Self-Supervised Graph Convolutional Network for Multi-View Clustering

Wei Xia, *Graduate Student Member, IEEE*, Qianqian Wang, Quanxue Gao, Xiangdong Zhang, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*—Despite the promising preliminary results, existing graph convolutional network (GCN) based multi-view learning methods directly use the graph structure as view descriptor, which may inhibit the ability of multi-view learning for multimedia data. The major reason is that, in real multimedia applications, the graph structure may contain outliers. Moreover, they fail to take advantage of the information embedded in the inaccurate clustering labels obtained from their proposed methods, resulting in inferior clustering results. These observations motivate us to study whether there is a better alternative GCN based framework for multi-view clustering. To this end, in this paper, we propose an end-to-end self-supervised graph convolutional network for multi-view clustering (SGCMC). Specifically, SGCMC constructs a new view descriptor for graph-structured data by mapping the raw node content into the complex space via Euler transformation, which not only suppresses outliers but also reveals non-linear patterns embedded in data. Meanwhile, the proposed SGCMC uses the clustering labels to guide the learning of the latent representation and coefficient matrix, and the latter in turn is used to conduct the subsequent node clustering. By this way, clustering and representation learning are seamlessly connected, with the aim to achieve better clustering results. Extensive experimental results indicate that the proposed SGCMC outperforms the state-of-the-art methods.**

***Index Terms*—Node clustering, graph representation learning, multi-view learning, subspace clustering, self-supervision.**

I. Introduction

With the development of social networking, multimedia data has become an important data resource in the domain of artificial intelligence [1]–[4]. Graph-structured data plays a significant role in multimedia social network data analysis. Owing to its superiority in processing graph-structured data, Graph Convolutional Network (GCN) [5] has been widely applied in various data mining tasks, such as action recognition [6], pose estimation [7], spammer detection [8], text classification [9], and node clustering [10]. We herein center on node clustering that is one of the most representative GCN based applications.

Clustering targets at dividing a group of unlabeled data into several disjoint clusters, such that the data in the same cluster have high correlation to each other [11], [12]. One of the most representative node clustering methods is graph auto-encoder (GAE) [13], which well characterizes the graph-structured data via encoding both the graph structure and node content, and finally obtains an interpretable latent representation for node clustering. As a variant of GAE, Pan *et al.* [14] proposed the adversarial regularized graph auto-encoder (ARGAE) by joint the adversarial learning. To learn a more robust node representation, Veličković *et al.* [15] took into account the importance of the neighbor nodes of the target node, and proposed the graph attention network (GAT). Similarly, Wang *et al.* [16] proposed the deep attentional embedded graph clustering approach (DAEGC). Despite the good performance, GAE, ARGAE, and DAEGC only reconstruct the graph structure via the inner product decoder. Therefore, the decoder cannot be learnable, resulting in degrading the capability of graph embedding. To this end, Salehi *et al.* [17] proposed graph attention auto-encoder (GATE) to simultaneously reconstruct the graph structure and node content, which makes the latent representation well preserve the graph structure as well as content information of nodes.

Numerous studies have shown that multi-view descriptors can provide complementary information embedded in multiple views [18]–[20], which is helpful for clustering. Motivated by this, Li *et al.* [21] proposed the co-training GCN (Co-GCN) with semi-supervised setting. When handling graph-structured data, Co-GCN first treats node content and graph structure as different view descriptors and constructs the nearest-neighbor graph for each view separately. Then, it trains a graph encoder for each view to obtain the common representation for the downstream task via aggregating the latent representation form each view.

Although the Co-GCN provides a new solution for multi-view learning and achieves impressive results, it still has the following shortcomings:

Manuscript received November 27, 2020; revised April 9, 2021 and June 4, 2021; accepted June 28, 2021. Date of publication July 2, 2021; date of current version July 12, 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 61773302, 62036007, and 62050175, in part by Natural Science Basic Research Plan in Shaanxi Province under Grant 2020JZ-19, in part by the Fundamental Research Funds for the Central Universities, the Innovation Fund of Xidian University, in part by the Natural Science Foundation of Ningbo under Grant 2018A610049. The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. M. Shamim Hossain. *(Corresponding author: Quanxue Gao.)*

Wei Xia, Qianqian Wang, and Xiangdong Zhang are with the State Key Laboratory of Integrated Services Networks, Xidian University, Shaanxi 710071, China (e-mail: xdweixia@gmail.com; qianqian174@foxmail.com; 578653865@qq.com).

Quanxue Gao is with the State Key Laboratory of ISN, Xidian University, Shaanxi 710071, China, and also with the Xidian-Ningbo Information Technology Institute, Ningbo 315000, China (e-mail: qxgao@xidian.edu.cn).

Xinbo Gao is with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xbgao@mail.xidian.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMM.2021.3094296.

Digital Object Identifier 10.1109/TMM.2021.3094296

1520-9210 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

1) When dealing with graph-structured data, it directly utilizes the graph structure as view descriptor, which may inhibit the ability of multi-view learning. In real applications, the outliers may be embedded in the graph structure. In general, different views can be represented by different descriptors of samples in multi-view learning [22]–[25], *e.g.*, Gabor feature, scale-invariant feature transform (SIFT) feature [26], local binary pattern (LBP) feature [27], GIST feature [28]. However, these descriptors are suitable for data with Euclidean structure, *e.g.*, face image, handwritten digit, object, and scene. For data with non-Euclidean structure, *e.g.*, graph structured data form citation networks, how to effectively construct view descriptors is the key to improving the clustering performance.
2) It neglects the useful information embedded in the inaccurate clustering labels. Although the clustering labels of some nodes are inaccurate, the partial accurate label is useful. However, to the best of our knowledge, similar investigations for multi-view graph-structured data clustering have been found lacking so far, which is one of the motivations behind this work.

Inspired by above insight analysis and the fact that kernel trick can capture the nonlinear features [29], [30], we propose a novel multi-view subspace clustering method, named self-supervised graph convolutional network for multi-view clustering (SGCMC). SGCMC constructs view descriptor for the graph-structured data by mapping the raw node content into a complex space via Euler transformation, which not only suppresses outliers but also reveals non-linear patterns embedded in data. Afterwards, the proposed SGCMC consists of two steps. The first step aims to learn the latent representation of each view and the coefficient matrix shared by different views, which is conducted to map the inputs of different views into a latent space in the forward pathway of SGCMC. The second step implements nodes clustering and uses the inaccurate clustering labels to guide the learning of the latent representation and the coefficient matrix. With such a strategy, even no ground-truth is provided, SGCMC can still be trained in an end-to-end pipeline. Meanwhile, as shown in our experiments, such a manner will lead to a better coefficient matrix and the superior clustering performance. In short summary, the major contributions of this paper are as follows:

1) By utilizing the inaccurate clustering label, we present a novel multi-view self-supervised clustering framework. To the best of our knowledge, this could be the first multi-view self-supervised graph convolutional clustering network. Hence, we assume that this paper could provide a novel insight toward multi-view GCN based unsupervised learning.
2) Our method uses the features extracted by Euler transform as a new view descriptor, different from the general method which employs graph structure as view descriptor, Euler transform maps the raw node content into an explicit space which has the same dimension as the raw node content. As a result, our method can be easily implemented in real applications. This helps to further study GCN based multi-view learning.
3) Experimental results over the four benchmark datasets indicate that SGCMC outperforms the state-of-the-art methods.

Notations: For convenience, we first introduce the notations used throughout the paper. We use bold upper case letters for matrices, *e.g.,* $\mathbf{Z}$, bold lower case letters for vectors, *e.g.,* $\mathbf{z}$, and upper case letters such as $\mathbf{Z}_{ij}$ for the entries of $\mathbf{Z}$. The Frobenius norm of $\mathbf{Z} \in \mathbb{R}^{N\times d}$ is $\|\mathbf{Z}\|_F = \sqrt{\sum_{p=1}^{N}\sum_{q=1}^{d}\mathbf{Z}_{pq}^2}$.

II. Related Works

As our proposed SGCMC is related to graph embedding learning and multi-view clustering. Hence, we review the literature on graph embedding learning and multi-view clustering.

A. Graph Embedding Learning

Graph embedding aims at learning a low-dimensional node representation while preserving both the content information and topology structure of the node. The past decade saw an upsurge of graph embedding methods which can be roughly grouped into two main categories according to the input information, *i.e.*, topological structure embedding (TSE) methods and content enhanced graph embedding (CEGE) methods. A comprehensive survey of recent graph representation learning methods can be found in [31], [32].

TSE methods only take the topological structure as input, and map it to learn low-dimensional node representation. For example, Perozzi *et al.* [33] presented the truncated random walk algorithm (DeepWalk) to learn node representation, in which the raw graph structure information was transformed into collections of linear sequences. Different from generating linear sequences [33], Cao *et al.* [34] presented deep neural networks for learning graph representations (DNGR), in which a random surfing model was proposed to exploit the topological structural information directly. Instead of embedding each individual node, Cavallari *et al.* [35] integrated community embedding, community detection and node embedding into a closed-loop, and proposed community embedding framework (ComE), which is beneficial to community-level applications, *e.g.*, graph visualization. To handle the unknown number of communities issue, Cavallari *et al.* [36] proposed to learn both finite and infinite communities embedding on graphs (ComE+). Although the aforementioned methods achieve impressive results, they only take graph structure into account, which limits their performances. To explore the graph structure with additional node content information, CEGE methods become a hot topic in the graph embedding learning.

CEGE methods simultaneously encode the graph structure and node content into a common space to obtain the node representation. For example, Yang *et al.* [37] extended DeepWalk model and proposed Text-Associated DeepWalk (TADW) to exploit node content features. Aiming at learning high-level representation from input graph structure and node content, GCN

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

has obtained impressive performances in scenarios of supervised and semi-supervised learning [6], [21], [38] due to the rapidly developing computational resources, *e.g.*, graphics processing units (GPUs) [39]. In contrast, less attention has been paid to unsupervised node clustering tasks [40]–[42]. Recently, some works [10], [13]–[17] have devoted to GCN based node clustering, and shown promising clustering results. Studies have shown multi-view data [43] is helpful to boost the clustering performance. However, all the aforementioned methods only exploit single-view graph structure and node content, resulting in inferior results.

Unlike the aforementioned graph embedding approaches, the proposed SGCMC leverages the Euler representation to construct a new node content descriptor and then learns a set of multi-view graph auto-encoder to map the input node content and graph structure into another space. Finally, SGCMC uses the coefficient matrix of nodes in the new latent space to calculate the affinity matrix for clustering.

B. Multi-View Clustering

In recent years, a large number of multi-view clustering methods have been devoted to learning a high quality latent representation or affinity matrix shared by different views [19], [20], [22]–[25], [44], [45], among which deep multi-view clustering approaches are widely concerned [41], [46]–[48] by researchers due to the outstanding representative capacity and fast inference speed of deep learning. For example, Andrew *et al.* [46] proposed a new multi-view clustering algorithm via combining deep encoder with Canonical Correlation Analysis (DCCA). To learn better multi-view representation, Wang *et al.* [47] extended DCCA by introducing deep decoder, and proposed deep canonically correlated auto-encoders (DCCAE). To better exploit the graph structure information from the multiple views, Li *et al.* [21] proposed GCN based multi-view learning approach, namely Co-GCN. However, Co-GCN is designed for semi-supervised clustering. To tackle this problem, Fan *et al.* [41] proposed the one to multi-graph auto-encoder for graph embedding clustering (O2MAC). Despite the success of O2MAC, it only encodes single-view node content information. When handling data with the single-view graph and node content, the performance is limited.

To further boost the performance of graph-structured data clustering, we herein study an important yet largely underexplored problem, *i.e.*, multi-view graph embedding clustering by incorporating GCN and self-supervision strategy.

III. The Proposed SGCMC

In this section, we first introduce how to effectively construct a new multi-view descriptor for graph-structured data, then introduce how the proposed SGCMC achieves nodes clustering in an end-to-end manner. After that, we will give the implementation details of SGCMC.

A. View Descriptor Construction

Let $\{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(V)}\}$ and $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)}\}$ be the multi-view dataset, where $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times N}$ denotes the node content matrix of the v-th view ($v = 1, \ldots, V$) and $\mathbf{A}^{(v)} \in \mathbb{R}^{N \times N}$ is the corresponding graph structure matrix. d_v and N denote the node content dimension and the number of nodes in the v-th view, respectively. V is the number of views. For brevity, we take two views as an example. Existing node clustering methods only have the raw node descriptor $\mathbf{X}^{(1)}$. For multi-view setting, we leverage Euler transform [29] to extract nonlinear features as a new view descriptior $\mathbf{X}^{(2)}$. More specifically, we map an arbitrary vector $\mathbf{x}_p \in \mathbb{R}^{d_v}$ onto the complex representation $\mathbf{z}_p \in \mathbb{C}^{d_v}$, where

$$\mathbf{z}_p = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\alpha\pi\mathbf{x}_{p_1}} \\ \vdots \\ e^{i\alpha\pi\mathbf{x}_{p_{d_v}}} \end{bmatrix} = \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{x}_p}, \tag{1}$$

where i is the unit imaginary number, $\alpha \in \mathbb{R}^+$ is the frequency value and is adjusted to suppress the values caused by outliers. $\mathbf{x}_p \in \mathbf{X}^{(1)}$ is the raw descriptor of p-th node, $\mathbf{z}_p \in \mathbf{Z}$ is the Euler representation of $\mathbf{x}_p$. In this paper, we have $\mathbf{X}^{(2)} = \mathbf{Z}$.

B. Subspace Node Clustering Module

Subspace clustering aims to learn a common coefficient representation matrix that is shared by different views, then we assign each node into one of K clusters in this new subspace. With regards to this, SGCMC gets the node clustering results with two joint modules. One is a subspace clustering module with a graph attention auto-encoder and the other is a self-supervised learning module by simultaneously supervising the latent representation and coefficient representation. Fig. 1 gives the overall architecture of our proposed SGCMC. In this paper, assuming that each view has the same graph structure $\mathbf{A}$, *i.e.*, $\mathbf{A} = \mathbf{A}^{(1)} = \mathbf{A}^{(2)}$. $\mathbf{F}^{(v)} \in \mathbb{R}^{d_{l2} \times N}$ is the corresponding latent representation learned by the graph attention encoder, where d_{l2} is the dimension of latent representation. $\widehat{\mathbf{A}}^{(v)}$ and $\widehat{\mathbf{X}}^{(v)}$ are the reconstructed graph structure and node content, respectively.

Specifically, SGCMC progressively maps the raw sample $\{\mathbf{X}^{(v)}, \mathbf{A}^{(v)}\}$ into the latent representation $\mathbf{F}^{(v)}$ via a series of nonlinear transformations. Here, the transformations are modeled by GATE [17]. In order to relieve the heterogeneous gap between different $\mathbf{F}^{(v)}$ and better align latent representation, we build a multi-view shared auto-encoder in the proposed SGCMC. The multi-view shared auto-encoder consists of a four-layer graph attention auto-encoder, *i.e.,* the two-layer encoder $\mathcal{E}[\cdot]$ and the two-layer decoder $\mathcal{D}[\cdot]$. We also utilize the inner product decoder to reconstruct graph structure $\mathbf{A}^{(v)}$ of each view. To ease of presentation, the latent representation of the v-th view can be represented by

$$\mathbf{F}^{(v)} = \mathcal{E}\left[(\mathbf{X}^{(v)}, \mathbf{A}^{(v)})|\Theta_{\mathcal{E}}\right], \tag{2}$$

where $\Theta_{\mathcal{E}}$ denotes the trainable parameters of the multi-view shared graph attention encoder.

To enforce the representation $\mathbf{F}^{(v)}$ more suitable for clustering than the raw data, SGCMC herein employs the good property of self-expressive learning to obtain a view-consensus coefficient representation. In more detail, to obtain a good coefficient matrix shared by different views, we employ the self-expressive operation on the latent representation $\mathbf{F}^{(v)}$ of the v-th view, which

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

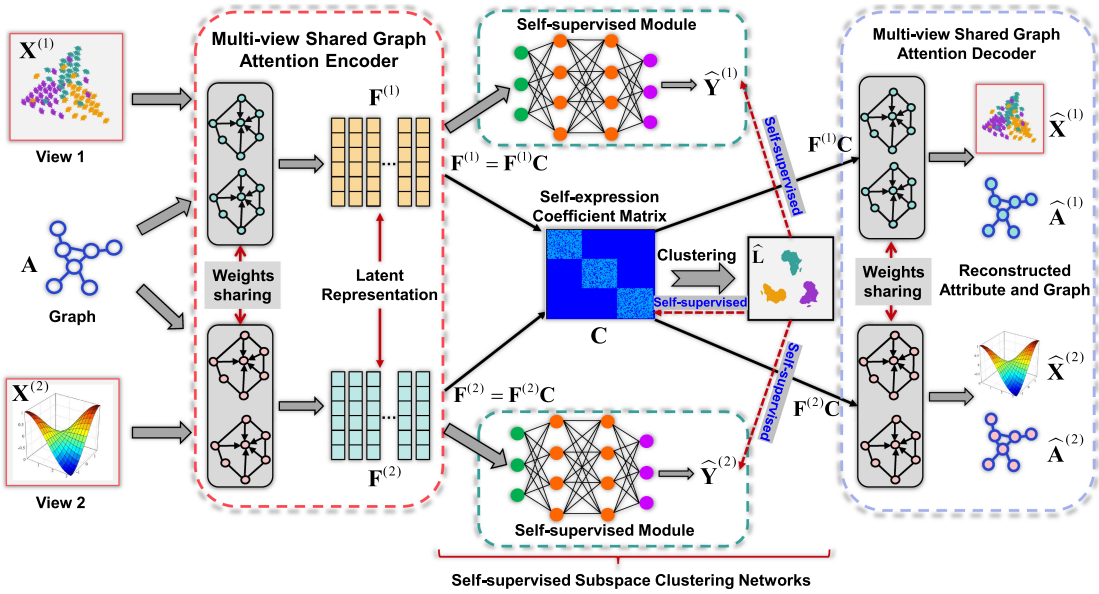


Fig. 1. The overall framework of the proposed SGCMC.

can be defined as

$$\mathbf{F}^{(v)} = \mathbf{F}^{(v)}\mathbf{C}, \quad \text{s.t.} \quad \mathbf{C}_{pp} = 0, p \in [1,\ N], \tag{3}$$

where $\mathbf{C} \in \mathbb{R}^{N\times N}$ is the view-consensus coefficient representation. Also, to prevent the trivial solution $\mathbf{C} = \mathbf{I}$, for the p-th node, we constrain $\mathbf{C}_{pp} = 0$. Thus, SGCMC minimizes the following objective function $\mathcal{L}_{\text{Sub}}$:

$$\mathcal{L}_{\text{Sub}} = \min_{\mathbf{C}} \lambda_1 \sum_{v=1}^{V} \left\| \mathbf{F}^{(v)}\mathbf{C} - \mathbf{F}^{(v)} \right\|_F^2 + \mu \left\| \mathbf{C} \right\|_p$$
$$\text{s.t.} \quad \mathbf{C}_{pp} = 0, p \in [1,\ N], \tag{4}$$

where λ_1 and μ are the two trade-off parameters. Here, $\|\mathbf{C}\|_p$ denotes the penalty term, *e.g.,* the sparsity penalty term $\|\mathbf{C}\|_1$, the nuclear norm penalty term$\|\mathbf{C}\|_*$, and the F-norm $\|\mathbf{C}\|_F$. In this paper, we set this penalty to $\|\mathbf{C}\|_1$. To make sure that the SGCMC can learn a consistent subspace $\mathbf{C}$ among different views, we herein employ a consistent representation constraint $\mathcal{L}_{\text{Con}}$ to capture the geometric relationship similarity embedded in different views. Thus, we have

$$\mathcal{L}_{\text{Con}} = \min_{\mathbf{F}^{(v)}} \sum_{v\neq m}^{V} \left\| \mathbf{F}^{(v)} - \mathbf{F}^{(m)} \right\|_F^2. \tag{5}$$

When we obtain the coefficient representation $\mathbf{C}$, the induced affinity matrix Δ can by calculated by $\Delta = \frac{1}{2}(|\mathbf{C}| + |\mathbf{C}^{\mathrm{T}}|)$. Finally, we can obtain the node clustering results, *i.e.,* the pseudo clustering label $\widehat{\mathbf{L}}$ by applying a spectral clustering algorithm on the affinity matrix Δ. In this paper, we employ normalized cut (NCut) [49] algorithm to get the nodes clustering labels.

Meanwhile, to make sure that the latent representations $\mathbf{F}^{(v)}$ preserve sufficient node content information and graph structure information, the new representation $\mathbf{F}^{(v)}\mathbf{C}$ of v-th view is subsequently fed into both the graph attention decoder (to reconstruct the original node content $\mathbf{X}^{(v)}$) and inner product decoder (to reconstruct the original graph structure $\mathbf{A}^{(v)}$). Hence, the graph attention decoder has a symmetrical structure to the encoder. To be exact, we train the graph auto-encoder with an attention mechanism by optimizing the node content reconstruction loss $\mathcal{L}_{\text{AR}}$ and graph structure reconstruction loss $\mathcal{L}_{\text{GR}}$. The definitions of these two objective functions are as follows:

$$\mathcal{L}_{\text{AR}} = \min_{\Theta_{\mathcal{E}},\Theta_{\mathcal{D}}} \sum_{v=1}^{V} \left\| \widehat{\mathbf{X}}^{(v)} - \mathbf{X}^{(v)} \right\|_F^2, \tag{6}$$

$$\mathcal{L}_{\text{GR}} = \min_{\Theta_{\mathcal{E}}} \lambda_2 \sum_{v=1}^{V} \mathbb{E}[\log(\Phi(\mathbf{F}^{(v)}\mathbf{F}^{(v)^{\mathrm{T}}}))]$$
$$= \min_{\Theta_{\mathcal{E}}} \lambda_2 \sum_{v=1}^{V} \mathbb{E}[\log(\widehat{\mathbf{A}}^{(v)})], \tag{7}$$

where λ_2 is a trade-off parameter, $\Theta_{\mathcal{D}}$ is the trainable parameters of the multi-view shared graph attention decoder. $\Phi(\cdot)$ is the Sigmoid activation function.

C. Self-Supervised Learning Module

To supervise the learning of the latent representation $\mathbf{F}^{(v)}$, we introduce the following cross-entropy based objective function:

$$\mathcal{L}_{\text{CE}} = \min_{\mathbf{F}^{(v)},\Theta_{\mathcal{H}}^{(v)}} \sum_{v=1}^{V} \text{cross\_entropy}\left(\widehat{\mathbf{L}}, \widehat{\mathbf{Y}}^{(v)}\right), \tag{8}$$

where $\widehat{\mathbf{Y}}^{(v)} \in \mathbb{R}^{K\times N}$ is the predicted label matrix obtained by $\mathbf{F}^{(v)}$, *i.e.,* $\widehat{\mathbf{Y}}^{(v)} = \mathcal{H}[\mathbf{F}^{(v)}|\Theta_{\mathcal{H}}^{(v)}]$. In this paper, we introduce a three-layer fully connected network (FCN), *i.e.,* a classifier $\mathcal{H}[\cdot]$, to supervise the learning of representation $\mathbf{F}^{(v)}$, where $\Theta_{\mathcal{H}}^{(v)}$ is the trainable parameters of classifier. $\widehat{\mathbf{L}} \in \mathbb{R}^{K\times N}$ is the one-hot format of the pseudo clustering label obtained by spectral clustering. Clearly, our objective function is proposed to achieve

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

self-supervision for representation learning by minimizing the discrepancy between the pseudo label matrix $\widehat{\mathbf{L}}$ and the predicted label matrix $\widehat{\mathbf{Y}}^{(v)}$. Noticed that, when optimizing model (8), $\widehat{\mathbf{L}}$ is fixed.

To take advantage of the information in the clustering label matrix, motivated by [50], we minimize the mismatch between the coefficients matrix $\mathbf{C}$ and the clustering label matrix $\widehat{\mathbf{L}}$. Specifically, we supervise the learning of coefficient $\mathbf{C}$ by

$$\mathcal{L}_{\text{Self}} = \min_{\mathbf{C}} \sum_{p,q=1}^{N} |\mathbf{C}_{pq}| \frac{\left\| \widehat{\mathbf{l}}_p - \widehat{\mathbf{l}}_q \right\|_2^2}{2}, \tag{9}$$

where $\widehat{\mathbf{l}}_p, \widehat{\mathbf{l}}_q \in \widehat{\mathbf{L}} \in \mathbb{R}^{K \times N}$. By optimizing model (9), SGCMC help to enforce the self-expression coefficient matrix $\mathbf{C}$ to be such that an entry $\mathbf{C}_{pq}$ is nonzero only if the p-th node and q-th node have the same clustering labels. Hence, the previous clustering results can provide the self-supervision information for fine-tuning the coefficient matrix $\mathbf{C}$, which is helpful for node subspace clustering.

D. Implementation Details

Consequently, we joint subspace node clustering and self-supervised learning in an end-to-end trainable framework. The objective function of the proposed SGCMC is induced as

$$\mathcal{L} = \min_{\substack{\Theta_{\mathcal{E}}, \Theta_{\mathcal{D}} \\ \mathbf{C}, \mathbf{F}^{(v)}, \Theta_{\mathcal{H}}^{(v)}}} \mathcal{L}_{\text{GAE}} + \mathcal{L}_{\text{Sub}} + \lambda_3 \mathcal{L}_{\text{SS}} + \mathcal{L}_{\text{Con}}, \tag{10}$$

where $\mathcal{L}_{\text{GAE}} = \frac{1}{N}(\mathcal{L}_{\text{AR}} + \mathcal{L}_{\text{GR}})$, $\mathcal{L}_{\text{SS}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{Self}}$.

We optimize $\mathcal{L}$ via Adam algorithm [51] with gradient clipping. The dimensions of graph attention auto-encoder are $d_v \to d_{l1} \to d_{l2} \to d_{l1} \to d_v$, where d_v is the dimension of raw node content space in v-th view. The dimensions of self-supervised model are $d_{l2} \to 512 \to K$. We utilize the Relu activation function for all layers except the output layer of the self-supervised network, Softmax activation function is employed for the output layer of the self-supervised network. The learning rate of SGCMC is set to 3.0×10^{-5}. Due to the clustering labels provided by spectral clustering are up to an unknown permutation, thus resulting in the class labels from two successive epochs might not be consistent. We herein adopt the Hungarian algorithm [52] to find an optimal align between the pseudo labels of previous iterations before feeding them into the self-supervision learning model. Fixing $\widehat{\mathbf{Y}}$, we update other parameters in SGCMC for T_0 epoches, and then update $\widehat{\mathbf{Y}}$ once for obtaining stable results.

Finally, the optimization procedure of SGCMC is summarized in Algorithm 1.

Algorithm 1: Procedure for Training SGCMC.

Input: Node content: $\{\mathbf{X}^{(v)}\}_{v=1}^{V} \in \mathbb{R}^{d_v \times N}$, graph structure: $\mathbf{A} \in \mathbb{R}^{N \times N}$, cluster number K, hyper-parameters λ_1, λ_2, λ_3, d_{l1}, d_{l2}, learning rate and maximum number of iterations $\mathrm{T}_{\max}$.
Output: Clustering label $\widehat{\mathbf{L}}$.

Initialize graph attention auto-encoder, self-expressive coefficient matrix $\mathbf{C}$ and self-supervised module;
// Obtain multi-view representation
Get the representation $\mathbf{F}^{(v)}$ by Eq. (2);
// Obtain clustering label
Run spectral clustering on $\Delta = \frac{1}{2}\left(|\mathbf{C}| + \left|\mathbf{C}^{\mathrm{T}}\right|\right)$ to get clustering label $\widehat{\mathbf{L}}$;
// Obtain the output of FCN
Get $\widehat{\mathbf{Y}}^{(v)}$ by $\widehat{\mathbf{Y}}^{(v)} = \mathcal{H}[\mathbf{F}^{(v)}|\Theta_{\mathcal{H}}^{(v)}]$;
for $\mathrm{T} = 1 : \mathrm{T}_{max}$ **do**
  for $\mathrm{T}_1 = 1 : \mathrm{T}_0$ **do**
    // Update auto-encoder and FCN
    Fix clustering label $\widehat{\mathbf{L}}$, and update other parameters of SGCMC by Eq. (10);
  end
  // Obtain coefficient matrix
  Get the coefficient matrix $\mathbf{C}$;
  // Update clustering label
  Run spectral clustering on $\Delta = \frac{1}{2}\left(|\mathbf{C}| + \left|\mathbf{C}^{\mathrm{T}}\right|\right)$ to update clustering label $\widehat{\mathbf{L}}$;
end
return: Clustering results $\widehat{\mathbf{L}}$.

IV. Experiments

In this section, we report the performance of our proposed SGCMC for node clustering and compare it with several state-of-the-art methods. For comprehensive studies, we adopt three metrics to evaluate the clustering quality.

A. Datasets and Experimental Settings

For the proposed SGCMC, we implement it in TensorFlow 1.13.1 platform based on Python 3.6.[1] All the experiments are conducted on a machine with a Intel (R) Core (TM) i7-9700 K CPU and dual NVIDIA GeForce RTX 2080-Ti GPUs.

1) Datasets: We leverage different clustering tasks to evaluate the performance of our proposed SGCMC. These different tasks involve the following four datasets:

- **Cora** dataset [53] includes 2708 documents from 7 classes. The graph has 5429 edges. The node content of 1-st view is a 1433-dimension (D) binary matrix, which indicates the presence of the corresponding word.
- **Citeseer** dataset [54] includes 6 categories with 3312 publications. The node content of 1-st view is described by a 3703-D binary matrix. The graph has 4732 edges.
- **Wiki** dataset [37] contains 2405 documents with 17 classes. The graph has 17 981 edges. The node content of 1-st view is the 4973-D Term Frequency Inverse Document Frequency (TFIDF) matrix.
- **Heterogeneity Human Activity Recognition** (**HHAR**) dataset [55] consists of sensor records of 6 categories human activities, 10 299 nodes in total, where the node content of 1-st view is described by a 561-D feature matrix.

[1] Our codes are available at: [Online]. Available: https://github.com/xdweixia/SGCMC

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

TABLE I
NODE CLUSTERING PERFORMANCE COMPARISONS WITH 13 CLUSTERING METHODS ON CORA, CITESEER, WIKI AND HHAR DATABASES

Database	Cora			Citeseer			Wiki			HHAR		
Metric	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
K-Means	0.500	0.317	0.239	0.544	0.312	0.285	0.417	0.440	0.151	0.599	0.589	0.461
SC	0.398	0.297	0.174	0.308	0.090	0.082	0.220	0.182	0.015	0.345	0.582	0.324
DNGR	0.419	0.318	0.142	0.326	0.180	0.043	0.376	0.359	0.180	0.516	0.470	0.307
VGAE	0.530	0.397	0.293	0.380	0.174	0.141	0.451	0.468	0.263	0.713	0.630	0.515
GATE	0.658	0.527	0.451	0.616	0.401	0.381	0.482	0.343	0.188	0.728	0.728	0.625
MGAE	0.684	0.511	0.448	0.661	0.412	0.414	0.515	**0.485**	0.349	0.742	0.745	0.647
ARGAE	0.640	0.449	0.352	0.573	0350	0.341	0.381	0.345	0.112	0.736	0.729	0.636
ARVGAE	0.638	0.450	0.374	0.544	0.261	0.245	0.387	0.339	0.107	0.722	0.715	0.615
DAEGC	0.704	0.528	0.496	0.672	0.397	0.410	0.521	0.432	0.337	0.765	0.691	0.604
ARGAE-AX	0.711	0.526	0.495	0.581	0.338	0.301	0.420	0.389	0.213	0.749	0.743	0.658
DCCA-best	0.621	0.439	0.356	0.523	0.291	0.247	0.367	0.328	0.125	0.684	0.721	0.618
DCCAE-best	0.616	0.414	0.326	0.546	0.321	0.254	0.385	0.336	0.133	0.713	0.742	0.634
CO-GCN-best	0.735	0.567	0.512	0.655	0.432	0.423	0.515	0.450	0.330	0.813	0.778	0.699
SGCMC-Gabor	0.694	0.498	0.423	0.636	0.399	0.370	0.487	0.409	0.297	0.754	0.698	0.598
SGCMC-FFT	0.658	0.502	0.382	0.612	0.336	0.366	0.478	0.436	0.275	0.794	0.746	0.667
SGCMC-Cartesian	0.733	0.563	0.495	0.692	0.416	0.398	0.555	0.454	0.353	0.848	0.769	0.709
SGCMC-Eular	**0.761**	**0.609**	**0.542**	**0.715**	**0.456**	**0.473**	**0.576**	0.471	**0.366**	**0.860**	**0.783**	**0.728**

Best Results are Highlighted With Bold Numbers.

The undirected nearest neighbor graph is constructed with top-5 neighbors for HHAR dataset.

For all datasets, the node content of the 2-nd view is the Euler representation of raw node content. In Eq. (1), α is set to 1.1 for all datasets.

2) Baseline Methods: According to the input of different approaches, we compare the proposed SGCMC with the following 13 methods:

- **Method using $\mathbf{X}^{(1)}$ only**: K-Means.
- **Methods using $\mathbf{A}$ only**: spectral clustering (SC) and DNGR [34].
- **Methods using both $\mathbf{X}^{(1)}$ and $\mathbf{A}$**: variational GAE (VGAE) [13], GATE [17], ARGAE [14], DAEGC [16], marginalized graph auto-encoder (MGAE) [10], ARVGAE with attribute reconstruction (ARVGA-AX) [56].
- **Methods using both $\mathbf{X}_1$ and $\mathbf{X}_2$**: DCCA [46], DCCAE [47].
- **Method using $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{A}$**: Co-GCN [21].

3) Evaluation Criteria: Following [10], three popular metrics are used to evaluate the node clustering performance, *i.e.,* accuracy (ACC), normalized mutual information (NMI) and adjusted rand index (ARI). For all three metrics, a higher value indicates better performance.

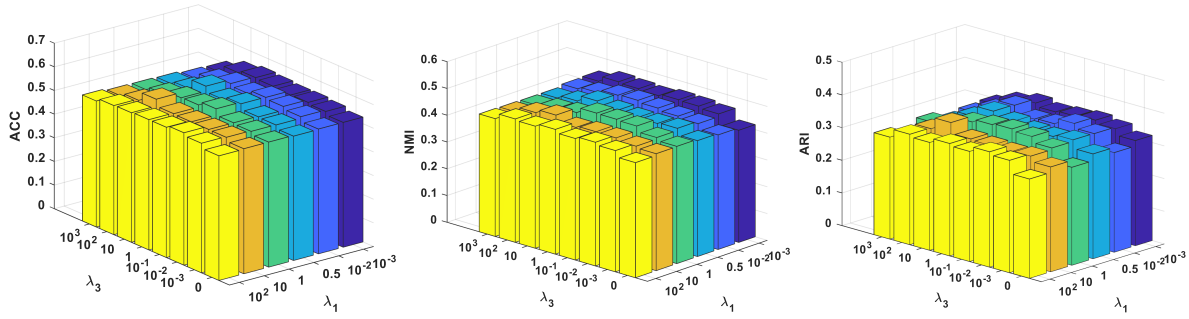
4) Parameter Setting: In the model (10), hyper-parameter λ_1 reflects the importance of the self-expression learning term, parameter λ_3 is used to balance the proportion of self-supervision. In the following experiments, we tune d^1 and d^2 in range of [128, 256, 512, 1024, 2048]. We tune λ_1 in the range of $[10^{-3}, 10^{-2}, 0.5, 1, 10, 10^2]$. We tune λ_3 in the range of $[0, 10^{-3}, 10^{-2}, 0.1, 1, 10, 10^2, 10^2]$ to get the best results. Specifically, d^1 and d^2 are set to 512, λ_1 is set to 100, and λ_3 is set to 10 on the Cora and HHAR datasets. d^1 is set to 1024, d^2 is set to 512, λ_1 is set to 100, and λ_3 is set to 10 on the Citeseer dataset. d^1 is set to 512, d^2 is set to 1024, λ_1 is set to 10, and λ_3 is set to 10 on the Wiki dataset. For all the compared methods, we follow the experiments settings in the corresponding papers.

B. Comparisons With State-of-The-Art Methods

To well estimate the clustering performance of our proposed SGCMC on node clustering task, we list the experimental results of SGCMC with three metrics in the aforementioned four datasets. Each experiment is run ten times, and we report the mean metric values in Table I. From these results, we have the following interesting observations:

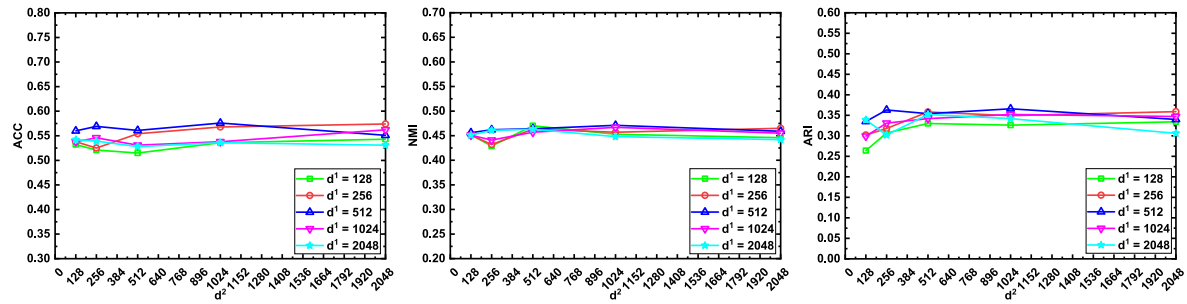
1) GCN based node clustering methods (GATE, MGAE, ARGAE, ARVGAE, ARGAE-AX, DAEGC, CO-GCN and SGCMC) are remarkably superior to other classical clustering methods. Even some single-view clustering methods based on GCN, *e.g.,* MGAE and ARVGAE, have better performance than those multi-view clustering methods, *e.g.,* DCCA and DCCAE. The reason may be that GCN based node clustering methods take advantage of the property of graph convolution network to process graph structure data. In contrast, GCN based methods can learn better node representation for clustering.
2) Our proposed SGCMC consistently obtains remarkable performances on Cora and Citeseer datasets, which shows its superiority in node clustering task. For example, on the Cora dataset, our method indicates a significant increase of 5.0%, 8.3%, and 4.7% *w.r.t.* ACC, NMI, and ARI compared to ARGAE-AX. The reason may be that our proposed SGCMC explicitly exploits the complementary information embedded in multi-view data, while single-view methods do not. Moreover, our method integrates coefficient matrix learning and node clustering into a unified framework, in which we explicitly consider the contribution of the clustering label by self-supervising coefficient matrix learning and latent representation learning. Thus, the learned coefficient matrix well characterizes the cluster structure.
3) Single-view GCN-based clustering methods are overall inferior to multi-view GCN-based clustering methods (CO-GCN and SGCMC). The reason may be that multi-view

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.



(a) ACC (b) NMI (c) ARI

Fig. 2. Parameter sensitivity of the trade-off parameters λ_1 and λ_3 on Wiki dataset. (a) ACC. (b) NMI. (c) ARI.



(a) ACC (b) NMI (c) ARI

Fig. 3. Parameter sensitivity of d^1 and d^2 on Wiki dataset. (a) ACC. (b) NMI. (c) ARI.

methods may leverage the complementary information embedded in multi-view data, while single-view methods do not.

4) For large-scale HHAR dataset, our method is still superior to state-of-the-art methods, which indicates that our method can effectively handle large-scale data.
5) The effect of different view descriptors: Take two views as an example to verify the effectiveness of different view descriptors, where the first view $\mathbf{X}^{(1)}$ is raw node content. We choose the second view descriptor $\mathbf{X}^{(2)}$ form different descriptors of the raw node content, including traditional multi-view descriptor, *e.g.*, Gabor, and some data conversion approaches, *e.g.*, Fast Fourier Transform (FFT), Cartesian product, and Euler transformation. We can observe that SGCMC achieves the best clustering performance with the raw node content and Euler transformation descriptor. This is because, compared with Euler transformation, other traditional multi-view descriptors are more suitable for characterizing Euclidean data, *e.g.*, face, handwritten digit, object, and scene. However, we handle the graph-structured data and that instead lies in an irregular domain. We map the raw node content into the complex space by Euler transform which not only suppresses outliers but also reveals non-linear patterns embedded in data.

C. Sensitivity Analysis of Different Parameters

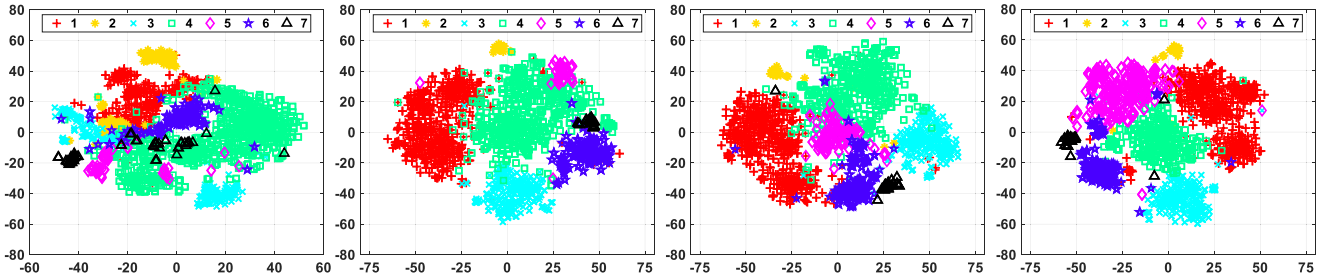
In the following experiments, we investigate the sensitivity of user-specified parameters, including the number of clusters, the trade-off parameters λ_1 and λ_3, the dimensions d^1 and d^2 of graph attention auto-encoder.

1) The Effect of the Trade-Off Parameters: We further make the experimental verification about the impact of the trade-off parameters λ_1 and λ_3 on Wiki dataset. The node clustering performances (ACC, NMI, and ARI) are shown in Fig. 2. It can be seen that, when the values of λ_1 and λ_3 are respectively 10 and 10, the optimal clustering performance is obtained. When $\lambda_3 = 0$, SGCMC is inferior to the best results with $\lambda_3 = 10$ on Wiki dataset, but its performance is still good. When $\lambda_3 > 0$, the clustering performance of our method remarkably increases. It indicates that λ_3 is important for improving clustering results, *i.e.*, the self-supervision via employing clustering label information is significant for clustering. Overall speaking, SGCMC obtains acceptable performances with most parameter combinations and is relatively robust for parameters λ_1 and λ_3.

2) The Effect of the Dimension of Different Layers: SGCMC has a symmetrical four-layer graph attention auto-encoder, d^1 and d^2 are the corresponding dimensions of the auto-encoder. We let d^1 and d^2 vary from [128, 256, 512, 1024, 2048] for Wiki dataset. Fig. 3 shows the variation of clustering ACC, NMI, and ARI with different d^1 and d^2. As reported in Fig. 3(a), (b), and (c), we can observe that our SGCMC gets a reasonable fluctuation when the dimensions of different layers range on a large scale. Therefore, SGCMC can keep stable when d^1 and d^2 vary within a reasonable range.

3) The Effect of the Number of Clusters: The number of clusters K is crucial for clustering, we analyze the stabilities of

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.



(a) 5 iteration (ACC = 0.534) (b) 30 iteration (ACC = 0.642) (c) 50 iteration (ACC = 0.716) (d) 80 iteration (ACC = 0.759)

Fig. 4. Performances *vs.* different cluster numbers on the Wiki dataset.

TABLE II
NODE CLUSTERING PERFORMANCE COMPARISONS WITH DIFFERENT STRATEGIES ON THREE DATASETS BASED ON THREE METRICS

Method				Cora			Citeseer			Wiki		
$\mathcal{L}_{\text{GAE}}$	$\mathcal{L}_{\text{Sub}}$	$\mathcal{L}_{\text{Con}}$	$\mathcal{L}_{\text{SS}}$	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
✓	✗	✗	✗	0.648	0.449	0.396	0.651	0.402	0.417	0.471	0.320	0.177
✓	✓	✗	✗	0.716	0.523	0.489	0.678	0.415	0.428	0.496	0.351	0.286
✓	✓	✓	✗	0.723	0.578	0.512	0.697	0.442	0.451	0.526	0.433	0.321
✓	✓	✓	✓	**0.761**	**0.609**	**0.542**	**0.715**	**0.456**	**0.473**	**0.576**	**0.471**	**0.366**

SGCMC on the Wiki dataset by varying K. Intuitively, as shown in Fig. 4, we vary K value in range of [5, 8, 10, 13, 16, 17, 20, 22, 25, 30]. One can observe that, as the number of clusters increases, *i.e.*, $K > 17$, the results of SGCMC generally decreases. This is because when the number of clusters increases, more uncertainty is triggered. Nevertheless, compared with most methods in Table I, the superiority of SGCMC still holds. This shows that SGCMC has adequate ability to tackle various clusters.

D. Ablation Study

We compare different strategies for training our SGCMC. For training a multi-view multi-layer graph attention auto-encoder, we analyze the following four cases:

1) Multi-view graph convolutional auto-encoder with reconstruction loss $\mathcal{L}_{\text{GAE}}$.
2) Case 1 with self-expression learning $\mathcal{L}_{\text{Sub}}$.
3) Case 2 with consistent representation constraint $\mathcal{L}_{\text{Con}}$.
4) Training of multi-view GCN-based subspace clustering and self-supervised learning.

Table II reports the results of different strategies for training SGCMC. It clearly demonstrates that each kind of strategy of SGCMC can improve the clustering performances effectively, especially after adding self-supervised learning in the multi-view GCN-based subspace clustering network. Fig. 6 demonstrates the significance of self-supervised learning strategy by comparing the visualization of the confusion matrix.

E. Visualization Verification

1) Visualizations on Real Dataset: By simultaneously exploiting the multi-view node contents and taking advantage of the clustering labels, SGCMC ought to learn a discriminative view-consensus coefficient matrix $\mathbf{C}$ and desirable clustering label at the same time. To illustrate how SGCMC achieves the goal, as shown in Fig. 5, we implement t-SNE [57] on the learned $\mathbf{C}$ at four different training iterations on Cora dataset, where

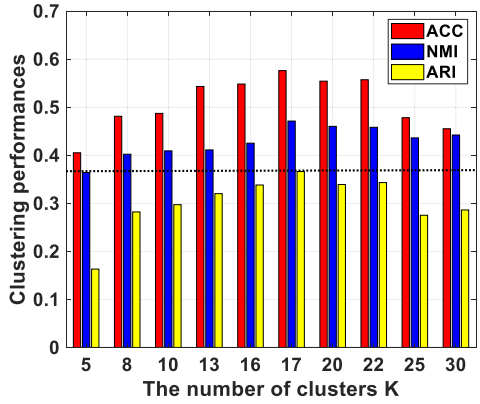
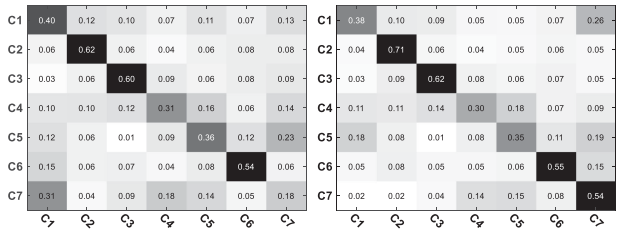


Fig. 5. The t-SNE visualizations on the Cora dataset with increasing training iteration. (a) 5 iteration (ACC = 0.534). (b) 30 iteration (ACC = 0.642). (c) 50 iteration (ACC = 0.716). (d) 80 iteration (ACC = 0.759).



(a) Without self-supervision (b) With self-supervision

Fig. 6. Confusion matrix on the Cora dataset. (a) Without self-supervision. (b) With self-supervision.

different colors indicated different clustering labels predicted by SGCMC. As observed, the cluster assignments become more reasonable, and different clusters scatter and gather more distinctly. These results indicate that the learned view-consensus

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

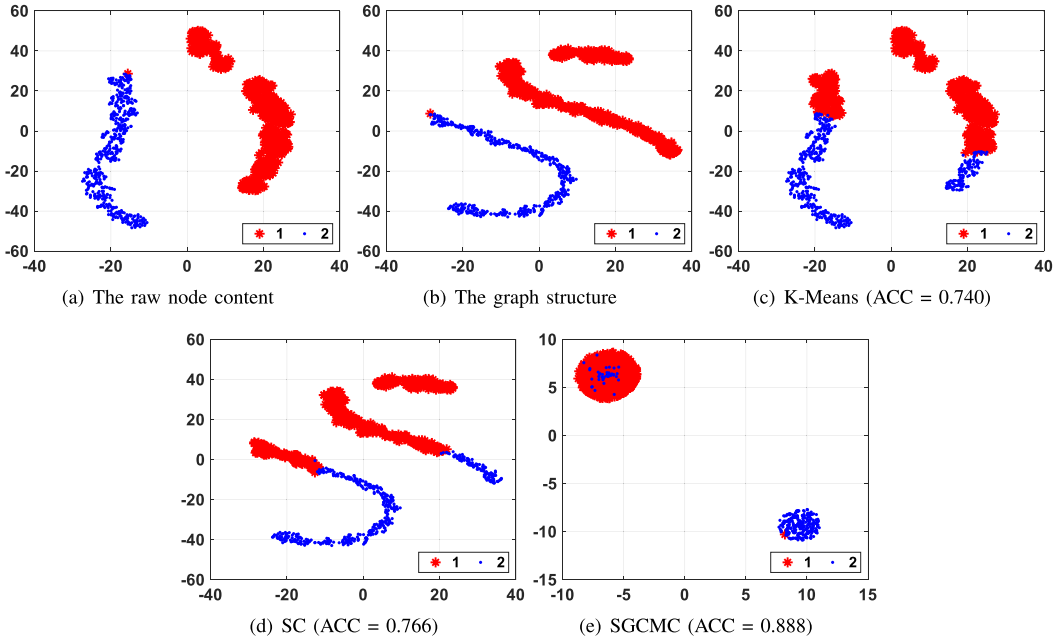


Fig. 7. The t-SNE visualizations on the two-moon toy dataset. In (a, c), the feature for t-SNE is the raw node content. In (b, d), the feature for t-SNE is the raw graph structure. In (e), the feature for t-SNE is the view-consensus coefficient matrix computed from SGCMC. In (a-b, d), the colors indicate the ground truth. In (c-d), the colors indicate the cluster assignment obtained from K-Means and SC, respectively. (a) The raw node content. (b) The graph structure. (c) K-Means (ACC = 0.740). (d) SC (ACC = 0.766). (e) SGCMC (ACC = 0.888).

coefficient matrix become more compact and discriminative with the increase of the iteration.

2) Visualizations on Synthetic Data: We also conduct visualization experiments on two-moon toy dataset[2] to validate the performance of the proposed SGCMC. To be specific, in our experiment, the used two-moon dataset has 1000 nodes, and each class has 500 nodes. The scale of raw node content belongs to $\mathbb{R}^{1000\times 2}$. Two view descriptors are raw node content and Euler representation of node content, respectively. We utilize the heat kernel approach to construct the corresponding graph structure. Fig. 7 shows the t-SNE visualization results. From these results, we can clearly observe that our method could separate the data into two clusters with higher accuracy. This is because the proposed SGCMC simultaneously exploit the multi-view node content and their corresponding graph structure. As shown in Fig. 7(e), although two clusters obtained by our proposed SGCMC scatter and gather more distinctly, there are still some nodes that are grouped incorrectly. The reason is that the dimension of raw node content of two-moon toy dataset is too small. It is challenging to learn a more suitable view-consensus coefficient matrix from a low-dimension node representation, and we will study this issue in the future.

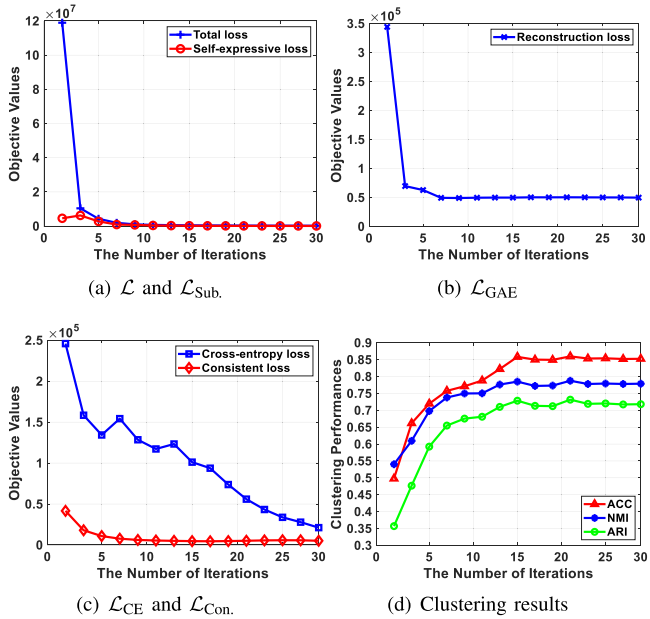


Fig. 8. The objective values and clustering performances of our proposed SGCMC with iterations on HHAR dataset. (a) $\mathcal{L}$ and $\mathcal{L}_{\text{Sub.}}$. (b) $\mathcal{L}_{\text{GAE}}$. (c) $\mathcal{L}_{\text{CE}}$ and $\mathcal{L}_{\text{Con.}}$. (d) Clustering results.

F. Convergence Analysis

Taking HHAR dataset as an example, we investigate the convergence of our proposed SGCMC. We record the objective values and clustering results of our proposed SGCMC with iteration and plot them in Fig. 8. As shown in Fig. 8(a)–(c), the objective values decrease a lot in the first 15 iterations, then continuously decrease until convergence. As for the clustering performances, as shown in Fig. 8(d), the ACC of our proposed SGCMC continuously increases to a maximum in the first 15 iterations, and generally maintain stable to slight variation. The curves in terms of NMI and ARI metrics have a similar trend. These observations clearly indicate that our proposed SGCMC usually converges quickly.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion

In this paper, we study the multi-view GCN based clustering, and propose a novel multi-view self-supervised graph convolutional subspace clustering network (SGCMC). SGCMC maps the original node content to an explicit Euler feature space and does not increase the dimensionality of features. To make full use of the inaccurate clustering label, SGCMC utilizes the clustering label to guide the learning of node representation and coefficient matrix learning, where the latter is used in turn to conduct the subsequent clustering. By this way, clustering and representation learning are seamlessly connected, with the aim to achieve better clustering performance. Extensive experimental results also show the effectiveness of such strategy. In the future, we are interested in exploring the multi-view GCN-based clustering with contrastive learning [58]–[60], which is helpful for learning more robust and effective graph embedding representation.

Acknowledgment

The authors would like to thank the anonymous reviewers and AE for their constructive comments and suggestions. Also deserving recognition Professor Chun-Guang Li for providing the code of his IEEE CVPR paper [50], Professor C. Shi for providing the HHAR dataset.

References

[1] J. Tang, X. Tang, and J. Yuan, "Traffic-optimized data placement for social media," *IEEE Trans. Multim.*, vol. 20, no. 4, pp. 1008–1023, Apr. 2018.
[2] L. Xu, T. Bao, L. Zhu, and Y. Zhang, "Trust-based privacy-preserving photo sharing in online social networks," *IEEE Trans. Multim.*, vol. 21, no. 3, pp. 591–602, Mar. 2019.
[3] P. Zhou, K. Wang, J. Xu, and D. O. Wu, "Differentially-private and trustworthy online social multimedia big data retrieval in edge computing," *IEEE Trans. Multim.*, vol. 21, no. 3, pp. 539–554, Mar. 2019.
[4] F. Xue *et al.*, "Knowledge-based topic model for multi-modal social event analysis," *IEEE Trans. Multim.*, vol. 22, no. 8, pp. 2098–2110, Aug. 2020.
[5] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.
[6] W. Peng, X. Hong, H. Chen, and G. Zhao, "Learning graph convolutional network for skeleton-based human action recognition by neural searching," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 2669–2676.
[7] Z. Qiu, K. Qiu, J. Fu, and D. Fu, "DGCN: Dynamic graph convolutional network for efficient multi-person pose estimation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 11924–11931.
[8] Y. Wu, D. Lian, Y. Xu, L. Wu, and E. Chen, "Graph convolutional networks with markov random field reasoning for social spammer detection," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1054–1061.
[9] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 7370–7377.
[10] C. Wang, S. Pan, G. Long, X. Zhu, and J. Jiang, "MGAE: Marginalized graph autoencoder for graph clustering," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 889–898.
[11] X. Yang, W. Yu, R. Wang, G. Zhang, and F. Nie, "Fast spectral clustering learning with hierarchical bipartite graph for large-scale data," *Pattern Recognit. Lett.*, vol. 130, pp. 345–352, 2020.
[12] W. Xia, Q. Gao, Q. Wang, and X. Gao, "Regression-based clustering network via combining prior information," *Neurocomputing*, vol. 448, pp. 324–332, 2021.
[13] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *in Proc. NIPS Workshop Bayesian Deep Learn.*, 2016.
[14] S. Pan *et al.*, "Adversarially regularized graph autoencoder for graph embedding," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 2609–2615.
[15] P. Veličković *et al.*, "Graph attention networks," *in Proc. Int. Conf. Learn. Representations*, 2018.
[16] C. Wang *et al.*, "Attributed graph clustering: A deep attentional embedding approach," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019, pp. 3670–3676.
[17] A. Salehi and H. Davulcu, "Graph attention auto-encoders," in *Proc. IEEE 32nd Int. Conf. Tools Artif. Intell.*, 2020, pp. 989–996.
[18] X. Shen *et al.*, "Semi-paired discrete hashing: Learning latent hash codes for semi-paired cross-view retrieval," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4275–4288, Dec. 2017.
[19] D. Xie *et al.*, "Multiview clustering by joint latent representation and similarity learning," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4848–4854, Nov. 2020.
[20] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, "Tensor-svd based graph learning for multi-view subspace clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 3930–3937.
[21] S. Li, W. Li, and W. Wang, "Co-GCN for multi-view semi-supervised learning," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 4691–4698.
[22] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2408–2414.
[23] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, Feb. 2020.
[24] C. Tang *et al.*, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Trans. Multim.*, vol. 21, no. 7, pp. 1724–1736, Jul. 2019.
[25] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Trans. Multim.*, vol. 22, no. 8, pp. 1985-1997, Aug. 2020.
[26] D. G. Lowe, "Distinctive image features from scaleinvariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[27] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[28] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[29] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 498–518, 2013.
[30] S. Liao *et al.*, "Discriminant analysis via joint euler transform and $\ell_{2,1}$-norm," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5668–5682, Nov. 2018.
[31] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition and applications," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3070843.
[32] Z. Wu *et al.*, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.
[33] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.
[34] S. Cao, W. Lu, and Q. Xu, "Deep neural networks for learning graph representations," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1145–1152.
[35] S. Cavallari, V. W. Zheng, H. Cai, K. C. Chang, and E. Cambria, "Learning community embedding with community detection and node embedding on graphs," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 377–386.
[36] S. Cavallari, E. Cambria, H. Cai, K. C. Chang, and V. W. Zheng, "Embedding both finite and infinite communities on graphs [application notes]," *IEEE Comput. Intell. Mag.*, vol. 14, no. 3, pp. 39–50, Aug. 2019.
[37] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Y. Chang, "Network representation learning with rich text information," in *Proc. Int. Joint Conf. Artif. Intell.*, 2015, pp. 2111–2117.
[38] K. Zhan and C. Niu, "Mutual teaching for graph convolutional networks," *Future Gener. Comput. Syst.*, vol. 115, pp. 837–843, 2021.
[39] H. N. Tran and E. Cambria, "A survey of graph processing on graphics processing units," *J. Supercomput.*, vol. 74, no. 5, pp. 2086–2115, 2018.
[40] D. Bo *et al.*, "Structural deep clustering network," in *Proc. ACM Web Conf.*, 2020, pp. 1400–1410.
[41] S. Fan *et al.*, "One2multi graph autoencoder for multi-view graph clustering," in *Proc. ACM Web Conf.*, 2020, pp. 3070–3076.
[42] J. Park, M. Lee, H. J. Chang, K. Lee, and J. Y. Choi, "Symmetric graph convolutional autoencoder for unsupervised graph representation learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6518–6527.
[43] X. Shen, W. Liu, I. W. Tsang, Q. Sun, and Y. Ong, "Multilabel prediction via cross-view search," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4324–4338, Sep. 2018.
[44] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.

[45] W. Xia *et al.*, "Multi-view subspace clustering by an enhanced tensor nuclear norm," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2021.3052352.
[46] G. Andrew, R. Arora, J. A. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, vol. 28, 2013, pp. 1247–1255.
[47] W. Wang, R. Arora, K. Livescu, and J. A. Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 1083–1092.
[48] Q. Wang, J. Cheng, Q. Gao, G. Zhao, and L. Jiao, "Deep multi-view subspace clustering with unified and discriminative learning," *IEEE Trans. Multim.*, to be published, doi: 10.1109/TMM.2020.3025666.
[49] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[50] J. Zhang *et al.*, "Self-supervised convolutional subspace clustering network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5473–5482.
[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR Track Proc.*, 2015.
[52] J. Munkres., "Normalized cuts and image segmentation," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.
[53] A. McCallum, K. Nigam, J. Rennie, and K. Seymore, "Automating the construction of internet portals with machine learning," *Inf. Retr.*, vol. 3, no. 2, pp. 127–163, 2000.
[54] C. L. Giles, K. D. Bollacker, and S. Lawrence, "CiteSeer: An automatic citation indexing system," in *Proc. ACM 3rd Conf. Digit. Lib.*, 1998, pp. 89–98.
[55] A. Stisen *et al.*, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 127–140.
[56] S. Pan *et al.*, "Learning graph embedding with adversarial training methods," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2475–2487, Jun. 2020.
[57] L. van der Maaten and G. Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.
[58] P. Velickovic *et al.*, "Deep graph infomax," in *Proc. ICLR*, 2019.
[59] Y. You *et al.*, "Graph contrastive learning with augmentations," in *Proc. NeurIPS*, 2020.
[60] Y. Li *et al.*, "Contrastive clustering," in *Proc. Conf. Artif. Intell.*, 2021, pp. 8547–8555.

Wei Xia (Graduate Student Member, IEEE) received the B.Eng. degree in communication engineering from the Lanzhou University of Technology, Lanzhou, China, in 2018. He is currently working toward the Ph.D. degree in communication and information system with Xidian University, Xi'an, China. His research interests include pattern recognition, machine learning, and deep learning.

Qianqian Wang received the B.Eng. degree in communication engineering from the Lanzhou University of Technology, Lanzhou, China, in 2014 and the Ph.D. degree from Xidian University, Xi'an China, in 2019. She is currently a Lecturer with the School of Telecommunications Engineering, Xidian University. Her research interests include pattern recognition, dimensionality reduction, sparse representation, and face recognition.

Quanxue Gao received the B.Eng. degree from Xi'an Highway University, Xi'an, China, in 1998, the M.S. degree from the Gansu University of Technology, Lanzhou, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an China, in 2005. From 2006 to 2007, he was an Associate Research with the Biometrics Center, The Hong Kong Polytechnic University, Hong Kong. From 2015 to 2016, he was a Visiting Scholar with the Department of Computer Science, The University of Texas at Arlington, Arlington, TX, USA. He is currently a Professor with the School of Telecommunications Engineering, Xidian University, Xi'an, China, and also a Key Member of State Key Laboratory of Integrated Services Networks. He has authored about 80 technical articles in refereed journals and proceedings, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, CVPR, AAAI, and IJCAI. His current research interests include pattern recognition and machine learning.

Xiangdong Zhang received the B.Eng., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 1992, 1995, and 1998, respectively. He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University. His current research interests include pattern recognition and machine learning.

Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electronic engineering, signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor with the Ministry of Education of P. R. China, a Professor of pattern recognition and intelligent system with Xidian University and a Professor of computer science and technology with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has authored or coauthored six books and about 300 technical articles in refereed journals and proceedings. His current research interests include Image processing, computer vision, multimedia analysis, machine learning, and pattern recognition. He is on the Editorial Boards of various journals, including the *Signal Processing (Elsevier)* and *Neurocomputing (Elsevier)*. He was the General Chair or Co-Chair, Program Committee Chair or Co-Chair, or a PC Member for about 30 major international conferences. He is a fellow of the Institute of Engineering and Technology and a fellow of the Chinese Institute of Electronics.

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on September 11,2023 at 11:55:55 UTC from IEEE Xplore. Restrictions apply.