HIGHER ORDER MULTIPLE GRAPH FILTERING FOR STRUCTURED GRAPH LEARNING

Liang Du[⋆]*, Xiaodong Li*[⋆]*, Yan Chen*[†]*, Gui Yang*[⋆]*, Mian Ilyas Ahmad*[⋆⋆]*, Peng Zhou*[††]

[⋆] School of Computer and Information Technology, Shanxi University, China

[†] School of Computer Science and Technology, Anhui University, China

[††] School of Computer Science, Sichuan University, China

[⋆⋆] National University of Sciences and Technology, Pakistan

ABSTRACT

In the field of machine learning, multi-view clustering endeavors to unveil concealed clustering patterns across diverse data perspectives. Traditional methodologies often fall short due to their reliance on low-order similarity data, rendering them susceptible to performance instability and vulnerability to noise and heterogeneity. In this context, we present a novel method that integrates multiple graph filter learning, approximated by Chebyshev polynomials, and consensus structural graph learning into a unified framework. This approach harnesses high-order statistical information from multiple data sources, enhancing multi-view clustering while introducing adaptive parameterization for the graph filters to ensure consistency across views. Our comprehensive experiments, conducted on multiple datasets, consistently demonstrate substantial performance improvements compared to conventional approaches.

Index Terms— Multi-view clustering, high-order graph filtering, similarity graph, clustering performance

1. INTRODUCTION

Multiple graph clustering (MGC) is a crucial domain in machine learning that focuses on clustering data points represented by multiple graphs or networks. These graphs capture different facets of the underlying data, and the objective is to uncover meaningful clusters that incorporate information from all the graphs [1, 2].

Nevertheless, conventional multiple graph clustering methods come with notable limitations. One major drawback is their failure to consider high-order relationships among data points. These methods typically rely on low-order pairwise similarities between data points, which often fall short in capturing intricate higher-order dependencies within the data [3]. This limitation can result in incomplete or inaccurate clustering outcomes, especially when dealing with data characterized by complex relationships that extend beyond pairwise comparisons.

Another common challenge is the lack of mechanisms to refine the quality of input graphs during the clustering process. Real-world scenarios frequently involve input graphs [4], contaminated with noise or irrelevant data, and traditional methods do not effectively address or enhance the quality of these graphs. Consequently, clustering performance may suffer due to noisy or poorly constructed graphs, leading to suboptimal results.

To enhance clustering performance, we propose a graph-based learning approach that leverages both complementary information between views [5] and local similarities within each view. We achieve this by using high-order Chebyshev polynomials [6]. In our method, we unify the learning of the consensus matrix and graph filters, seamlessly integrating these two processes into a unified optimization framework. In this framework, the consensus graph guides the learning of graph filters, approximated by Chebyshev polynomials in each candidate graph. Simultaneously, we refine the candidate higher-order graph filters to capture long-range relationships effectively from the vertex domain and consider the influence of eigenvalues from the spectral domain [7]. This refinement results in a better consensus graph, enabling the learned consensus matrix to guide the learning of graph filters effectively, and vice versa. In the subsequent optimization steps, we introduce a simple alternative algorithm to efficiently optimize our objective, learning both an optimal consensus graph and optimal filters. Our method's superiority is validated through extensive experiments on multi-view data, consistently outperforming most existing multi-view clustering approaches. These innovations significantly enhance the accuracy and robustness of clustering results, particularly when dealing with intricate and noisy real-world data. Furthermore, this method holds potential practical value in domains such as molecular chemistry and social networks, where modeling complex relationships is essential.

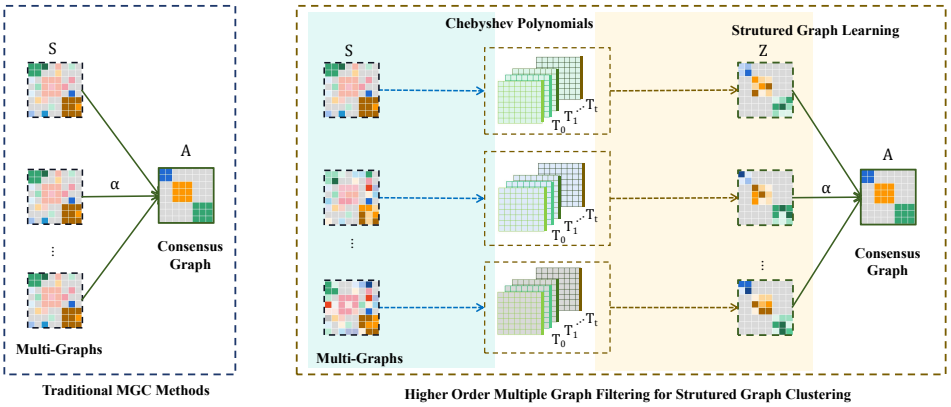


Fig. 1. Example of placing a figure with experimental results.

The main contributions of this paper are as follows:

- **Enhanced High-Order Graph Filtering**: Our method introduces high-order graph filters using Chebyshev polynomials, effectively capturing complex relationships among data points, a critical feature often lacking in traditional approaches.
- **Unified Optimization Framework**: We unify the learning of the consensus matrix and graph filters, establishing a bidirectional learning process that improves both components. This framework enhances the quality of the consensus graph and the accuracy of the graph filters.
- **Superior Multi-View Clustering Performance**: Through extensive experiments on multi-view data, our method consistently outperforms existing multi-view clustering approaches. It offers robust and efficient clustering results, making it a valuable tool for real-world applications.

2. THE PROPOSED METHOD

As mentioned earlier, traditional methods often face limitations when relying on low-order similarities in the data, leading to unstable performance and susceptibility to noise and heterogeneity. In this paper,We integrate multiple graph filter learning techniques using Chebyshev polynomials and consensus structural graph learning into a unified framework, as illustrated in Fig. 1, enhancing our ability to capture complex relationships and patterns in graph data.

2.1. First Order Neighbor Graph Construction

For a multi-view dataset consisting of v views, denoted as $\{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(v)}\}$, where $\mathbf{X}^{(v)} \in \mathbf{R}^{n \times d_v}$, with n representing the number of instances, and d_v indicating the feature dimension in the v-th view, it is crucial to construct high-quality graphs $\{\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(v)}\}$, with $\mathbf{S}^{(v)} \in \mathbf{R}^{n \times n}$, for further clustering. In this paper, we adopt a simple yet effective neighbor assignment method [6]. The entries $\mathbf{s}_{ij}$ of the similarity graph matrix $\mathbf{S}^v$ can be obtained as follows

$$\mathbf{S}^v_{i,j} = \begin{cases} \frac{\mathbf{h}(\mathbf{x}_i^{(v)}, \mathbf{x}_{k+1}^{(v)}) - \mathbf{h}(\mathbf{x}_i^{(v)}, \mathbf{x}_j^{(v)})}{\sum_{j'=1}^{k} (\mathbf{h}(\mathbf{x}_i^{(v)}, \mathbf{x}_{k+1}^{(v)}) - \mathbf{h}(\mathbf{x}_i^{(v)}, \mathbf{x}_{j'}^{(v)}))}, & j \leq k, \\ 0, & j > k. \end{cases} \quad (1)$$

where $\mathbf{h}(\cdot, \cdot)$ measures the Euclidean distance between two samples and k is the neighborhood size.

2.2. Higher Order Chebyshev Polynomials

Given an affinity matrix $\mathbf{S}$, which represents pairwise similarities or weights between nodes in a graph, Chebyshev polynomials $\{T_i(\mathbf{S})\}$ are defined as a sequence of polynomials, where i varies from 0 onwards. These polynomials are computed recursively as follows:

$$\begin{aligned} T_0(\mathbf{S}) &= \mathbf{I}, \\ T_1(\mathbf{S}) &= \mathbf{S}, \\ T_t(\mathbf{S}) &= 2\mathbf{S}T_{t-1}(\mathbf{S}) - T_{t-2}(\mathbf{S}), \quad \text{for } t \geq 2. \end{aligned}$$

Here, $\mathbf{I}$ signifies the identity matrix, and $\mathbf{S}$ represents the affinity matrix. These Chebyshev polynomials are element-wise operations on $\mathbf{S}$, finding applications in various graph signal processing tasks.

In contrast to the first-order affinity graph $\mathbf{S}$, Chebyshev polynomials offer distinct advantages in graph signal processing. They provide a higher-order representation of graph data, allowing for the capture of intricate relationships beyond pairwise connections. Chebyshev filtering proves computationally efficient, especially for large graphs, while maintaining numerical stability to reduce potential errors. Moreover, Chebyshev filters enable localized filtering, facilitating the examination of specific local patterns within the graph, and they efficiently extract spectral information,

making them versatile tools for diverse graph-based applications. These advantages establish Chebyshev polynomials as a potent choice for graph signal processing tasks.

In practical applications, Chebyshev polynomials can be employed to filter a graph signal $\mathbf{x}$ using the following expression:

$$\mathbf{y} = \sum_{t=0}^{\hat{t}} \beta_t T_t(\mathbf{S})\mathbf{x} \tag{2}$$

Here, $\mathbf{y}$ represents the filtered signal, β_t denotes filter coefficients, and $\hat{t}$ signifies the order of the Chebyshev filter.

2.3. Multiple Higher Order Structural graph learning

For each affinity graph, we can build a set of Chebyshev polynomials in the form of

$$\left\{ \sum_{t=0}^{\hat{t}} \beta_t^1 T_t(\mathbf{S}^{(1)}), \sum_{t=0}^{\hat{t}} \beta_t^2 T_t(\mathbf{S}^{(2)}), \ldots, \sum_{t=0}^{\hat{t}} \beta_t^{(v)} T_t(\mathbf{S}^{(v)}) \right\}. \tag{3}$$

We aim to minimize the weighted sum of squared differences between the linear combinations of Chebyshev polynomials applied to the affinity matrices $\mathbf{S}^{(v)}$ and the final consensus affinity matrix $\mathbf{A}$ by solving the following optimization problem for the v-th view:

$$\begin{aligned} \min_{\boldsymbol{\beta}^v, \mathbf{A}} \quad & \left\| \sum_{t=0}^{\hat{t}} \beta_t^v T_t(\mathbf{S}^{(v)}) - \mathbf{A} \right\|_2^2 \\ \text{s.t.} \quad & 0 \le a_{ij} \le 1, \sum_{j=1}^{n} a_{ij} = 1. \end{aligned} \tag{4}$$

Here, β_t^v is the coefficient associated with each Chebyshev polynomial, and a_{ij} is the weight associated with the elements of matrix $\mathbf{A}$, with $0 \le a_{ij} \le 1$ for all i and j.

Subsequently, we aggregate the weighted errors across all views and orders of Chebyshev polynomials by solving the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{A}} \quad & \sum_{v=1}^{m} \alpha_v^2 \left\| \sum_{t=0}^{\hat{t}} \beta_t^v T_t(\mathbf{S}^{(v)}) - \mathbf{A} \right\|_2^2 \\ \text{s.t.} \quad & 0 \le \alpha_v \le 1, \sum_{v=1}^{m} \alpha_v = 1, \\ & 0 \le a_{ij} \le 1, \sum_{j=1}^{n} a_{ij} = 1, \end{aligned} \tag{5}$$

where α_v represents the weight assigned to the v-th view.

To enforce the target matrix $\mathbf{A}$ to have the desired number of connected components, we introduce the constraint on the rank of the Laplacian matrix $\mathbf{L_A}$ as $\text{rank}(\mathbf{L_A}) = n - c$ based on Ky Fan's theorem. Furthermore, we introduce the sum of the top c smallest eigenvalues of $\mathbf{L_A}$ to control its rank.

The Higher Order Multiple Graph Filtering for Structured Graph Learning and Clustering (HMGC) method can be summarized as follows:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{A}, \mathbf{F}} \quad & \sum_{v=1}^{m} \alpha_v^2 \left\| \sum_{t=0}^{\hat{t}} \beta_t^v T_t(\mathbf{S}^{(v)}) - \mathbf{A} \right\|_2^2 + \lambda \text{tr}(\mathbf{F}^T \mathbf{L_A} \mathbf{F}) \\ \text{s.t.} \quad & 0 \le \alpha_v \le 1, \sum_{v=1}^{m} \alpha_v = 1, \\ & 0 \le a_{ij} \le 1, \sum_{j=1}^{n} a_{ij} = 1, \mathbf{F}^T \mathbf{F} = \mathbf{I}_c. \end{aligned} \tag{6}$$

$\mathbf{F} \in \mathbf{R}^{n \times c}$ consists of c eigenvectors corresponding to c eigenvalues of $\mathbf{L_A}$. In HMGC, we use Chebyshev polynomials to minimize the weighted sum of squared differences between linear combinations of these polynomials applied to multiple affinity matrices $\mathbf{S}^{(v)}$ and a consensus matrix $\mathbf{A}$. We achieve this through two optimization steps, optimizing Chebyshev coefficients and $\mathbf{A}$ for each view and then aggregating errors across all views and polynomial orders. HMGC provides a powerful framework for structured graph learning and clustering, allowing control over the number of connected components in the final consensus matrix $\mathbf{A}$. In summary, HMGC enhances graph-based data analysis by combining higher-order Chebyshev polynomials and multi-view affinity matrices.

3. OPTIMIZATION PROCESS

In this paper we adopt the block coordinate descend method to update a single variable while keeping other variables fixed.

Update $\boldsymbol{\alpha}$. Fixing $\boldsymbol{\beta}$, $\mathbf{F}$,and $\mathbf{A}$, the rest problem w.r.t. $\boldsymbol{\alpha}$ reduces to:

$$\min_{\boldsymbol{\alpha}} \sum_{v=1}^{m} \alpha_v^2 e_v \quad \text{s.t.} \quad 0 \le \alpha_v \le 1, \sum_{v=1}^{m} \alpha_v = 1, \tag{7}$$

where $e_v = || \sum_{t=0}^{\hat{t}} \beta_t^v T_t(\mathbf{S}^{(v)}) - \mathbf{A}||_2^2$. We can obtain its closed-form solution as follows:

$$\alpha_v = \frac{e_v^{-1}}{\sum_{v'=1}^{m} e_{v'}^{-1}}. \tag{8}$$

Update $\boldsymbol{\beta}$. Fixing $\boldsymbol{\alpha}$,$\mathbf{F}$, and $\mathbf{A}$, the rest problem w.r.t. $\boldsymbol{\beta} \in \mathbb{R}^{\hat{t} \times m}$ is column decoupled. The sub-problem of the v-th column w.r.t. $\boldsymbol{\beta}^v$ can be written as:

$$\min_{\boldsymbol{\beta}^v} \quad \boldsymbol{\beta}^{vT} \mathbf{H}^{(v)} \boldsymbol{\beta}^v + \boldsymbol{\beta}^{vT} \mathbf{f}^v, \tag{9}$$

where $\mathbf{H}^{(v)} \in \mathbb{R}^{\hat{t} \times \hat{t}}$ with $\mathbf{H}_{ij}^{(v)} = \text{tr}(T_i(\mathbf{S}^{(v)})^T T_j(\mathbf{S}^{(v)}))$, and $\mathbf{f}^v \in \mathbb{R}^{\hat{t} \times 1}$ with $f_t^v = -2\text{tr}(T_t(\mathbf{S}^{(v)})^T \mathbf{A})$. Eq.(9) can be easily solved by off-the-shelf solvers.

Update $\mathbf{A}$. Fixing $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{F}$, the rest problem w.r.t. $\mathbf{A}$ reduces to:

$$\min_{\mathbf{A}} \quad ||\mathbf{A}-\mathbf{B}||_2^2 \quad \text{s.t.} \quad 0 \leq a_{ij} \leq 1, \sum_{j=1}^{n} a_{ij} = 1, \quad (10)$$

where $\mathbf{B} = \frac{\sum_{v=1}^{m} \alpha_v^2 \sum_{t=0}^{\hat{t}} \beta_t^v T_t(\mathbf{S}^{(v)}) - \frac{\lambda}{4}\mathbf{D_F}}{\sum_{v=1}^{m} \alpha_v^2}$ and $\mathbf{D_F} \in \mathbb{R}^{n \times n}$ denotes the pairwise distance on $\mathbf{F}$. Eq.(10) is row decoupled and each sub-problem can be solved by the Euclidean projection on the simplex [6].

Update $\mathbf{F}$. Fixing $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{A}$, the rest problem w.r.t. $\mathbf{F}$ reduces to:

$$\min_{\mathbf{F}} \quad tr(\mathbf{F}^T \mathbf{L_A} \mathbf{F}) \quad \text{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_c. \quad (11)$$

The optimal solution for $\mathbf{F}$ corresponds to the c smallest eigenvalues of $\mathbf{L_A}$ and their corresponding eigenvectors.

After the convergence of the iterative algorithm, the final clustering results can be obtained from $\mathbf{A}$. We omit the convergence and complexity analysis due to space limitation.

4. EXPERIMENTS

4.1. Benchmark Datasets

We conducted experiments on 10 benchmark datasets, primarily covering two categories: documents and images. In the document category, we included the news dataset 3SOURCES[1] and TEXAS[2], which encompasses web pages and hyperlinks. For the image category, it includes 100LEAVES[3] (Comprising 100 types of plant images), NUSWIDE [8] (Network images), MNIST [9] (Handwritten images), COIL[4] (Color images), PBMC[5] (Cell images), SCENE[6] , and CORNELL[7] (Object images), as well as SUNRGBD [10] (RGB images related to scene understanding).The detailed information of these datasets is shown in Table 1.

4.2. Experiment Setup

In our experiments, we evaluate the performance of various clustering algorithms on a diverse set of benchmark datasets. These datasets, summarized in Table 1, encompass a wide range of characteristics, including the number of samples, features, and classes.

[1] http://mlg.ucd.ie/datasets/3sources.html

[2] https://starling.utdallas.edu/datasets/webkb

[3] https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set

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

[5] https://satijalab.org/seurat/vignettes.html

[6] https://figshare.com/articles/dataset/15-SceneImageDataset/7007177

[7] http://www.cs.cornell.edu/\textasciitildecristian/Cornell_Movie-Dialogs_Corpus.html

Table 1. Summary of the Data Sets

ID	Data Name	Samples	Features	Classes
D1	3SOURCES	169	3560,3631,3068	6
D2	TEXAS	187	1703,187	5
D3	CORNELL	195	1703,195	5
D4	COIL	1440	1024,944,4096,576	20
D5	100LEAVES	1600	64,64,64	100
D6	NUSWIDE	2000	64,225,144,73,128	31
D7	MNIST4	4000	30,9,30	4
D8	PBMC	4271	100,200,400,100,200,400,64,128,256	8
D9	SCENE	4485	20,59,40	15
D10	SUNRGBD	10335	4096,4096	45

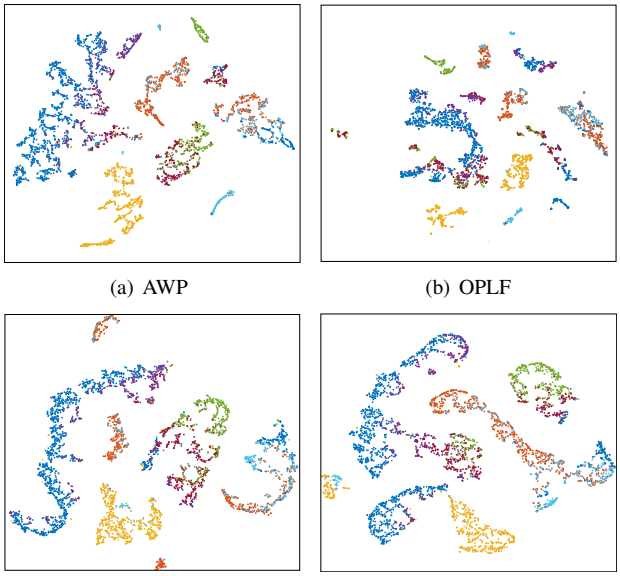
To demonstrate the effectiveness of our proposed method, we conducted comparative experiments with nine state-of-the-art multi-view clustering approaches. These approaches include AWP [11], MCGC [12], GMC [13], LMVSC [14], 2CMV [15], OPLF [16], AONGR [17], EOMSC [18] and LSRMSC [19]. Across all datasets, we set the number of clusters equal to the number of classes. For our method, we simply fix the neighborhood size regularization $k = 10$ and conduct grid search on the parameter $\lambda \in [10^{-3}, \ldots, 10^{3}]$ and the order of ChebyshevPolynomials $\hat{t} \in [3, \ldots, 10]$. The clustering performance is assessed through the utilization of two commonly employed metrics: Accuracy (ACC) and Normalized Mutual Information (NMI). Due to space limitations, we only report the clustering accuracy in Table 2 and omit the results for other metrics and time consumption.

4.3. Experimental Results Analysis

The impractical time complexity of LSRMSC on the D10 restricts its feasible execution within a reasonable timeframe. In contrast, our proposed HMGC exhibits superior performance across all datasets. This is evident from consistently higher accuracy scores, as shown in Table 2. Notably, on challenging datasets (e.g., D4 and D5), HMGC achieves significantly better accuracy compared to its counterparts. Furthermore, as shown in Fig. 2, we present the t-SNE [20] results of AWP, OPLF, LMVSC, and our HGMC on D8. We can see that our method achieves better embedding results because it can more clearly partition the data into different categories.This underscores the robustness and effectiveness of HMGC in achieving accurate and reliable clustering results.

4.4. Parameter Sensitivity

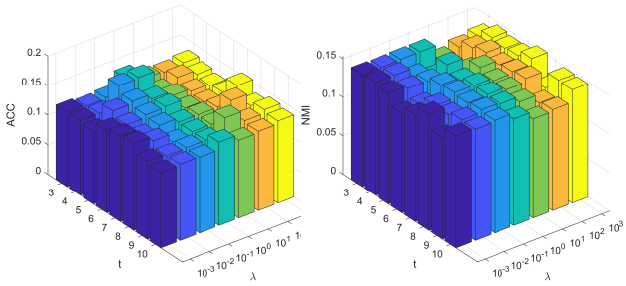
The clustering efficiency and effectiveness of HMGC are influenced by two parameters: λ and the order of Chebyshev polynomials. In this paper, we illustrate the impact of different values of these two parameters on HMGC using D6 and D8 as examples in Fig. 3 and Fig. 4.



(a) AWP (b) OPLF

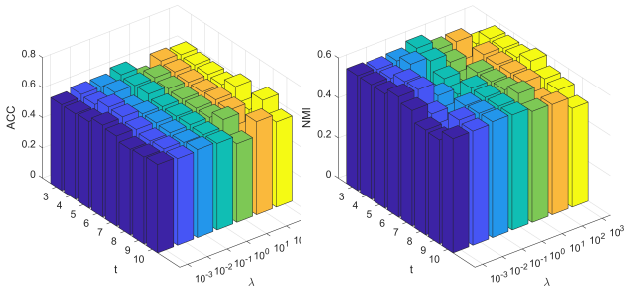
(c) LMVSC (d) HMGC

Fig. 2. We respectively present the t-SNE results of embeddings using AWP, OPLF, LMVSC, and our HGMC on D8.



(a) Clustering ACC on D6 (b) Clustering NMI on D6

Fig. 3. Clustering ACC and NMI of HMGC with different parameters on the D6.



(a) Clustering ACC on D8 (b) Clustering NMI on D8

Fig. 4. Clustering ACC and NMI of HMGC with different parameters on the D8.

The results on other datasets are similar. Fig. 3 and Fig. 4 indicates that the proposed HMGC can achieve relatively good results when $\lambda \in [10^{-3}, \ldots, 1]$ and $t \in [3, \ldots, 6]$.

Table 2. Accuracy results on all datasets. Bold indicates the best results, and underline indicates the second-best results.

Methods	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
AWP	62.72	50.27	42.05	69.86	78.88	15.15	60.65	63.59	36.63	17.23
MCGC	74.56	55.61	52.82	80.69	72.75	14.90	80.25	55.07	14.74	18.23
GMC	69.23	52.94	52.82	80.35	82.38	14.90	92.03	62.61	14.00	12.77
LMVSC	54.44	55.61	50.77	65.83	60.63	13.70	64.18	57.01	35.88	18.49
OPLF	60.80	55.61	50.77	54.37	75.19	14.05	87.37	54.88	37.53	8.92
LSRMSC	63.31	54.55	48.72	59.72	3.81	14.00	50.58	46.66	22.92	-
2CMV	34.32	54.10	49.66	67.50	70.86	12.45	72.51	53.45	33.36	18.65
EOMSC	33.14	56.68	44.62	60.97	42.06	16.02	88.65	67.82	32.95	**23.70**
AONGR	52.07	58.82	47.69	69.10	88.19	15.50	64.93	56.52	40.69	19.98
HMGC	**76.92**	**63.64**	**56.05**	**95.12**	**93.16**	**16.13**	**92.35**	**68.51**	**41.53**	20.29

4.5. Ablation Study

We introduce a first-order graph clustering method denoted as MGC, which utilizes only $\{T_1(\mathbf{S}^{(1)}), T_1(\mathbf{S}^{(2)}), \ldots, T_1(\mathbf{S}^{(m)})\}$ as defined in Eq. (6). The results of first-order MGC and higher-order HMGC are presented in Table 3. These results clearly demonstrate that leveraging higher-order graphs with optimized coefficient learning significantly enhances clustering performance.

Table 3. Ablation Study on First and Higher order MGC

Dataset	ACC		NMI	
	MGC	HMGC	MGC	HMGC
D1	72.78	**76.42**	63.48	**73.48**
D2	61.50	**63.64**	31.14	**35.70**
D3	53.33	**56.05**	36.83	**41.00**
D4	87.36	**95.12**	93.90	**97.48**
D5	87.59	**93.16**	94.04	**96.88**
D6	14.38	**16.13**	14.87	**15.94**
D7	92.23	**92.35**	81.02	**81.42**
D8	64.84	**68.51**	56.56	**60.44**
D9	39.98	**41.53**	39.12	**42.07**
D10	17.54	**20.29**	18.78	**23.78**

5. CONCLUSION

HMGC introduces a promising direction in multi-graph clustering, showcasing its remarkable capability to unveil concealed structures within intricate datasets. By harnessing Chebyshev polynomials constructed from adjacency matrices, our approach effectively captures higher-order, nonlinear relationships present in the data. This innovative strategy leads to a substantial enhancement in clustering performance. In the future, we plan to investigate the spectral property of the Chebyshev polynomials for the task of MGC.

6. REFERENCES

[1] Hao Wang, Yan Yang, Bing Liu, and Hamido Fujita, "A study of graph-based system for multi-view clustering," *Knowledge-Based Systems*, vol. 163, pp. 1009–1019, 2019.

[2] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang, "Multi-view clustering in latent embedding space," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 3513–3520.

[3] 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*, vol. 45, no. 1, pp. 1055–1069, 2022.

[4] Steffen Bickel and Tobias Scheffer, "Multi-view clustering.," in *ICDM*. Citeseer, 2004, vol. 4, pp. 19–26.

[5] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao, "Late fusion incomplete multi-view clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 10, pp. 2410–2423, 2018.

[6] Feiping Nie, Xiaoqian Wang, and Heng Huang, "Clustering and projected clustering with adaptive neighbors," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 977–986.

[7] Durgesh Singh, Ahcène Boubekki, Robert Jenssen, and Michael C Kampffmeyer, "Supercm: Revisiting clustering for semi-supervised learning," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[8] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng, "Nus-wide: a real-world web image database from national university of singapore," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, Jul 2009.

[9] Li Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE signal processing magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[10] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, "Learning deep features for scene recognition using places database," *Neural Information Processing Systems,Neural Information Processing Systems*, Dec 2014.

[11] Feiping Nie, Lai Tian, and Xuelong Li, "Multiview clustering via adaptively weighted procrustes," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 2022–2030.

[12] Kun Zhan, Feiping Nie, Jing Wang, and Yi Yang, "Multiview consensus graph clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1261–1270, 2018.

[13] Hao Wang, Yan Yang, and Bing Liu, "Gmc: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1116–1129, 2019.

[14] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu, "Large-scale multi-view subspace clustering in linear time," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 4412–4419.

[15] Khanh Luong and Richi Nayak, "A novel approach to learning consensus and complementary information for multi-view data clustering," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 2020, pp. 865–876.

[16] 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*, 2021, pp. 6850–6859.

[17] Mingyu Zhao, Weidong Yang, and Feiping Nie, "Auto-weighted orthogonal and nonnegative graph reconstruction for multi-view clustering," *Information Sciences*, vol. 632, pp. 324–339, 2023.

[18] 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*, 2022, vol. 36, pp. 7576–7584.

[19] Shudong Huang, Yixi Liu, Yazhou Ren, Ivor W Tsang, Zenglin Xu, and Jiancheng Lv, "Learning smooth representation for multi-view subspace clustering," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 3421–3429.

[20] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.