize that this decline results from an imbalance in the loss function. Moreover, the model achieves optimal performance when the trade-off parameters are set around 1.0.
- The results also indicate that α has a more significant impact on model performance, suggesting that the causal model plays a crucial role in enhancing the overall effectiveness of the approach.

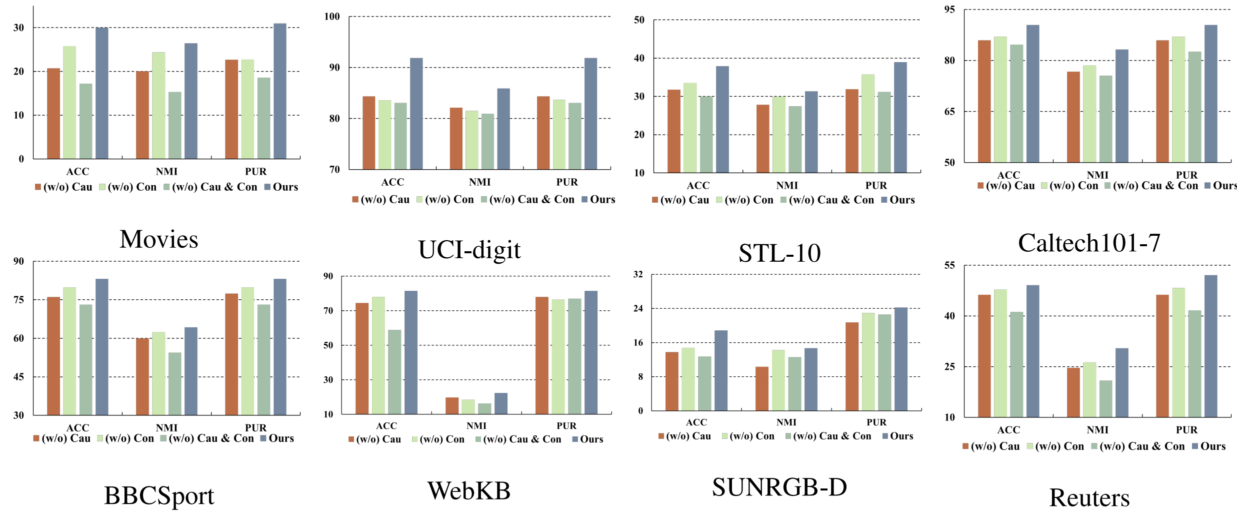

Figure 2. Ablation studies on eight datasets with fully aligned data. “(w/o) Cau”, “(w/o) Con”, “(w/o) Cau&Con”, and “Ours” correspond to reduced models by individually removing the causal module, the contrastive regularizer, and all aforementioned modules combined, respectively.

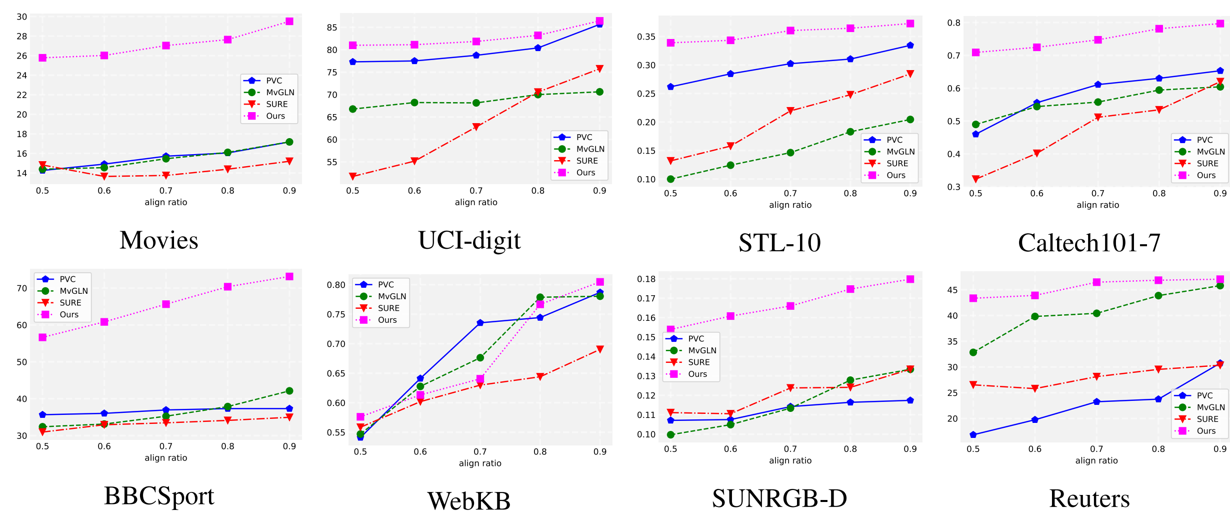

Figure 3. Clustering performance on eight Datasets with different aligned ratios in accuracy metrics.

References

[1] Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Advances in neural information processing systems*, 31, 2018. 2

[2] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20:197–243, 1995. 2

[3] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2016. 2

[4] Dong Huang, Chang-Dong Wang, Jian-Sheng Wu, Jian-Huang Lai, and Chee-Keong Kwoh. Ultra-scalable spectral clustering and ensemble clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1212–1226, 2019. 1

[5] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering. *Advances in Neural Information Processing Systems*, 33:2892–2902, 2020. 1

[6] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. *arXiv preprint arXiv:2303.15689*, 2023. 1

[7] Murat Kocaoglu, Christopher Snyder, Alexandros G Di-

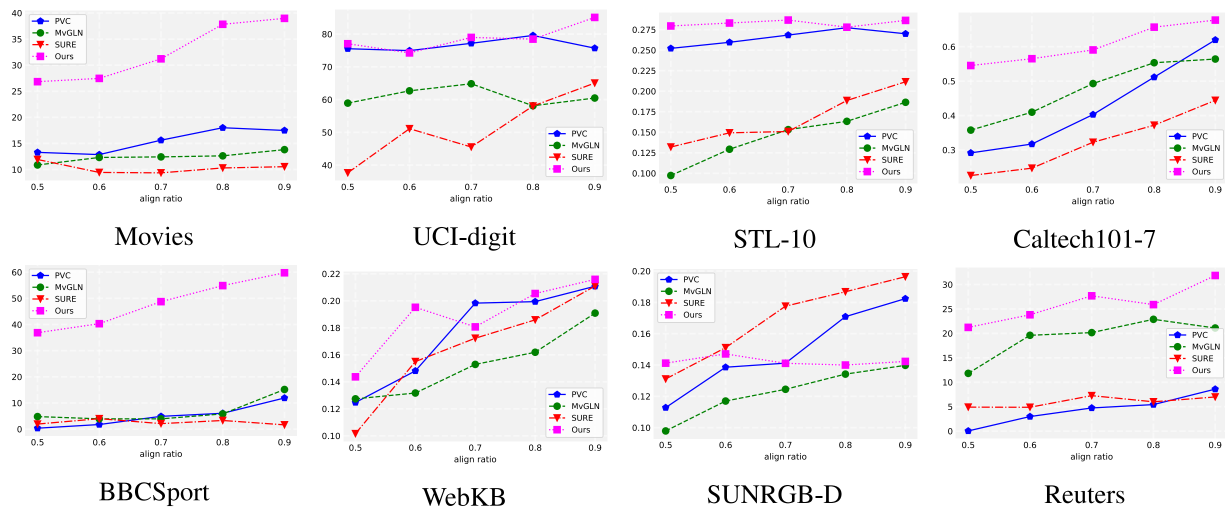

Figure 4. Clustering performance on eight Datasets with different aligned ratios in NMI metrics.

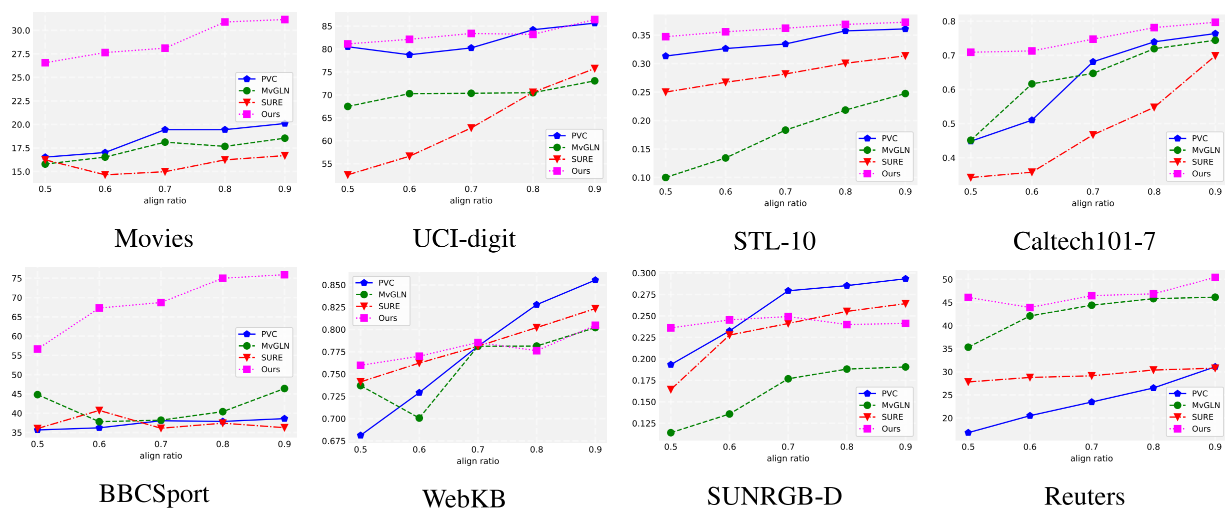

Figure 5. Clustering performance on eight Datasets with different aligned ratios in PUR metrics.

makis, and Sriram Vishwanath. Causalgan: Learning causal implicit generative models with adversarial training. *arXiv preprint arXiv:1709.02023*, 2017. 2

[8] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu. Consensus graph learning for multi-view clustering. *IEEE Transactions on Multimedia*, 24: 2461–2472, 2021. 1

[9] Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7576–7584, 2022. 1

[10] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024. 1

[11] Xinwang Liu. Simplemkkm: Simple multiple kernel k-means. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1

[12] Xinwang Liu, Li Liu, Qing Liao, Siwei Wang, Yi Zhang, Wenxuan Tu, Chang Tang, Jiyuan Liu, and En Zhu. One pass late fusion multi-view clustering. In *International Conference on Machine Learning*, pages 6850–6859. PMLR, 2021. 1

[13] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. In *AAAI Conference on Artificial Intelligence*, 2022. 1

[14] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Zhen Wang, Ke Liang, Wenxuan Tu, Liang Li, Jingcan Duan, and Cancan Chen. Hard sample aware network for contrastive deep graph clustering. In *Proceedings of the AAAI confer-*

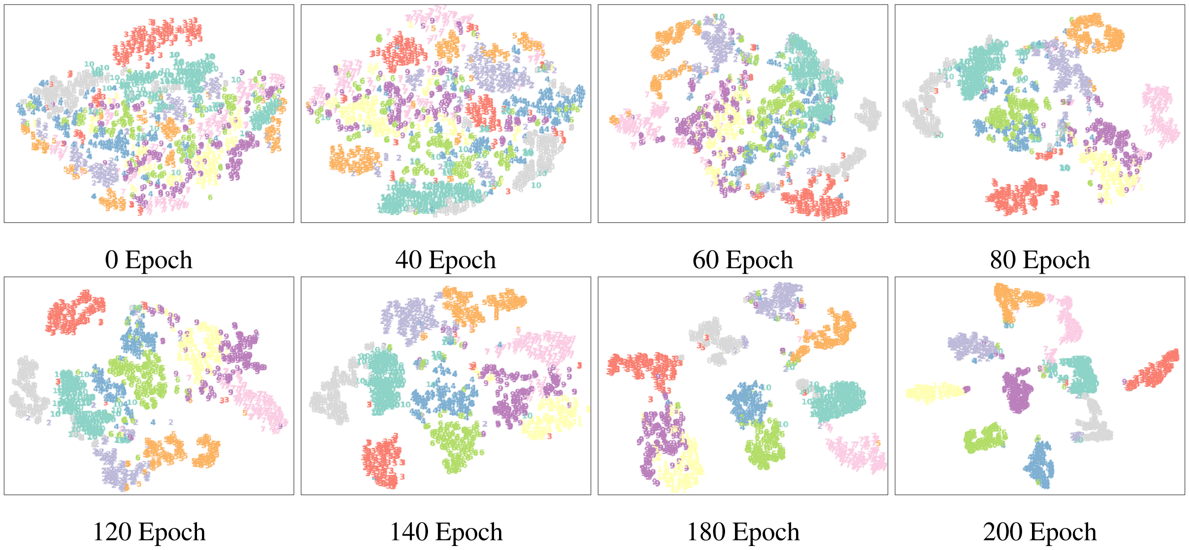

Figure 6. Visualization of the representations during the training process on UCI-digit dataset.

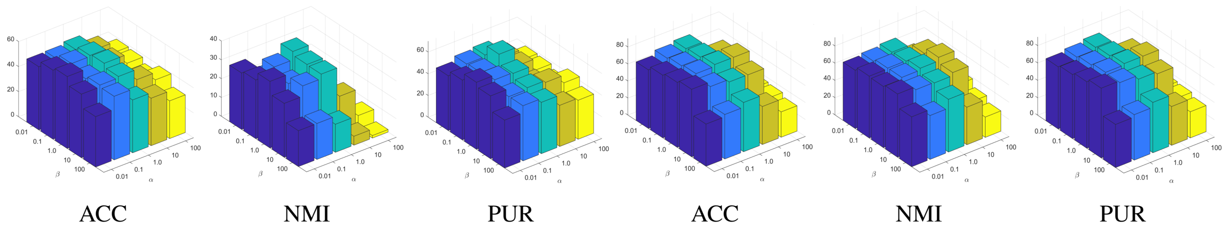

Figure 7. Sensitivity analysis of the hyper-parameter α and β on BBCSport and UCI-digit datasets.

ence on artificial intelligence, pages 8914–8922, 2023. 1

[15] Zhen Long, Ce Zhu, Pierre Comon, Yazhou Ren, and Yipeng Liu. Feature space recovery for efficient incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2023. 1

[16] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006. 2

[17] Raphael Suter, Djordje Miladinovic, Bernhard Schölkopf, and Stefan Bauer. Robustly disentangled causal mechanisms: Validating deep representations for interventional robustness. In *International Conference on Machine Learning*, pages 6056–6065. PMLR, 2019. 2

[18] Xinhang Wan, Jiyuan Liu, Weixuan Liang, Xinwang Liu, Yi Wen, and En Zhu. Continual multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3676–3684, 2022. 1

[19] Jie Wen, Zheng Zhang, Yong Xu, and Zuofeng Zhong. Incomplete multi-view clustering via graph regularized matrix factorization. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1

[20] Yi Wen, Siwei Wang, Qing Liao, Weixuan Liang, Ke Liang, Xinhang Wan, and Xinwang Liu. Unpaired multi-view graph clustering with cross-view structure matching. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1, 2

[21] Yang Xihong, Jing Heming, Zhang Zixing, Wang Jindong, Niu Huakang, Wang Shuaiqiang, Lu Yu, Wang Junfeng, Yin Dawei, Liu Xinwang, Zhu En, Lian Defu, and Min Erxue. Darec: A disentangled alignment framework for large language model and recommender system. In *2025 IEEE 41rd International Conference on Data Engineering (ICDE)*. IEEE, 2025. 1

[22] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1134–1143, 2021. 1

[23] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022. 1, 2

[24] Xihong Yang, Xiaochang Hu, Sihang Zhou, Xinwang Liu, and En Zhu. Interpolation-based contrastive learning for few-label semi-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–12, 2022. 1

[25] Xihong Yang, Jiaqi Jin, Siwei Wang, Ke Liang, Yue Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En

Zhu. Dealmvc: Dual contrastive calibration for multi-view clustering. In *Proceedings of the 31th ACM International Conference on Multimedia*, 2023. 1

[26] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Wenxuan Tu, Qun Zheng, Xinwang Liu, Liming Fang, and En Zhu. Cluster-guided contrastive graph clustering network. In *Proceedings of the AAAI conference on artificial intelligence*, pages 10834–10842, 2023. 1

[27] Xihong Yang, Cheng Tan, Yue Liu, Ke Liang, Siwei Wang, Sihang Zhou, Jun Xia, Stan Z Li, Xinwang Liu, and En Zhu. Convert: Contrastive graph clustering with reliable augmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 319–327, 2023.

[28] Xihong Yang, Erxue Min, Ke Liang, Yue Liu, Siwei Wang, Sihang Zhou, Huijun Wu, Xinwang Liu, and En Zhu. Graphlearner: Graph node clustering with fully learnable augmentation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5517–5526, 2024.

[29] Xihong Yang, Yiqi Wang, Yue Liu, Yi Wen, Lingyuan Meng, Sihang Zhou, Xinwang Liu, and En Zhu. Mixed graph contrastive network for semi-supervised node classification. *ACM Transactions on Knowledge Discovery from Data*, 2024. 1

[30] Xihong Yang, Siwei Wang, Fangdi Wang, Jiaqi Jin, Suyuan Liu, Yue Liu, En Zhu, Xinwang Liu, and Yueming Jin. Automatically identify and rectify: Robust deep contrastive multi-view clustering in noisy scenarios. In *International Conference on Machine Learning*. PMLR, 2025. 1

[31] Mingjia Yin, Hao Wang, Xiang Xu, Likang Wu, Sirui Zhao, Wei Guo, Yong Liu, Ruiming Tang, Defu Lian, and Enhong Chen. Apgl4sr: A generic framework with adaptive and personalized global collaborative information in sequential recommendation. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pages 3009–3019, 2023. 1

[32] Mingjia Yin, Hao Wang, Wei Guo, Yong Liu, Suojuan Zhang, Sirui Zhao, Defu Lian, and Enhong Chen. Dataset regeneration for sequential recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3954–3965, 2024. 1

[33] Hong Yu, Jia Tang, Guoyin Wang, and Xinbo Gao. A novel multi-view clustering method for unknown mapping relationships between cross-view samples. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2075–2083, 2021. 2

[34] Qun Zheng, Xihong Yang, Siwei Wang, Xinru An, and Qi Liu. Asymmetric double-winged multi-view clustering network for exploring diverse and consistent information. *Neural Networks*, 179:106563, 2024. 1

[35] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems*, 31, 2018. 2