# Toward Multi-Diversified Ensemble Clustering of High-Dimensional Data: From Subspaces to Metrics and Beyond

Dong Huang, *Member, IEEE,* Chang-Dong Wang, *Member, IEEE,*
Jian-Huang Lai, *Senior Member, IEEE,* and Chee-Keong Kwoh, *Senior Member, IEEE*

***Abstract*—The rapid emergence of high-dimensional data in various areas has brought new challenges to current ensemble clustering research. To deal with the curse of dimensionality, recently considerable efforts in ensemble clustering have been made by means of different subspace-based techniques. However, besides the emphasis on subspaces, rather limited attention has been paid to the potential diversity in similarity/dissimilarity metrics. It remains a surprisingly open problem in ensemble clustering how to create and aggregate a large population of diversified metrics, and furthermore, how to jointly investigate the multi-level diversity in the large populations of metrics, subspaces, and clusters in a unified framework. To tackle this problem, this paper proposes a novel multi-diversified ensemble clustering approach. In particular, we create a large number of diversified metrics by randomizing a scaled exponential similarity kernel, which are then coupled with random subspaces to form a large set of metric-subspace pairs. Based on the similarity matrices derived from these metric-subspace pairs, an ensemble of diversified base clusterings can thereby be constructed. Further, an entropy-based criterion is utilized to explore the cluster-wise diversity in ensembles, based on which three specific ensemble clustering algorithms are presented by incorporating three types of consensus functions. Extensive experiments are conducted on 30 high-dimensional datasets, including 18 cancer gene expression datasets and 12 image/speech datasets, which demonstrate the superiority of our algorithms over the state-of-the-art. The source code is available at https://github.com/huangdonghere/MDEC.**

***Index Terms*—Ensemble clustering, Consensus clustering, High-dimensional data, Random subspaces, Diversified metrics.**

## I. Introduction

THE last decade has witnessed significant progress in the development of the ensemble clustering technique [1]–[22], which is typically featured by its ability of combining multiple base clusterings into a probably better and more robust consensus clustering and has recently shown promising advantages in discovering clusters of arbitrary shapes, dealing with noisy data, coping with data from multiple sources, and producing robust clustering results [7].

Dong Huang is with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China, and also with Pazhou Lab, Guangzhou, China. E-mail: huangdonghere@gmail.com.

Chang-Dong Wang and Jian-Huang Lai are with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, and also with Guangdong Key Laboratory of Information Security Technology, Guangzhou, China, and also with Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China. E-mail: changdongwang@hotmail.com, stsljh@mail.sysu.edu.cn.

Chee-Keong Kwoh is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. E-mail: asckkwoh@ntu.edu.sg.

In recent years, with high-dimensional data widely appearing in various areas, new challenges have been brought to the conventional ensemble clustering algorithms, which, however, often lack the ability to well address the high-dimensional issues. As is called the curse of dimensionality, it is highly desired but very difficult to find the inherent cluster structure hidden in the huge dimensions, especially when it is frequently coupled with quite low sample size. Recently some efforts have been devoted to ensemble clustering of high-dimensional data, which typically exploit some subspace-based (or feature-based) techniques (such as random subspace sampling [13], [23]–[26], stratified subspace sampling [6], and subspace projection [27]) to explore the diversity in high-dimensionality. Inherently, these subspace-based techniques select or linearly combine data features into different subsets (i.e., subspaces) by a variety of strategies to seek more perspectives for finding cluster structures.

Besides the issue of subspaces (or features), the choice of similarity/dissimilarity metrics is another crucial factor in dealing with high-dimensional data [28], [29]. The existing ensemble clustering methods generally adopt one or a few pre-selected metrics, which are often selected explicit or implicitly based on the expert's knowledge or some prior assumptions. However, few, if not none, of them have considered the potentially huge opportunities hidden in randomized metric spaces. On the one hand, it is very difficult to select or learn an optimal metric for a given dataset without human supervision or implicit assumptions. On the other hand, with different metrics capable of reflecting different perspectives on data, the joint use of a large number of randomized/diversified metrics may reveal more valuable information hidden in high-dimensionality. However, it is surprisingly still an open problem in ensemble clustering how to produce and aggregate a large number of diversified metrics to enhance the consensus performance. Furthermore, starting from the metric diversification problem, another crucial challenge arises as to how to jointly exploit multiple levels of diversity in the large number of metrics, subspaces, and clusters, in a unified ensemble clustering framework.

To tackle the above-mentioned problem, this paper presents a novel ensemble clustering approach termed multi-diversified ensemble clustering (MDEC) by jointly exploiting large populations of diversified metrics, random subspaces, and weighted clusters. Particularly, a scaled exponential similarity kernel is utilized as the seed kernel, which has advantage in parameter

flexibility and neighborhood adaptivity and is randomized to breed a large set of diversified metrics. Then, the set of diversified metrics are coupled with random subspaces to form a large number of metric-subspace pairs, which contribute to the jointly randomized ensemble generation process where the set of diversified base clusterings can thereby be produced with the help of spectral clustering. Further, to exploit the cluster-wise diversity in the multiple base clusterings, an entropy-based cluster validity strategy is adopted to evaluate and weight each base cluster by considering the distribution of clusters in the entire ensemble, based on which three specific ensemble clustering algorithms are therefore proposed by incorporating three types of consensus functions. Extensive experiments are conducted on 30 real-world high-dimensional datasets, including 18 cancer gene expression datasets and 12 image/speech datasets. The experimental results have shown the superiority of our approach against the state-of-the-art ensemble clustering approaches for clustering high-dimensional data.

For clarity, the main contributions of this work are summarized as follows:

1) This paper for the first time, to our knowledge, shows that the joint use of a large population of randomly diversified metrics can significantly benefit the ensemble clustering of high-dimensional data in a purely unsupervised manner.
2) A new metric diversification strategy is proposed by randomizing the scaled exponential similarity kernel with both parameter flexibility and neighborhood adaptivity considered, which is further coupled with random subspace sampling for the jointly randomized generation of base clusterings.
3) A new ensemble clustering framework termed MDEC is presented, which is capable of simultaneously exploiting large populations of diversified metrics, random subspaces, and weighted clusters in a unified model for high-dimensional data.
4) Three specific algorithms are designed under the proposed MDEC framework by incorporating three types of consensus functions. Experiments conducted on a variety of real-world high-dimensional datasets have confirmed the advantages of the proposed algorithms over the state-of-the-art.

The remainder of the paper is organized as follows. The related work is reviewed in Section II. The proposed ensemble clustering framework is described in Section III. The experimental results are reported in Section IV. Finally, the paper is concluded in Section V.

## II. Related Work

Due to its ability of combining multiple base clusterings into a probably better and more robust consensus clustering, the ensemble clustering technique has been receiving increasing attention in recent years. Many ensemble clustering algorithms have been developed from different technical perspectives [1]–[5], [7]–[12], [15], [30]–[42], which can be classified into three main categories, namely, the pair-wise co-occurrence based methods, the graph partitioning based methods, and the median partition based methods.

The pair-wise co-occurrence based methods [30], [37], [38] typically construct a co-association matrix by considering the frequency that two data samples occur in the same cluster among the multiple base clusterings. The co-association matrix is then used as the similarity matrix for the data samples, upon which some clustering algorithms can thereby be performed to obtain the final clustering result. Fred and Jain [30] first introduced the concept of the co-association matrix and proposed the evidence accumulation clustering (EAC) method, which applied a hierarchical agglomerative clustering algorithm [43] on the co-association matrix to build the consensus clustering. To extend the EAC method, Wang et al. [37] took the cluster sizes into consideration and proposed the probability accumulation method. Yi et al. [38] dealt with the uncertain entries in the co-association matrix by first labeling them as unobserved, and then recovering the unobserved entries by the matrix completion technique. Liu et al. [11] proved that the spectral clustering of the co-association matrix is equivalent to a weighted version of $K$-means, and proposed the spectral ensemble clustering (SEC) method to effectively and efficiently obtain the consensus result.

The graph partitioning based methods [8], [39], [40] generally construct a graph model for the ensemble of multiple base clusterings, and then partition the graph into several disjoint subsets to obtain the final clustering result. Strehl and Ghosh [39] solved the ensemble clustering problem by using three graph partitioning based algorithms, namely, cluster-based similarity partitioning algorithm (CSPA), hypergraph partitioning algorithm (HGPA), and meta-clustering algorithm (MCLA). Fern and Brodley [40] formulated a bipartite graph model by treating both clusters and data samples as nodes, and partitioned the graph by the METIS algorithm [44] to obtain the consensus result. Huang et al. [8] dealt with the ensemble clustering problem by sparse graph representation and random walk trajectory analysis, and presented the probability trajectory based graph partitioning (PTGP) method.

The median partition based methods [9], [41], [42] typically formulate the ensemble clustering problem into an optimization problem which aims to find the median partition such that the similarity between the base partitions (i.e., base clusterings) and the median partition is maximized. The median partition problem is NP-hard [41]. To find an approximate solution, Topchy et al. [41] cast the median partition problem into a maximum likelihood problem and solved it by the EM algorithm. Franek and Jiang [42] reduced the ensemble clustering problem to an Euclidean median problem and solved it by the Weiszfeld algorithm [45]. Huang et al. [9] formulated the ensemble clustering problem into a binary linear programming problem and obtained an approximate solution based on the factor graph model and the max-product belief propagation [46].

Although in recent years significant advances have been made in the research of ensemble clustering [1]–[5], [7]–[12], [15], [30]–[42], yet the existing methods are mostly devised for general-purpose scenarios and lack the desirable ability to appropriately address the clustering problem of high-

dimensional data. More recently, some efforts have been made to deal with the curse of dimensionality, where subspace-based (or feature-based) techniques are often exploited. Jing et al. [6] adopted stratified feature sampling to generate a set of subspaces, which are further incorporated into several ensemble clustering algorithms to build the consensus clustering for high-dimensional data. Yu et al. [24] proposed a novel subspace-based ensemble clustering framework termed AP$^2$CE, which integrates random subspaces, affinity propagation, normalized cut, and five candidate distance metrics. Further, Yu et al. [25] proposed a semi-supervised subspace-based ensemble clustering framework by incorporating random subspaces, constraint propagation, incremental ensemble selection, and normalized cut into the framework. Fern and Brodley [27] exploited random subspace projection to build a set of subspaces, which in fact are obtained by (randomly) linear combination of features (or feature sets). Chu et al. [47] used three clustering results (generated by three different clustering algorithms) to guide the feature learning process. These methods [6], [24], [25], [27], [47] typically exploit the diversity in high-dimensionality by various subspace-based or feature learning based techniques, but few of them have fully considered the potentially huge diversity in metric spaces. The existing methods [6], [24], [25], [27], [47] generally use one or a few preselected similarity/disimilarity metrics, which are selected implicitly based on the expert's knowledge or some prior assumptions. Although the method in [24] proposed to randomly select a metric out of the five candidate metrics at each time, yet it still failed to go beyond a few metrics to explore the huge potential hidden in a large number of diversified metrics, which may play a crucial role in clustering high-dimensional data. The key challenge here lies in how to create such a large number of highly diversified metrics, and further how to jointly exploit the diversity in the large number of metrics, together with subspace-wise diversity and cluster-wise diversity, to achieve a unified ensemble clustering framework for high-dimensional data.

## III. Proposed Framework

This section describes the overall algorithm of the proposed ensemble clustering framework. A brief overview is provided in Section III-A. The metric diversification process is presented in Section III-B. The jointly randomized ensemble generation is introduced in Section III-C. Finally the consensus functions are given in Section III-D.

### A. Brief Overview

In this paper, we propose a novel multi-diversified ensemble clustering (MDEC) framework (as shown in Fig. 1). First, we create a large number of diversified metrics by randomizing a scaled exponential similarity kernel, and combine the diversified metrics with the random subspaces to form a large set of random metric-subspace pairs. Second, with each random metric-subspace pair, we construct a similarity matrix for the data samples. The spectral clustering is then performed on these similarity matrices derived from metric-subspace pairs to obtain an ensemble of base clusterings. Third, to exploit the cluster-wise diversity in the ensemble of multiple base clusterings, we adopt an entropy based criterion to evaluate and weight the clusters by considering the distribution of cluster labels in the entire ensemble. With the weighted clusters, a locally weighted co-association matrix is further constructed to serve as a summary of the ensemble. Finally, to obtain the consensus clustering, three types of consensus functions are presented based on hierarchical clustering (HC), spectral clustering (SC), and bipartite graph (BG) model, respectively, leading to three specific ensemble clustering algorithms termed MDEC-HC, MDEC-SC, and MDEC-BG, respectively. It is noteworthy that we simultaneously incorporate three levels of diversity, i.e., metric-wise diversity, subspace-wise diversity, and cluster-wise diversity, in a unified framework, which has shown its advantages in dealing with high-dimensional data when compared to the state-of-the-art ensemble clustering algorithms. In the following sections, we will further introduce each component of the proposed framework.

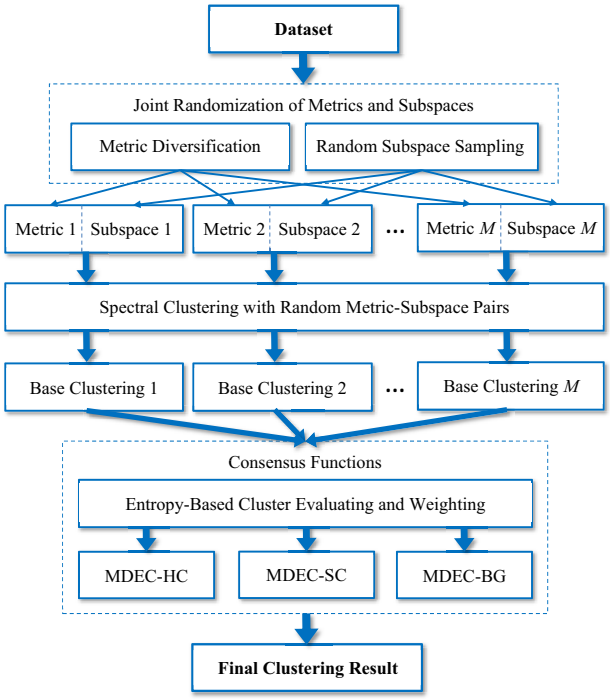



Fig. 1. Flow diagram of the proposed MDEC approach.

### B. Diversification of Metrics

The choice of similarity/dissimilarity metrics plays a crucial role in the field of data mining and machine learning [48]–[53], especially for high-dimensional data analysis where the high-dimensionality further complicates the use of metrics [49], [53]. Some supervised or semi-supervised learning techniques have been developed to learn a suitable metric for some specific applications [48]–[53], but in unsupervised scenarios it is still very difficult to discover one or a few proper metrics given a task without prior knowledge.

Instead of relying on one or a few selected or learned metrics, this paper proposes to jointly use a large number of

randomly diversified metrics, which will be further coupled with a large number of random subspaces, in a unified ensemble clustering framework for high-dimensional data. Here, two sub-problems are involved, namely, how to create a large number of diversified metrics, and how to collectively exploit them in ensemble clustering.

To create the diversified metrics, we take advantage of the kernel trick with randomization incorporated. The kernel similarity metrics have been proved to be a powerful tool for clustering complex data [48], [51], which, however, suffer from the difficulties in selecting proper kernel parameters. The kernel parameters can be learned by some metric learning techniques [48], [51] with supervision or semi-supervision. But without human supervision, it is often extremely difficult to decide proper kernel parameters. This is a critical disadvantage of kernel methods for conventional (unsupervised) applications, which, nevertheless, just becomes an important advantage in our situation where what is highly desired is not the selection of a good kernel similarity metric, but the flexibility to create a large number of diversified ones.

Specifically, in this paper, we adopt a scaled exponential similarity kernel [54] as the seed kernel, which has advantage in parameter flexibility and neighborhood adaptivity and is then randomized to breed a large population of diversified metrics. Given a set of $N$ data samples $\mathcal{X} = \{x_1, \cdots, x_N\}$, where $x_i \in \mathbb{R}^D$ is the $i$-th sample and $D$ is the number of features. The kernel similarity between samples $x_i$ and $x_j$ is defined as:

$$\varphi^{\mathcal{X}}(x_i, x_j) = \begin{cases} \exp\left(-\frac{d(x_i,x_j)}{\mu\varepsilon_{ij}}\right), & \text{if } x_i \in N_k(x_j) \\ & \text{or } x_j \in N_k(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $\mu$ is a hyperparameter, $\varepsilon_{ij}$ is a scaling term, $d(x_i, x_j)$ is the Euclidean distance between $x_i$ and $x_j$, and $N_k(x_i)$ denotes the set of the $k$ nearest neighbors of $x_i$. The average distance between $x_i$ and its $k$ nearest neighbors can be computed as

$$\rho_k(x_i) = \frac{1}{k} \cdot \sum_{x_p \in N_k(x_i)} d\left(x_i, x_p\right). \quad (2)$$

Then, as suggested in [54], to simultaneously take into consideration the neighborhood of $x_i$, the neighborhood of $x_j$, and their distance, the scaling term $\varepsilon_{ij}$ is defined as the average of $\rho_k(x_i)$, $\rho_k(x_j)$, and $d(x_i, x_j)$. That is

$$\varepsilon_{ij} = \frac{\rho_k(x_i) + \rho_k(x_j) + d(x_i, x_j)}{3}. \quad (3)$$

The above kernel is a variant of the Gaussian kernel. It has two free parameters, i.e., the hyperparameter $\mu$ and the number of nearest neighbors $k$. The motivation to adopt the SES kernel as the seed kernel in our approach is two-fold. First, with the influence of the scaling term where the $k$-nearest-neighbors' information is incorporated, the kernel has the adaptivity to the neighborhood structure among data samples. Moreover, with each $k$ value corresponding to a specific neighborhood size, by randomizing the parameter $k$, multi-scale neighborhood information can be explored to enhance the diversity. Second, the two free parameters $k$ and $\mu$ in the kernel provide high flexibility for adjusting the influence of the kernel and can contribute to the desired diversity of the generated metrics by randomly perturbing the two parameters.

Specifically, we randomly select the two parameters $\mu \in [\mu_{min}, \mu_{max}]$ and $k \in [k_{min}, k_{max}]$, respectively, as follows:

$$\mu = \mu_{min} + \sigma_1(\mu_{max} - \mu_{min}), \quad (4)$$

$$k = k_{min} + \lfloor \sigma_2(k_{max} - k_{min}) \rfloor, \quad (5)$$

where $\sigma_1 \in [0, 1]$ and $\sigma_2 \in [0, 1]$ are two uniform random variables, and $\lfloor\rfloor$ outputs the floor of a real number.

Note that our objective is not to find a good pair of parameters $\mu$ and $k$, but to randomize them to yield a large population of diversified metrics. The parameters $\mu$ and $k$ are suggested to be randomly selected in a wide range to enhance the diversity.

By performing the random selection $M$ times, a set of $M$ pairs of $\mu$ and $k$ are obtained, which correspond to $M$ randomized kernel similarity metrics for the dataset $\mathcal{X}$, denoted as

$$\varphi^{\mathcal{X}}_{\mu_1,k_1}(\cdot,\cdot), \varphi^{\mathcal{X}}_{\mu_2,k_2}(\cdot,\cdot), \cdots, \varphi^{\mathcal{X}}_{\mu_M,k_M}(\cdot,\cdot), \quad (6)$$

where $\mu_i$ and $k_i$ are the $i$-th pair of randomized parameters.

### C. Ensemble Generation by Joint Randomization

In high-dimensional data, the cluster structures are often hidden in different low-dimensional subspaces [49], [53]. Besides the subspaces, different metrics may also provide complementary information for investigating the high-dimensional space. In this section, with the set of diversified metrics generated, we proceed to couple the large number of diversified metrics with random subspaces to explore the rich information in various subspaces in high-dimensional data.

Specifically, let $\mathcal{F} = \{f_1, \cdots, f_D\}$ be the set of features in the dataset $\mathcal{X}$, where $f_i$ denotes the $i$-th feature. A random subspace is a set of a certain number of features that are randomly sampled from the original feature set. The cluster structure of high-dimensional data may be hidden in differen feature subspaces as well as in different metric spaces. In this paper, we propose to jointly exploit large populations of diversified metrics and random subspaces. Specifically, we perform random subspace sampling $M$ times to obtain $M$ random subspaces, denoted as $\mathcal{F}_1, \cdots, \mathcal{F}_M$, which lead to $M$ component datasets, denoted as $\mathcal{X}_1, \cdots, \mathcal{X}_M$. Note that each component dataset $\mathcal{X}_i$ has the same number of data samples as the original dataset $\mathcal{X}$, but its feature set $\mathcal{F}_i$ only consists of $d \leq D$ attributes that are randomly sampled from $\mathcal{F}$ with a sampling ratio $\tau \in (0, 1]$. Obviously, if $\tau = 1$, then it means every subspace is in fact the original feature space, i.e., no sub-sampling actually happens. Here, with the random subspaces generated, we can couple each of them with a randomly diversified metric (as describe in Section III-B), and thus obtain $M$ random metric-subspace pairs, denoted as

$$\varphi^{\mathcal{X}_1}_{\mu_1,k_1}(\cdot,\cdot), \varphi^{\mathcal{X}_2}_{\mu_2,k_2}(\cdot,\cdot), \cdots, \varphi^{\mathcal{X}_M}_{\mu_M,k_M}(\cdot,\cdot). \quad (7)$$

In terms of the $m$-th metric-subspace pair $\varphi^{\mathcal{X}_m}_{\mu_m,k_m}(\cdot,\cdot)$, the similarity between samples $x_i$ and $x_j$ is computed by first mapping $x_i$ and $x_j$ onto the subspace associated with

the component dataset $\mathcal{X}_m$ and then computing their kernel similarity with the randomly selected parameters $\mu_m$ and $k_m$. Thus, we can obtain $M$ similarity matrices in terms of the $M$ metric-subspace pairs as follows:

$$\mathcal{S} = \{S^{(1)}, S^{(2)}, \cdots, S^{(M)}\}, \tag{8}$$

where the $m$-th similarity matrix (i.e., $S^{(m)}$) is constructed w.r.t. the $m$-th metric-subspace pair $\varphi^{\mathcal{X}_m}_{\mu_m,k_m}(\cdot,\cdot)$, denoted as

$$S^{(m)} = \{S^{(m)}_{ij}\}_{N\times N}, \tag{9}$$

where

$$S^{(m)}_{ij} = \varphi^{\mathcal{X}_m}_{\mu_m,k_m}(x_i, x_j). \tag{10}$$

is the $(i,j)$-th entry in $S^{(m)}$. Obviously, according to the definition of the kernel, it holds that $S^{(m)}_{ij} \in (0,1]$ for any $x_i, x_j \in \mathcal{X}$. If samples $x_i$ and $x_j$ have the same feature values in the subspace associated with $\mathcal{X}_m$, then their similarity $S^{(m)}_{ij}$ reaches its maximum 1.

Having constructed $M$ similarity matrices with diversified metric-subspace pairs, we then exploit the spectral clustering [55] to construct the ensemble of base clusterings. Specifically, for the $m$-th similarity matrix $S^{(m)}$, we treat each data sample as a graph node and build a similarity graph as follows:

$$G^{(m)} = (V, E^{(m)}), \tag{11}$$

where $V = \mathcal{X}$ is the node set, and $E^{(m)}$ is the edge set. The edge weights are decided by the similarity matrix $S^{(m)}$, i.e., for any $x_i, x_j \in \mathcal{X}$, we have $E^{(m)}_{ij} = S^{(m)}_{ij}$. Let $K_m$ denote the number of clusters in the $m$-base clustering. The objective of spectral clustering is to partition the graph $G^{(m)}$ into $K_m$ disjoint subsets. To this end, we construct the normalized graph Laplacian $L_m$ as follows:

$$L_m = I - D_m^{-1/2} S^{(m)} D_m^{-1/2}, \tag{12}$$

where the degree matrix $D_m \in \mathbb{R}^{N\times N}$ is a diagonal matrix with its $(i,i)$-th entry defined as the sum of the $i$-th row of $S^{(m)}$. The eigenvectors corresponding to the first $K_m$ eigenvalues of $L_m$ are computed and then stacked to form a new matrix $U_m \in \mathbb{R}^{N\times K_m}$, where the $i$-th column of $U_m$ is the eigenvector corresponding to the $i$-th eigenvalue of $L_m$. Thereafter, the matrix $T_m \in \mathbb{R}^{N\times K_m}$ can be obtained from $U_m$ by normalizing the rows to norm 1. By treating each row of $T_m$ as a data point in $\mathbb{R}^{K_m}$, we can cluster the rows into $K_m$ clusters by $K$-means discretization, and thereby obtain the $m$-th base clustering based on the similarity matrix $S^{(m)}$. Formally, the $m$-th base clustering is denoted as

$$\pi^m = \{C^m_1, C^m_2, \cdots, C^m_{K_m}\}, \tag{13}$$

where $C^m_i$ is the $i$-th cluster in $\pi^m$. It is obvious that the $K_m$ clusters in a base clustering cover the entire dataset, i.e., $\bigcup_{i=1}^{K_m} C^m_i = \mathcal{X}$, and two clusters in the same base clustering will not overlap with each other, i.e., $\forall i \neq j, C^m_i \bigcap C^m_j = \emptyset$.

Finally, based on the $M$ diversified similarity matrices, we can construct an ensemble of $M$ base clusterings, denoted as

$$\Pi = \{\pi^1, \pi^2, \cdots, \pi^M\}, \tag{14}$$

where $\pi^m$ is the $m$-th base clustering in the ensemble $\Pi$.

### D. Consensus Functions

This paper aims to integrate multiple levels of diversity in a unified ensemble clustering framework for high-dimensional data. With the metric-wise and subspace-wise diversity exploited in the jointly randomized ensemble generation, in this section, we proceed to explore the cluster-wise diversity in ensembles and incorporate three types of consensus functions to combine the multiple base clusterings into the final consensus clustering result.

With each base clustering consisting of a certain number of clusters, the entire ensemble can also be viewed as a large set of clusters from different base clusterings. To exploit the different reliability of different clusters and incorporate the cluster-wise diversity in the consensus function, here we adopt a local weighting strategy [15] to evaluate and weight the base clusters by jointly considering the distribution of cluster labels in the entire ensemble using an entropic criterion. Formally, we denote the ensemble of clusters as

$$\mathcal{C} = \{C_1, C_2, \cdots, C_{N_c}\}, \tag{15}$$

where $C_i$ is the $i$-th cluster and $N_c$ is the total number of clusters in the ensemble $\Pi$. Note that $N_c = \sum_{m=1}^{M} K_m$.

Each cluster is a set of data samples. To estimate the uncertainty of different clusters, the concept of entropy is utilized here [15]. Given a cluster $C_i \in \mathcal{C}$ and a base clustering $\pi^m \in \Pi$, the uncertainty (or entropy) of $C_i$ w.r.t. $\pi^m$ can be computed as

$$H^{\pi^m}(C_i) = -\sum_{j=1}^{K_m} p(C_i, C^m_j) \log_2 p(C_i, C^m_j), \tag{16}$$

where

$$p(C_i, C^m_j) = \frac{|C_i \bigcap C^m_j|}{|C_i|}, \tag{17}$$

is the proportion of data samples in $C_i$ that also appear in $C^m_j$. It is obvious that $p(C_i, C^m_j) \in [0,1]$, which leads to $H^{\pi^m}(C_i) \in [0, +\infty)$. If and only if all data samples in $C_i$ also occur in the same cluster in $\pi^m$, the uncertainty of $C_i$ w.r.t. $\pi^m$ reaches its minimum 0.

With the general assumption that the set of base clusterings are independent of each other, we can obtain the uncertainty (or entropy) of $C_i$ w.r.t. the entire ensemble $\Pi$ as follows

$$H^{\Pi}(C_i) = \sum_{m=1}^{M} H^{\pi^m}(C_i). \tag{18}$$

Intuitively, higher uncertainty indicates lower reliability for a cluster, which implies that the ensemble of base clusterings tend to disagree with the cluster and accordingly a smaller weight can be associated with it [15]. In particular, we proceed to compute a reliability index from the above-mentioned uncertainty measure, and exploit it as a cluster weighting term in our consensus function. The experimental analysis about the efficacy of the cluster weighting term will also be provided in Section IV-E. Specifically, the ensemble-driven cluster index (ECI) is computed as an indication for the reliability of each cluster in the ensemble, which is defined as follows:

$$ECI(C_i) = \exp\left(-\frac{H^{\Pi}(C_i)}{M}\right). \tag{19}$$

Obviously, for any $\pi^m \in \Pi$, it holds that $H^{\pi^m}(C_i) \in [0, +\infty)$, then we have $H^{\Pi}(C_i) \in [0, +\infty)$ and thereby $ECI(C_i) \in (0, 1]$. Note that a larger value of ECI is associated with a cluster of lower uncertainty (i.e., higher reliability). If and only if the data samples in $C_i$ appear in the same cluster in all of the $M$ base clusterings (i.e., all base clusterings agree that the data samples in $C_i$ should belong to the same cluster), the uncertainty of $C_i$ w.r.t. $\Pi$ reaches its minimum 0 and the ECI of $C_i$ reaches its maximum 1.

The ECI measure serves as a reliability index for different clusters in the ensemble $\Pi$. By using ECI as a cluster-weighting term, the locally weighted co-association matrix can be obtained as follow:

$$A = \{A_{ij}\}_{N \times N}, \tag{20}$$

$$A_{ij} = \frac{1}{M} \cdot \sum_{m=1}^{M} w_i^m \cdot \delta_{ij}^m, \tag{21}$$

$$w_i^m = ECI\left(Cls^m(x_i)\right), \tag{22}$$

$$\delta_{ij}^m = \begin{cases} 1, & \text{if } Cls^m(x_i) = Cls^m(x_j), \\ 0, & \text{otherwise,} \end{cases} \tag{23}$$

where $Cls^m(x_i)$ denotes the cluster in $\pi^m$ that $x_i$ belongs to. Note that $w_i^m$ is the cluster weighting term which weights each cluster according to its ECI value, while $\delta_{ij}^m$ is the pair-wise co-occurrence term that indicates whether two samples occur in the same cluster in a base clustering $\pi^m$.

With the cluster-wise diversity explored, we further exploit three types of consensus functions to obtain the consensus clustering result, and thereby present three specific ensemble clustering algorithms, termed MDEC-HC, MDEC-SC, and MDEC-BG, respectively, under our MDEC framework.

*1) MDEC-HC:* In MDEC-HC, a hierarchical clustering based consensus function is incorporated, which constructs a dendrogram (i.e., a hierarchical cluster tree) by iterative region merging. Specifically, given a dataset $\mathcal{X}$, the $N$ data samples in $\mathcal{X}$ are treated as the set of initial regions, denoted as

$$\mathcal{R}^{(0)} = \{R_1^{(0)}, \cdots, R_{|\mathcal{R}^{(0)}|}^{(0)}\}, \tag{24}$$

where $R_i^{(0)} = \{x_i\}$ is the $i$-th initial region and $|\mathcal{R}^(0)| = N$ is the number of initial regions. Here, the locally weighted co-association matrix $A$ in Eq. (20) is used as the similarity matrix for the initial regions, denoted as $Z^{(0)} = A$, upon which the iterative region merging can then be performed. In each iteration, the two regions with the highest similarity will be merged into a new and larger region.

Let $\mathcal{R}^{(t)} = \{R_1^{(t)}, \cdots, R_{|\mathcal{R}^{(t)}|}^{(t)}\}$ denote the set of regions after the $t$-th iteration, whose similarity matrix can be updated according to the average-linkage criterion after the region merging process, that is

$$Z^{(t)} = \{Z_{ij}^{(t)}\}_{|\mathcal{R}^{(t)}| \times |\mathcal{R}^{(t)}|}, \tag{25}$$

$$Z_{ij}^{(t)} = \frac{1}{|R_i^{(t)}| \cdot |R_j^{(t)}|} \sum_{x_u \in R_i^{(t)}, x_v \in R_j^{(t)}} a_{uv}, \tag{26}$$

where $|\mathcal{R}^{(t)}|$ denotes the number of regions in the region set $\mathcal{R}^{(t)}$, and $|R_i^{(t)}|$ denotes the number of data samples in the region $R_i^{(t)}$. Note that, in each iteration, the number of regions decrements by one. Obviously, it holds that $|\mathcal{R}^{(t)}| = N - t$, and the dendrogram will be completed after exactly $N - 1$ iterations. Each level of the dendrogram corresponds to a clustering result with a certain number of clusters, and therefore the final consensus clustering can be obtained by specifying a level in the dendrogram.

*2) MDEC-SC:* In MDEC-SC, a spectral clustering based consensus function is exploited. We first construct a graph with the data samples treated as graph nodes and the locally weighted co-association matrix $A$ used as the adjacent matrix, denoted as:

$$G = (V, E), \tag{27}$$

where $V = \mathcal{X}$ is the node set, and $E$ is the edge set. The edge weight between nodes $x_i$ and $x_j$ is defined as $E_{ij} = A_{ij}$. The normalized graph Laplacian is computed as

$$L = I - D^{-1/2} A D^{-1/2}, \tag{28}$$

where $D \in \mathbb{R}^{N \times N}$ is the degree matrix. Let $K$ be the number of clusters that we aim to obtain. Then we compute the eigenvectors corresponding to the first $K$ eigenvalues of $L$, which are further stacked to form a new matrix $U \in \mathbb{R}^{N \times K}$ with each eigenvector being a column of it. Thereafter, the matrix $U$ will be row-normalized, upon which $K$-means discretization will be performed to obtain the final clustering result.

*3) MDEC-BG:* In MDEC-BG, a bipartite graph based consensus function is presented. Different from MDEC-SC, we build a bipartite graph with both data samples and base clusters treated as graph nodes, and then perform bipartite graph partitioning to obtain the consensus clustering. Specifically, the bipartite graph is defined as

$$\tilde{G} = (U, V, \tilde{E}), \tag{29}$$

where $U \bigcup V$ is the node set (with $U = \mathcal{X}$ and $V = \mathcal{C}$), and $\tilde{E}$ is the edge set. An edge between two nodes exists if and only if one of them is a data sample and the other one is a base cluster that contains this data sample. Typically, given two nodes $u_i \in \mathcal{X}$ and $v_j \in \mathcal{C}$, their edge weight will be decided by two factors, i.e., their belonging-to relationship and the reliability of the connected cluster, which can be estimated by the ECI index. Formally, the edge weight between $u_i \in \mathcal{X}$ and $v_j \in \mathcal{C}$ is defined as

$$\tilde{E}_{ij} = \begin{cases} ECI(v_j), & \text{if } u_i \in v_j, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

Note that the ECI value reflects the reliability of a cluster w.r.t. the entire ensemble of base clusterings (as shown in Eqs. (16), (17), (18), and (19)). By taking advantage of ECI to define the edge weights in the bipartite graph $\tilde{G}$, the clusters with higher reliability can exhibit greater influence on the bipartite graph and thus play a bigger role in the consensus process. Then, with the bipartite graph constructed, we further adopt the transfer cut [56] to efficiently partition the graph nodes into several subsets. By treating the data samples in each subset as a final cluster, the consensus clustering result can be obtained.

## IV. Experiments

In this section, we conduct experiments on a variety of real-world high-dimensional datasets to compare the proposed three ensemble clustering algorithms. against nine state-of-the-art algorithms.

### A. Datasets and Evaluation Measures

In our experiments, 30 real-world high-dimensional datasets are used, including 18 cancer gene expression datasets [57] and 12 image/speech datasets. Among the 12 speech/image datasets, ISOLET [58] is a speech recognition dataset, while the other 11 datasets, including COIL20 [59], MSRA25 [60], Binary Alphadigits [61], YaleB32 [60], Semeion [58], UMist [62], Multiple Features [58], Flowers17 [63], MNIST [61], USPS [61], and Gisette [58], are image datasets. To simplify the description, in the following, the 30 benchmark datasets will be abbreviated as *DS-1* to *DS-30*, respectively (as shown in Tables I and II).

To evaluate the quality of the clustering result, two widely used evaluation measures are adopted, namely, normalized mutual information (NMI) [39] and adjusted Rand index (ARI) [64]. Note that greater values of NMI and ARI indicate better clustering results.

TABLE I
Description of the eighteen cancer gene expression datasets.

| Dataset | Abbr. | #Sample | Dimension | #Class |
|---|---|---|---|---|
| Bhattacharjee-2001 | *DS-1* | 203 | 1,543 | 5 |
| Bredel-2005 | *DS-2* | 50 | 1,739 | 3 |
| Chowdary-2006 | *DS-3* | 104 | 182 | 2 |
| Garber-2001 | *DS-4* | 66 | 4,553 | 4 |
| Golub-1999-v1 | *DS-5* | 72 | 1,877 | 2 |
| Golub-1999-v2 | *DS-6* | 72 | 1,877 | 3 |
| Nutt-2003-v1 | *DS-7* | 50 | 1,377 | 4 |
| Pomeroy-2002-v2 | *DS-8* | 42 | 1,379 | 5 |
| Su-2001 | *DS-9* | 174 | 1,571 | 10 |
| West-2001 | *DS-10* | 49 | 1,198 | 2 |
| Yeoh-2002-v1 | *DS-11* | 248 | 2,526 | 2 |
| Yeoh-2002-v2 | *DS-12* | 248 | 2,526 | 6 |
| Alizadeh-2000-v2 | *DS-13* | 62 | 2,093 | 3 |
| Alizadeh-2000-v3 | *DS-14* | 62 | 2,093 | 4 |
| Armstrong-2002-v2 | *DS-15* | 72 | 2,194 | 3 |
| Ramaswamy-2001 | *DS-16* | 190 | 1,363 | 14 |
| Risinger-2003 | *DS-17* | 42 | 1,771 | 4 |
| Tomlins-2006-v1 | *DS-18* | 104 | 2,315 | 5 |

TABLE II
Description of the twelve image/speech datasets.

| Dataset | Abbr. | #Sample | Dimension | #Class |
|---|---|---|---|---|
| ISOLET | *DS-19* | 7,797 | 617 | 26 |
| COIL20 | *DS-20* | 1,440 | 1,024 | 20 |
| MSRA25 | *DS-21* | 1,799 | 256 | 12 |
| Binary Alphadigits | *DS-22* | 1,404 | 320 | 36 |
| YaleB32 | *DS-23* | 2,414 | 1,024 | 38 |
| Semeion | *DS-24* | 1,593 | 256 | 10 |
| UMist | *DS-25* | 575 | 10,304 | 20 |
| Multiple Features | *DS-26* | 2,000 | 649 | 10 |
| Flowers17 | *DS-27* | 1,360 | 30,000 | 17 |
| MNIST | *DS-28* | 5,000 | 784 | 10 |
| Gisette | *DS-29* | 7,000 | 5,000 | 2 |
| USPS | *DS-30* | 11,000 | 256 | 10 |

### B. Baseline Methods and Experimental Settings

In our experiments, we compare the proposed ensemble clustering algorithms with nine baseline algorithms, i.e., stratified sampling based cluster-based similarity partitioning algorithm (SSCSPA) [6], stratified sampling based hypergraph partitioning algorithm (SSHGPA) [6], stratified sampling based meta-clustering algorithm (SSMCLA) [6], $K$-means based consensus clustering (KCC) [7], probability trajectory based graph partitioning (PTGP) [8], spectral ensemble clustering (SEC) [11], entropy based consensus clustering (ECC) [12], locally weighted graph partitioning (LWGP) [15], and self-paced clustering ensemble (SPCE) [65].

For the baseline algorithms, the parameters will be set as suggested by their corresponding papers. For the proposed algorithms, to produce a set of diversified metrics, the two kernel parameters $\mu$ and $k$ are randomized in the ranges of $[0.2, 0.8]$ and $[5, 20]$, respectively. To generate the ensemble of base clusterings, the ensemble size $M = 100$ and the sampling ratio $\tau = 0.5$ are used. The number of clusters in each base clustering is randomly selected in the range of $[2, \sqrt{N}]$. Further, the performances of our algorithms w.r.t. different ensemble sizes $M$ and different sampling ratios $\tau$ will also be evaluated in Sections IV-D and IV-E, respectively.

### C. Comparison with Other Ensemble Clustering Methods

In this section, we evaluate the proposed ensemble clustering algorithms, namely, MDEC-HC, MDEC-SC, and MDEC-BG, against nine baseline ensemble clustering algorithms. For each test algorithm, the number of clusters is set to the true number of classes on the dataset, which is a commonly-adopted experimental protocol in the literature [11], [12]. If an algorithm is not computationally feasible on a dataset due to the out-of-memory error, its score on this dataset will be labeled as "OM". For each dataset, we run every test algorithm 100 times, and report their average NMI and ARI scores in Tables III and IV, respectively.

In terms of NMI, as shown in Table III, the proposed algorithms outperform the baseline algorithms on most of the benchmark datasets. The SPCE achieves the best NMI scores on 5 out of the 30 datasets, but the three proposed algorithms outperform SPCE on most of the other datasets. It is noteworthy that the proposed MDEC-HC algorithm achieves the best NMI scores on 16 out of the 30 datasets. Also, the proposed MDEC-HC, MDEC-SC, and MDEC-BG algorithms rank in the top three positions 26, 27, and 25 times, respectively, out of the total of 30 datasets, while the fourth best algorithm (i.e., SPCE) ranks in the top three positions just 7 times. In terms of ARI, as shown in Table III, our three algorithms also exhibit clear advantages over the baselines, ranking in the top three positions 28, 25, and 25 times, respectively, while the fourth best algorithm only ranks in the top three positions 4 times.

Further, to provide a summary view across the 30 benchmark datasets, we report the *average score* and *average rank* of different algorithms in the last two rows in Tables III and IV.

TABLE III

AVERAGE PERFORMANCES (W.R.T. NMI) OVER 100 RUNS BY DIFFERENT ENSEMBLE CLUSTERING ALGORITHMS. ON EACH DATASET, THE HIGHEST THREE SCORES ARE HIGHLIGHTED IN **BOLD**, WHILE THE HIGHEST ONE IN [**BOLD AND BRACKETS**].

| Dataset | SSCSPA | SSHGPA | SSMCLA | KCC | PTGP | SEC | ECC | LWGP | SPCE | MDEC-HC | MDEC-SC | MDEC-BG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DS-1* | $20.63_{\pm 2.11}$ | $19.06_{\pm 2.15}$ | $26.71_{\pm 2.95}$ | $26.88_{\pm 6.93}$ | $39.86_{\pm 1.55}$ | $27.56_{\pm 3.04}$ | $35.90_{\pm 1.23}$ | $42.10_{\pm 4.82}$ | [$\mathbf{53.53}_{\pm 2.27}$] | $\mathbf{52.41}_{\pm 1.68}$ | $\mathbf{51.72}_{\pm 1.25}$ | $51.64_{\pm 1.11}$ |
| *DS-2* | $\mathbf{40.53}_{\pm 2.79}$ | $32.54_{\pm 7.33}$ | $33.88_{\pm 2.94}$ | $36.12_{\pm 2.79}$ | $31.50_{\pm 3.62}$ | $36.01_{\pm 3.67}$ | $36.52_{\pm 5.94}$ | $31.56_{\pm 3.28}$ | $\mathbf{40.64}_{\pm 5.27}$ | [$\mathbf{45.23}_{\pm 3.35}$] | $38.44_{\pm 3.19}$ | $38.58_{\pm 3.02}$ |
| *DS-3* | $60.26_{\pm 4.47}$ | $73.58_{\pm 4.94}$ | $47.99_{\pm 17.36}$ | $60.13_{\pm 3.26}$ | $8.15_{\pm 0.00}$ | $50.27_{\pm 11.41}$ | $61.18_{\pm 1.25}$ | $8.15_{\pm 0.00}$ | $23.83_{\pm 18.80}$ | [$\mathbf{85.97}_{\pm 0.00}$] | [$\mathbf{85.97}_{\pm 0.00}$] | [$\mathbf{85.97}_{\pm 0.00}$] |
| *DS-4* | $15.02_{\pm 0.84}$ | $17.03_{\pm 1.87}$ | $14.49_{\pm 2.40}$ | $15.34_{\pm 4.08}$ | $14.20_{\pm 2.49}$ | $13.57_{\pm 2.67}$ | $14.89_{\pm 3.15}$ | $12.44_{\pm 2.59}$ | [$\mathbf{27.76}_{\pm 1.53}$] | $21.77_{\pm 0.82}$ | $\mathbf{22.38}_{\pm 4.05}$ | $\mathbf{22.86}_{\pm 0.79}$ |
| *DS-5* | $42.42_{\pm 5.77}$ | $46.91_{\pm 6.64}$ | $62.53_{\pm 9.31}$ | $55.65_{\pm 3.73}$ | $57.11_{\pm 6.67}$ | $49.37_{\pm 6.05}$ | $53.46_{\pm 2.11}$ | $56.45_{\pm 2.69}$ | $51.02_{\pm 0.60}$ | [$\mathbf{74.92}_{\pm 13.84}$] | $\mathbf{73.75}_{\pm 7.77}$ | $\mathbf{71.79}_{\pm 11.81}$ |
| *DS-6* | $48.35_{\pm 3.59}$ | $51.73_{\pm 7.02}$ | $42.31_{\pm 15.19}$ | $53.68_{\pm 8.42}$ | $67.32_{\pm 5.32}$ | $59.31_{\pm 6.68}$ | $67.51_{\pm 4.18}$ | $65.98_{\pm 3.90}$ | $45.22_{\pm 2.62}$ | [$\mathbf{76.17}_{\pm 5.42}$] | $\mathbf{75.14}_{\pm 3.33}$ | $\mathbf{76.04}_{\pm 2.86}$ |
| *DS-7* | $40.61_{\pm 3.50}$ | $37.22_{\pm 6.00}$ | $36.71_{\pm 3.90}$ | $46.31_{\pm 3.48}$ | $\mathbf{48.40}_{\pm 0.98}$ | $46.53_{\pm 2.67}$ | $43.21_{\pm 1.58}$ | $44.79_{\pm 0.60}$ | [$\mathbf{49.78}_{\pm 2.27}$] | $\mathbf{48.26}_{\pm 0.82}$ | $45.50_{\pm 3.89}$ | $47.36_{\pm 1.04}$ |
| *DS-8* | $50.30_{\pm 4.17}$ | $56.29_{\pm 5.27}$ | $40.85_{\pm 12.12}$ | $57.55_{\pm 5.32}$ | $58.67_{\pm 2.73}$ | $52.92_{\pm 8.03}$ | $56.58_{\pm 3.09}$ | $55.53_{\pm 2.78}$ | $63.20_{\pm 2.72}$ | [$\mathbf{66.16}_{\pm 2.44}$] | $\mathbf{65.53}_{\pm 2.77}$ | $\mathbf{64.80}_{\pm 1.14}$ |
| *DS-9* | $52.05_{\pm 1.40}$ | $52.72_{\pm 3.49}$ | $44.66_{\pm 10.09}$ | $57.22_{\pm 2.78}$ | $57.02_{\pm 3.24}$ | $56.91_{\pm 4.53}$ | $56.95_{\pm 2.76}$ | $53.15_{\pm 2.26}$ | $58.26_{\pm 7.31}$ | $\mathbf{66.57}_{\pm 2.27}$ | [$\mathbf{69.55}_{\pm 0.93}$] | $\mathbf{69.41}_{\pm 1.32}$ |
| *DS-10* | $29.05_{\pm 4.68}$ | [$\mathbf{35.75}_{\pm 1.25}$] | $29.21_{\pm 4.49}$ | $31.17_{\pm 3.09}$ | $31.18_{\pm 1.33}$ | $28.29_{\pm 4.29}$ | $31.22_{\pm 0.00}$ | $4.08_{\pm 0.00}$ | $21.37_{\pm 5.62}$ | $32.03_{\pm 3.92}$ | $\mathbf{32.45}_{\pm 2.21}$ | $\mathbf{34.29}_{\pm 2.37}$ |
| *DS-11* | $20.20_{\pm 0.30}$ | $18.60_{\pm 1.37}$ | $4.56_{\pm 4.57}$ | $79.18_{\pm 11.23}$ | $84.56_{\pm 2.01}$ | $56.61_{\pm 35.73}$ | $85.59_{\pm 6.47}$ | $1.92_{\pm 10.32}$ | $41.78_{\pm 22.53}$ | [$\mathbf{94.35}_{\pm 2.49}$] | $\mathbf{92.91}_{\pm 2.50}$ | $\mathbf{93.64}_{\pm 2.13}$ |
| *DS-12* | $31.48_{\pm 2.51}$ | $32.65_{\pm 4.85}$ | $9.42_{\pm 5.32}$ | $33.40_{\pm 1.31}$ | $34.93_{\pm 4.53}$ | $32.85_{\pm 2.87}$ | $36.09_{\pm 3.73}$ | $26.44_{\pm 1.05}$ | $41.48_{\pm 1.17}$ | $\mathbf{45.98}_{\pm 5.39}$ | $\mathbf{50.81}_{\pm 5.34}$ | [$\mathbf{57.01}_{\pm 3.79}$] |
| *DS-13* | $57.11_{\pm 1.02}$ | $48.88_{\pm 3.43}$ | $56.71_{\pm 2.28}$ | $65.40_{\pm 18.11}$ | $\mathbf{97.56}_{\pm 7.89}$ | $59.52_{\pm 13.51}$ | $70.32_{\pm 20.54}$ | $71.92_{\pm 3.87}$ | $91.10_{\pm 9.48}$ | [$\mathbf{98.34}_{\pm 3.44}$] | $\mathbf{94.86}_{\pm 4.31}$ | $93.38_{\pm 3.74}$ |
| *DS-14* | $42.77_{\pm 3.38}$ | $49.40_{\pm 2.65}$ | $49.02_{\pm 3.82}$ | $52.26_{\pm 4.44}$ | $\mathbf{63.97}_{\pm 2.07}$ | $50.11_{\pm 9.60}$ | $63.47_{\pm 2.07}$ | $53.01_{\pm 4.60}$ | [$\mathbf{65.99}_{\pm 2.49}$] | $\mathbf{65.28}_{\pm 1.46}$ | $63.61_{\pm 1.63}$ | $63.02_{\pm 1.42}$ |
| *DS-15* | $66.82_{\pm 2.16}$ | $49.13_{\pm 6.83}$ | $66.67_{\pm 3.09}$ | $50.34_{\pm 5.45}$ | $75.00_{\pm 4.99}$ | $41.77_{\pm 7.09}$ | $72.08_{\pm 5.32}$ | $58.02_{\pm 9.73}$ | $61.86_{\pm 8.91}$ | [$\mathbf{80.90}_{\pm 2.71}$] | $\mathbf{80.42}_{\pm 2.11}$ | $\mathbf{78.56}_{\pm 2.16}$ |
| *DS-16* | $50.44_{\pm 1.65}$ | $54.22_{\pm 1.67}$ | $18.71_{\pm 4.94}$ | $45.13_{\pm 2.79}$ | $38.42_{\pm 2.16}$ | $38.05_{\pm 5.40}$ | $43.57_{\pm 4.85}$ | $29.19_{\pm 0.86}$ | $52.51_{\pm 1.15}$ | $\mathbf{65.56}_{\pm 1.53}$ | $\mathbf{66.60}_{\pm 1.18}$ | [$\mathbf{66.64}_{\pm 1.16}$] |
| *DS-17* | $30.98_{\pm 5.61}$ | $30.90_{\pm 4.37}$ | $16.20_{\pm 9.87}$ | $29.09_{\pm 3.41}$ | $31.32_{\pm 2.60}$ | $28.12_{\pm 3.68}$ | $31.93_{\pm 3.47}$ | $15.57_{\pm 0.79}$ | $32.05_{\pm 1.04}$ | [$\mathbf{36.79}_{\pm 2.93}$] | $\mathbf{32.77}_{\pm 1.44}$ | $\mathbf{33.13}_{\pm 1.17}$ |
| *DS-18* | $38.12_{\pm 1.90}$ | $35.95_{\pm 3.24}$ | $42.68_{\pm 3.19}$ | $44.86_{\pm 5.88}$ | $44.66_{\pm 1.78}$ | $37.92_{\pm 5.01}$ | $40.50_{\pm 1.83}$ | $42.61_{\pm 2.06}$ | $46.57_{\pm 2.01}$ | [$\mathbf{61.11}_{\pm 2.59}$] | $\mathbf{53.30}_{\pm 2.24}$ | $\mathbf{54.30}_{\pm 2.04}$ |
| *DS-19* | $69.12_{\pm 1.52}$ | $60.32_{\pm 1.56}$ | $73.81_{\pm 1.00}$ | $68.43_{\pm 1.04}$ | $72.96_{\pm 0.70}$ | $68.71_{\pm 1.25}$ | $70.06_{\pm 0.88}$ | $74.91_{\pm 0.53}$ | $72.35_{\pm 1.86}$ | [$\mathbf{77.09}_{\pm 0.77}$] | $\mathbf{76.41}_{\pm 0.29}$ | $\mathbf{76.48}_{\pm 0.50}$ |
| *DS-20* | $78.42_{\pm 1.17}$ | $76.77_{\pm 2.20}$ | $75.38_{\pm 4.97}$ | $72.33_{\pm 1.98}$ | $67.00_{\pm 2.65}$ | $72.06_{\pm 3.01}$ | $74.26_{\pm 1.37}$ | $80.16_{\pm 1.10}$ | $86.01_{\pm 1.28}$ | [$\mathbf{91.66}_{\pm 0.75}$] | $\mathbf{90.15}_{\pm 1.06}$ | $\mathbf{91.25}_{\pm 0.78}$ |
| *DS-21* | $59.41_{\pm 1.50}$ | $61.66_{\pm 2.84}$ | $65.07_{\pm 2.07}$ | $61.05_{\pm 1.61}$ | $63.67_{\pm 1.97}$ | $59.79_{\pm 3.59}$ | $62.70_{\pm 2.39}$ | $66.29_{\pm 1.59}$ | $69.31_{\pm 1.56}$ | $\mathbf{70.37}_{\pm 1.05}$ | [$\mathbf{71.78}_{\pm 1.64}$] | $\mathbf{71.03}_{\pm 1.28}$ |
| *DS-22* | $55.31_{\pm 0.97}$ | $46.99_{\pm 1.35}$ | $54.45_{\pm 1.35}$ | $54.21_{\pm 0.73}$ | $55.06_{\pm 0.57}$ | $56.41_{\pm 0.60}$ | $54.79_{\pm 0.59}$ | $51.67_{\pm 0.87}$ | $\mathbf{60.57}_{\pm 0.65}$ | $58.06_{\pm 1.31}$ | [$\mathbf{62.54}_{\pm 0.62}$] | $\mathbf{62.33}_{\pm 0.62}$ |
| *DS-23* | $10.63_{\pm 0.31}$ | $11.86_{\pm 0.52}$ | $11.10_{\pm 0.50}$ | $10.11_{\pm 0.40}$ | $10.15_{\pm 0.40}$ | $10.39_{\pm 0.45}$ | $10.47_{\pm 0.34}$ | $9.40_{\pm 0.33}$ | $27.29_{\pm 1.23}$ | $\mathbf{34.12}_{\pm 1.36}$ | $\mathbf{35.40}_{\pm 0.98}$ | [$\mathbf{37.54}_{\pm 1.25}$] |
| *DS-24* | $55.69_{\pm 3.29}$ | $35.82_{\pm 2.07}$ | $57.51_{\pm 2.33}$ | $53.22_{\pm 1.77}$ | $62.64_{\pm 0.69}$ | $54.39_{\pm 3.28}$ | $53.61_{\pm 1.74}$ | $61.93_{\pm 3.57}$ | $62.27_{\pm 3.03}$ | [$\mathbf{71.33}_{\pm 0.64}$] | $\mathbf{67.85}_{\pm 1.41}$ | $\mathbf{68.51}_{\pm 1.25}$ |
| *DS-25* | $57.93_{\pm 1.22}$ | $65.52_{\pm 2.53}$ | $58.69_{\pm 3.61}$ | $60.85_{\pm 1.74}$ | $62.67_{\pm 1.11}$ | $60.49_{\pm 1.39}$ | $61.37_{\pm 1.20}$ | $62.50_{\pm 1.53}$ | $69.84_{\pm 1.34}$ | [$\mathbf{83.97}_{\pm 1.77}$] | $\mathbf{82.24}_{\pm 2.73}$ | $\mathbf{83.15}_{\pm 2.45}$ |
| *DS-26* | $81.34_{\pm 3.10}$ | $69.94_{\pm 3.07}$ | $86.33_{\pm 2.11}$ | $73.08_{\pm 3.97}$ | $83.45_{\pm 2.14}$ | $64.14_{\pm 6.06}$ | $75.24_{\pm 1.57}$ | $87.02_{\pm 1.64}$ | $82.00_{\pm 4.02}$ | $\mathbf{90.68}_{\pm 2.09}$ | [$\mathbf{94.19}_{\pm 0.30}$] | $\mathbf{93.99}_{\pm 0.26}$ |
| *DS-27* | $21.75_{\pm 0.75}$ | $18.75_{\pm 1.24}$ | $23.74_{\pm 0.59}$ | $24.15_{\pm 0.47}$ | $24.96_{\pm 0.39}$ | $24.18_{\pm 0.49}$ | $24.16_{\pm 0.39}$ | $21.60_{\pm 0.75}$ | [$\mathbf{37.09}_{\pm 0.38}$] | $25.44_{\pm 0.64}$ | $\mathbf{27.56}_{\pm 0.39}$ | $\mathbf{27.14}_{\pm 0.38}$ |
| *DS-28* | $53.13_{\pm 3.25}$ | $38.54_{\pm 2.66}$ | $58.34_{\pm 2.20}$ | $53.46_{\pm 3.18}$ | $64.53_{\pm 1.64}$ | $53.10_{\pm 2.26}$ | $53.42_{\pm 2.23}$ | $60.13_{\pm 1.22}$ | $51.77_{\pm 8.95}$ | [$\mathbf{78.47}_{\pm 2.19}$] | $\mathbf{76.04}_{\pm 1.20}$ | $\mathbf{77.92}_{\pm 0.96}$ |
| *DS-29* | $20.44_{\pm 6.59}$ | $19.28_{\pm 10.95}$ | $47.98_{\pm 4.69}$ | $41.41_{\pm 9.87}$ | $49.82_{\pm 0.87}$ | $10.73_{\pm 6.22}$ | $46.07_{\pm 2.01}$ | $5.62_{\pm 14.21}$ | $9.23_{\pm 0.54}$ | $\mathbf{67.23}_{\pm 5.28}$ | $\mathbf{67.43}_{\pm 0.68}$ | [$\mathbf{67.94}_{\pm 0.34}$] |
| *DS-30* | $47.98_{\pm 2.88}$ | $34.33_{\pm 2.08}$ | $53.71_{\pm 2.89}$ | $51.17_{\pm 1.52}$ | $65.00_{\pm 1.84}$ | $47.13_{\pm 2.49}$ | $52.74_{\pm 1.77}$ | $59.30_{\pm 1.41}$ | OM | [$\mathbf{77.58}_{\pm 1.37}$] | $\mathbf{74.91}_{\pm 1.26}$ | $\mathbf{75.33}_{\pm 2.27}$ |
| Avg. score | 44.94 | 42.74 | 43.65 | 48.77 | 52.19 | 44.91 | 51.33 | 43.78 | - | **64.79** | **64.07** | **64.50** |
| Avg. rank | 8.80 | 9.13 | 8.73 | 8.00 | 6.13 | 9.13 | 7.00 | 8.40 | 5.53 | **2.00** | **2.63** | **2.37** |

* Note that OM indicates the out-of-memory error.

Note that the average score (across 30 datasets) is computed by taking the average on the NMI (or ARI) scores, while the average rank is obtained by taking the average on the ranking positions, for each algorithm across all the benchmark datasets. Note that if an algorithm is not computationally feasible on all datasets, it will not have an average score, but will still have an average rank, where the *lost* score on a dataset (due to the out-of-memory error) will be considered as the last position on this dataset. As can be seen in Table III, the proposed three algorithms achieve average NMI(%) scores of $64.79$, $64.07$, and $64.50$, respectively, which are significantly higher than the fourth best average score of $52.19$. In terms of the ranking positions in Table III, the proposed three algorithms obtain average ranks of $2.00$, $2.63$, and $2.37$, respectively, while the fourth best algorithm only obtains an average rank of $5.53$. Similar advantages can also be observed in terms of ARI. The average ARI scores and the average ranks (across the 30 datasets) of the proposed algorithms significantly outperform the nine baseline algorithms (as shown in Table IV).

### *D. Robustness to Ensemble Sizes*

In this section, we compare the performances of different ensemble clustering algorithms with varying ensemble sizes. Specifically, we perform the test algorithms on the datasets with the ensemble size varying from 20 to 300, and report their average NMI and ARI scores in Figs. 2 and 3, respectively.

In terms of NMI, as can be seen in Fig. 2, the proposed algorithms yield stably high performances across the 30 benchmark datasets, with varying ensemble sizes. Although the SPCE algorithm outperforms our algorithms on the *DS-7* and *DS-27* datasets, yet on most of the other datasets our algorithms achieve better or comparable performance when compared to the baseline ensemble clustering algorithms. Especially, on the *DS-3*, *DS-4*, *DS-5*, *DS-6*, *DS-9*, *DS-16*, *DS-18*, *DS-20*, *DS-23*, *DS-24*, *DS-25*, *DS-26*, *DS-28*, *DS-29*, and *DS-30* datasets, our three proposed algorithms have shown clear advantages over the baseline algorithms as the ensemble size goes from 20 to 300. Similarly, in terms of ARI, our algorithms also exhibit very competitive performances on most of the benchmark datasets with varying ensemble sizes (as can be seen in Fig. 3).

Further, in Fig. 4, we illustrates the average NMI and ARI curves (across the 30 benchmark datasets) by different ensemble clustering algorithms with varying ensemble sizes. Specifically, Fig. 4(a) is obtained by taking the average of the 30 sub-figures in Fig. 2, while Fig. 4(b) by taking the average of the 30 sub-figures in Fig. 3. As can be observed in Fig. 4, the proposed algorithms achieve significantly better performances (w.r.t. both NMI and ARI) than the baseline ensemble clustering algorithms across the 30 benchmark datasets. Even when compared to the best two baseline algorithms, i.e., PTGP and ECC, the proposed three algorithms can still consistently achieve approximately 20% higher average NMI/ARI scores.

### *E. Influence of Metrics, Subspaces, and Clusters*

In this paper, we jointly exploit large populations of diversified metrics, random subspaces, and weighted clusters in

TABLE IV
AVERAGE PERFORMANCES (W.R.T. ARI) OVER 100 RUNS BY DIFFERENT ENSEMBLE CLUSTERING ALGORITHMS. ON EACH DATASET, THE HIGHEST THREE SCORES ARE HIGHLIGHTED IN **BOLD**, WHILE THE HIGHEST ONE IN [**BOLD AND BRACKETS**].

| Dataset | SSCSPA | SSHGPA | SSMCLA | KCC | PTGP | SEC | ECC | LWGP | SPCE | MDEC-HC | MDEC-SC | MDEC-BG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DS-1* | $8.25_{\pm 1.68}$ | $7.01_{\pm 1.67}$ | $12.26_{\pm 3.33}$ | $13.74_{\pm 7.37}$ | $23.96_{\pm 1.57}$ | $12.35_{\pm 3.34}$ | $22.08_{\pm 1.33}$ | $30.83_{\pm 10.37}$ | [**51.88**$_{\pm 4.14}$] | **50.42**$_{\pm 2.51}$ | **46.94**$_{\pm 2.15}$ | $46.71_{\pm 1.81}$ |
| *DS-2* | $36.46_{\pm 2.66}$ | $29.64_{\pm 6.81}$ | $30.95_{\pm 3.64}$ | $35.63_{\pm 2.62}$ | $33.18_{\pm 6.35}$ | $33.71_{\pm 5.18}$ | $35.52_{\pm 4.90}$ | $22.98_{\pm 7.25}$ | **47.03**$_{\pm 9.54}$ | [**53.89**$_{\pm 4.35}$] | $46.82_{\pm 2.67}$ | **46.99**$_{\pm 2.42}$ |
| *DS-3* | $65.20_{\pm 3.14}$ | $80.99_{\pm 3.44}$ | $52.52_{\pm 20.09}$ | $69.40_{\pm 4.03}$ | $6.57_{\pm 0.00}$ | $58.71_{\pm 13.38}$ | $71.06_{\pm 1.35}$ | $6.57_{\pm 0.00}$ | $29.80_{\pm 28.79}$ | [**92.38**$_{\pm 0.00}$] | [**92.38**$_{\pm 0.00}$] | [**92.38**$_{\pm 0.00}$] |
| *DS-4* | $8.97_{\pm 0.50}$ | $9.38_{\pm 1.89}$ | **17.26**$_{\pm 6.14}$ | $14.44_{\pm 8.72}$ | $8.85_{\pm 2.55}$ | $11.18_{\pm 7.49}$ | $14.64_{\pm 6.83}$ | $15.99_{\pm 3.22}$ | [**21.94**$_{\pm 4.88}$] | **18.42**$_{\pm 0.95}$ | $16.89_{\pm 3.03}$ | $16.98_{\pm 0.78}$ |
| *DS-5* | $44.73_{\pm 4.34}$ | $45.15_{\pm 10.61}$ | $73.83_{\pm 8.22}$ | $64.01_{\pm 4.32}$ | $65.53_{\pm 6.71}$ | $57.42_{\pm 5.94}$ | $61.72_{\pm 2.50}$ | $65.00_{\pm 2.85}$ | $59.14_{\pm 0.65}$ | **80.99**$_{\pm 16.82}$ | [**82.41**$_{\pm 9.41}$] | **78.53**$_{\pm 15.58}$ |
| *DS-6* | $46.12_{\pm 3.25}$ | $49.59_{\pm 10.14}$ | $39.78_{\pm 18.56}$ | $51.89_{\pm 14.45}$ | $68.29_{\pm 6.54}$ | $60.02_{\pm 9.10}$ | $72.63_{\pm 5.64}$ | $65.13_{\pm 3.73}$ | $56.21_{\pm 5.55}$ | **79.45**$_{\pm 6.05}$ | **80.14**$_{\pm 3.37}$ | [**80.92**$_{\pm 3.17}$] |
| *DS-7* | $24.54_{\pm 2.60}$ | $23.19_{\pm 5.12}$ | $20.48_{\pm 5.47}$ | $34.18_{\pm 4.19}$ | **36.43**$_{\pm 1.79}$ | $35.02_{\pm 3.46}$ | $26.20_{\pm 2.62}$ | $35.56_{\pm 0.59}$ | $32.87_{\pm 0.69}$ | [**39.86**$_{\pm 1.24}$] | $34.12_{\pm 5.93}$ | **38.57**$_{\pm 1.57}$ |
| *DS-8* | $36.14_{\pm 6.23}$ | $44.96_{\pm 7.56}$ | $24.64_{\pm 13.25}$ | $47.44_{\pm 6.69}$ | $47.29_{\pm 4.63}$ | $42.17_{\pm 10.56}$ | $46.25_{\pm 4.74}$ | $45.62_{\pm 3.86}$ | $50.18_{\pm 3.11}$ | [**58.47**$_{\pm 2.85}$] | **57.68**$_{\pm 2.86}$ | **57.04**$_{\pm 1.31}$ |
| *DS-9* | $36.17_{\pm 1.94}$ | $35.40_{\pm 4.20}$ | $28.27_{\pm 11.94}$ | $42.58_{\pm 3.68}$ | $41.44_{\pm 4.38}$ | $42.03_{\pm 5.91}$ | $40.05_{\pm 4.08}$ | $38.63_{\pm 2.83}$ | $38.14_{\pm 8.43}$ | **52.65**$_{\pm 2.57}$ | **55.98**$_{\pm 1.60}$ | [**56.48**$_{\pm 2.35}$] |
| *DS-10* | $35.96_{\pm 5.65}$ | [**43.82**$_{\pm 1.40}$] | $30.87_{\pm 5.69}$ | $38.30_{\pm 3.28}$ | $38.67_{\pm 1.66}$ | $33.22_{\pm 6.87}$ | $38.75_{\pm 0.00}$ | $0.01_{\pm 0.00}$ | $25.41_{\pm 10.12}$ | $39.53_{\pm 4.62}$ | **40.12**$_{\pm 2.49}$ | **42.19**$_{\pm 2.65}$ |
| *DS-11* | $12.36_{\pm 0.14}$ | $8.65_{\pm 3.13}$ | $13.12_{\pm 8.07}$ | $88.94_{\pm 14.60}$ | $93.09_{\pm 1.10}$ | $56.16_{\pm 49.17}$ | $93.54_{\pm 3.52}$ | $0.22_{\pm 12.44}$ | $53.28_{\pm 35.18}$ | [**97.90**$_{\pm 1.04}$] | **97.27**$_{\pm 1.13}$ | **97.60**$_{\pm 0.94}$ |
| *DS-12* | $23.12_{\pm 2.21}$ | $24.41_{\pm 4.01}$ | $6.17_{\pm 5.31}$ | $18.46_{\pm 2.35}$ | $18.15_{\pm 5.31}$ | $15.03_{\pm 3.78}$ | $20.46_{\pm 3.96}$ | $17.53_{\pm 0.64}$ | $17.68_{\pm 1.57}$ | **30.83**$_{\pm 6.96}$ | **32.85**$_{\pm 5.88}$ | [**41.49**$_{\pm 3.12}$] |
| *DS-13* | $43.00_{\pm 0.74}$ | $39.27_{\pm 2.45}$ | $43.39_{\pm 2.01}$ | $55.44_{\pm 23.52}$ | **97.68**$_{\pm 9.43}$ | $48.03_{\pm 17.25}$ | $61.88_{\pm 26.89}$ | $76.94_{\pm 9.01}$ | $94.16_{\pm 6.26}$ | [**99.00**$_{\pm 2.08}$] | **96.88**$_{\pm 2.61}$ | $95.98_{\pm 2.27}$ |
| *DS-14* | $26.77_{\pm 4.98}$ | $33.34_{\pm 3.38}$ | $35.32_{\pm 3.29}$ | $40.71_{\pm 7.03}$ | $41.90_{\pm 2.85}$ | $31.78_{\pm 7.48}$ | $41.49_{\pm 1.70}$ | $42.84_{\pm 3.49}$ | [**50.15**$_{\pm 1.88}$] | **45.52**$_{\pm 1.30}$ | **43.09**$_{\pm 1.23}$ | $42.87_{\pm 1.22}$ |
| *DS-15* | $68.50_{\pm 2.50}$ | $48.65_{\pm 7.02}$ | $65.12_{\pm 3.81}$ | $46.04_{\pm 6.69}$ | $76.81_{\pm 6.68}$ | $34.81_{\pm 8.69}$ | $72.54_{\pm 7.32}$ | $56.48_{\pm 10.58}$ | $65.09_{\pm 12.86}$ | [**84.57**$_{\pm 2.99}$] | **84.10**$_{\pm 2.24}$ | **82.13**$_{\pm 2.33}$ |
| *DS-16* | $28.99_{\pm 1.67}$ | $36.03_{\pm 1.93}$ | $7.45_{\pm 3.67}$ | $17.68_{\pm 4.70}$ | $6.17_{\pm 1.50}$ | $10.47_{\pm 4.51}$ | $16.87_{\pm 7.24}$ | $3.07_{\pm 0.38}$ | $6.86_{\pm 0.68}$ | [**57.54**$_{\pm 3.33}$] | **48.82**$_{\pm 2.27}$ | **48.74**$_{\pm 3.14}$ |
| *DS-17* | $17.65_{\pm 6.12}$ | **21.11**$_{\pm 3.90}$ | $5.48_{\pm 9.14}$ | $10.69_{\pm 5.22}$ | $13.90_{\pm 3.04}$ | $12.29_{\pm 5.34}$ | $14.27_{\pm 3.02}$ | $-6.19_{\pm 0.48}$ | $8.74_{\pm 2.76}$ | [**22.16**$_{\pm 3.96}$] | $18.87_{\pm 2.57}$ | **19.48**$_{\pm 2.14}$ |
| *DS-18* | $24.15_{\pm 1.51}$ | $24.24_{\pm 3.64}$ | $29.01_{\pm 4.07}$ | $30.23_{\pm 5.14}$ | $29.89_{\pm 3.72}$ | $22.30_{\pm 4.04}$ | $25.08_{\pm 2.68}$ | $27.22_{\pm 2.30}$ | $26.13_{\pm 2.58}$ | [**51.33**$_{\pm 4.08}$] | **40.39**$_{\pm 3.17}$ | **41.91**$_{\pm 2.49}$ |
| *DS-19* | $46.24_{\pm 2.84}$ | $33.48_{\pm 2.31}$ | **53.59**$_{\pm 1.74}$ | $44.11_{\pm 2.07}$ | $46.03_{\pm 2.12}$ | $43.64_{\pm 2.18}$ | $44.63_{\pm 1.85}$ | **54.18**$_{\pm 1.68}$ | $44.69_{\pm 5.44}$ | [**56.57**$_{\pm 1.90}$] | $49.05_{\pm 0.60}$ | $50.49_{\pm 1.68}$ |
| *DS-20* | $64.16_{\pm 1.84}$ | $59.10_{\pm 3.94}$ | $56.54_{\pm 9.01}$ | $52.50_{\pm 3.59}$ | $37.76_{\pm 4.80}$ | $51.94_{\pm 5.44}$ | $54.40_{\pm 2.30}$ | $64.33_{\pm 1.99}$ | $64.99_{\pm 3.39}$ | [**81.58**$_{\pm 1.30}$] | **77.21**$_{\pm 2.72}$ | **80.21**$_{\pm 1.68}$ |
| *DS-21* | $39.90_{\pm 1.88}$ | $41.42_{\pm 2.85}$ | $43.39_{\pm 2.54}$ | $38.72_{\pm 1.92}$ | $37.71_{\pm 2.89}$ | $35.75_{\pm 5.65}$ | $39.99_{\pm 3.26}$ | $42.29_{\pm 2.83}$ | $42.64_{\pm 4.96}$ | **44.16**$_{\pm 1.61}$ | [**46.13**$_{\pm 2.49}$] | **45.08**$_{\pm 1.85}$ |
| *DS-22* | $25.06_{\pm 1.23}$ | $15.59_{\pm 1.15}$ | $24.82_{\pm 1.88}$ | $25.01_{\pm 1.21}$ | $27.02_{\pm 0.52}$ | $27.16_{\pm 0.77}$ | $25.46_{\pm 1.04}$ | $24.94_{\pm 0.94}$ | $24.21_{\pm 2.45}$ | **30.56**$_{\pm 1.61}$ | **34.27**$_{\pm 0.90}$ | [**34.62**$_{\pm 0.93}$] |
| *DS-23* | $0.63_{\pm 0.09}$ | $1.05_{\pm 0.24}$ | $0.91_{\pm 0.14}$ | $0.61_{\pm 0.07}$ | $0.84_{\pm 0.09}$ | $0.80_{\pm 0.12}$ | $0.64_{\pm 0.15}$ | $0.72_{\pm 0.06}$ | $0.41_{\pm 0.08}$ | **7.77**$_{\pm 0.58}$ | **11.23**$_{\pm 0.54}$ | [**13.11**$_{\pm 0.89}$] |
| *DS-24* | $45.44_{\pm 4.83}$ | $21.30_{\pm 1.95}$ | $46.18_{\pm 3.41}$ | $39.94_{\pm 3.42}$ | $48.25_{\pm 1.13}$ | $39.33_{\pm 4.61}$ | $38.02_{\pm 3.31}$ | $51.21_{\pm 3.99}$ | $49.28_{\pm 3.89}$ | [**60.23**$_{\pm 1.13}$] | **53.71**$_{\pm 2.35}$ | **55.74**$_{\pm 2.64}$ |
| *DS-25* | $27.98_{\pm 1.33}$ | $35.84_{\pm 3.76}$ | $27.62_{\pm 4.08}$ | $30.77_{\pm 2.06}$ | $31.17_{\pm 1.11}$ | $29.95_{\pm 1.88}$ | $31.24_{\pm 1.45}$ | $33.48_{\pm 1.69}$ | $37.60_{\pm 3.33}$ | [**67.41**$_{\pm 2.82}$] | **61.32**$_{\pm 6.10}$ | **63.91**$_{\pm 5.02}$ |
| *DS-26* | $78.94_{\pm 4.63}$ | $61.55_{\pm 4.35}$ | $84.82_{\pm 3.40}$ | $62.73_{\pm 6.98}$ | $79.62_{\pm 4.36}$ | $48.52_{\pm 8.63}$ | $65.22_{\pm 2.48}$ | $85.68_{\pm 2.94}$ | $77.17_{\pm 7.15}$ | **88.10**$_{\pm 4.07}$ | [**94.46**$_{\pm 0.44}$] | **94.19**$_{\pm 0.40}$ |
| *DS-27* | $7.98_{\pm 0.40}$ | $6.88_{\pm 0.85}$ | $8.97_{\pm 0.51}$ | $9.51_{\pm 0.32}$ | $9.34_{\pm 0.24}$ | $9.22_{\pm 0.37}$ | **9.70**$_{\pm 0.44}$ | $9.69_{\pm 0.47}$ | $4.03_{\pm 0.59}$ | $9.48_{\pm 0.38}$ | **10.08**$_{\pm 0.30}$ | [**10.17**$_{\pm 0.35}$] |
| *DS-28* | $42.13_{\pm 4.44}$ | $25.53_{\pm 3.13}$ | $46.71_{\pm 3.13}$ | $40.52_{\pm 3.60}$ | $53.84_{\pm 2.85}$ | $40.97_{\pm 3.11}$ | $40.84_{\pm 3.37}$ | $48.54_{\pm 1.94}$ | $34.18_{\pm 12.43}$ | [**70.13**$_{\pm 4.54}$] | **65.83**$_{\pm 2.05}$ | **69.39**$_{\pm 1.88}$ |
| *DS-29* | $26.82_{\pm 8.18}$ | $24.92_{\pm 13.52}$ | $58.48_{\pm 5.61}$ | $50.80_{\pm 11.04}$ | $60.54_{\pm 0.92}$ | $11.89_{\pm 6.18}$ | $56.48_{\pm 2.15}$ | $6.77_{\pm 17.60}$ | $5.27_{\pm 2.76}$ | **76.30**$_{\pm 4.91}$ | **77.15**$_{\pm 0.64}$ | [**77.59**$_{\pm 0.29}$] |
| *DS-30* | $33.95_{\pm 4.16}$ | $19.23_{\pm 1.81}$ | $38.28_{\pm 3.47}$ | $33.41_{\pm 1.81}$ | $52.57_{\pm 3.11}$ | $30.04_{\pm 3.59}$ | $37.33_{\pm 2.13}$ | $43.99_{\pm 1.67}$ | OM | [**67.35**$_{\pm 2.13}$] | **61.86**$_{\pm 2.41}$ | **62.52**$_{\pm 3.90}$ |
| Avg. score | 34.21 | 31.69 | 34.21 | 38.28 | 41.08 | 32.86 | 40.63 | 33.68 | - | **57.15** | **55.27** | **56.13** |
| Avg. rank | 8.60 | 8.87 | 8.23 | 7.77 | 6.57 | 9.03 | 7.00 | 7.33 | 7.53 | **1.93** | **2.73** | **2.27** |

a unified ensemble clustering framework, and propose three specific ensemble clustering algorithms, termed MDEC-HC, MDEC-SC, and MDEC-BG, respectively. In this section, we will evaluate the influence of the three key factors (i.e., diversified metrics, random subspaces, and weighted clusters) upon the proposed three algorithms.

First, we compare the diversified metrics with several widely used similarity metrics, namely, cosine similarity, correlation coefficient, and Spearman correlation coefficient. For each of the three proposed algorithms, we test its performance of using the diversified metrics against the above three metrics. Taking MDEC-HC as an example, as shown in Fig. 5, using diversified metrics leads to an average NMI(%) score $67.49$, while using the three conventional similarity metrics leads to average NMI(%) scores of $50.99$, $48.23$, and $41.34$, respectively. Similar advantages of using diversified metrics can also be observed in the cases of MDEC-SC and MDEC-BG, which confirm that the use of diversified metrics in the proposed algorithms can substantially benefits the consensus clustering performance (as shown in Fig. 5).

Second, we evaluate the performances of the proposed ensemble clustering algorithms with different subspace sampling ratio $\tau$, which varies from $0.1$ to $1$. As can be observed in Figs. 6 and 7, moderate values of $\tau$ generally lead to better consensus clustering performance. When the sampling ratio $\tau$ goes from $0.8$ to $1$, the performance declines, which suggests that the use of random subspaces exhibits a positive influence when compared to using the full feature sets (by setting $\tau = 1$). At the other extreme, when setting $\tau$ to very small values, e.g., in the range of [0.1, 0.3], the performance also declines, due to the fact that the subspaces generated by a very small sampling ratio may not well reflect the underlying structure of the dataset. Empirically, it is suggested that the sampling ratio $\tau$ be set in the range of $[0.4, 0.8]$, which strikes a balance between diversity and quality. In this paper, the sampling ratio $\tau = 0.5$ is used in the experiments on all the benchmark datasets.

Third, we evaluate the performances of the proposed algorithms *with* and *without* the weighted clusters. Note that the performances of the proposed algorithms without weighted clusters are obtained by setting all cluster weights equal to one. As shown in Figs. 6 and 7, in terms of both NMI and ARI, the proposed three algorithms with weighted clusters exhibit consistently better average performances (across the 30 benchmark datasets) than that without weighted clusters.

From the comparison results in Figs. 5, 6, and 7, we can have two main observations: 1) the performance of our approach benefits from the use of diversified metrics, random subspaces, and weighted clusters; 2) out of the three beneficial factors, the diversified metrics play the possibly most important role in the consensus clustering performance, with consideration to the clear improvement (w.r.t. both NMI and ARI) that they lead to.

### *F. Execution Time*

In this section, we evaluate the efficiency of different ensemble clustering algorithms and report their execution times

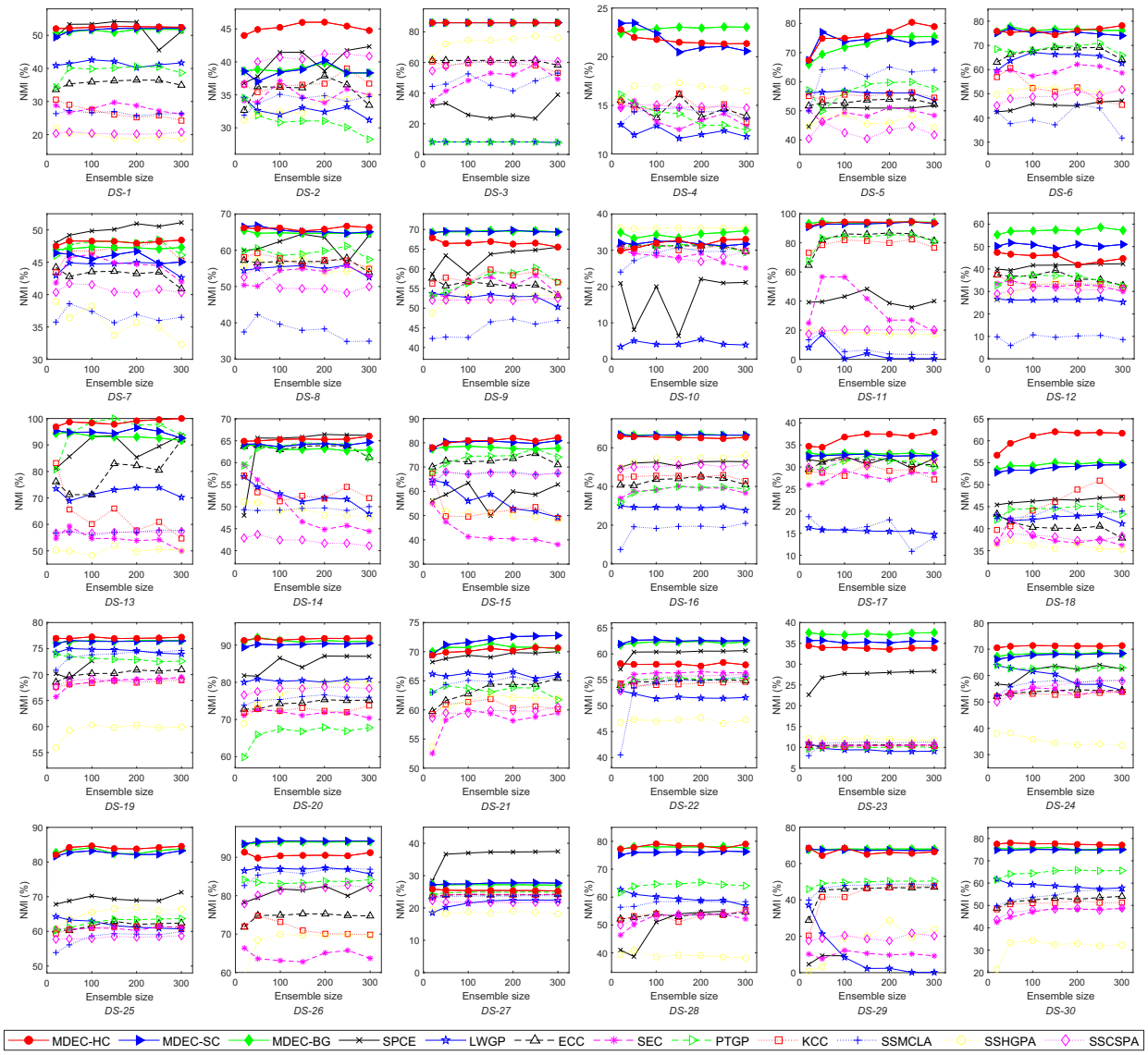



Fig. 2. The average performances (w.r.t. NMI) over 20 runs by different ensemble clustering algorithms with varying ensemble sizes $M$. Note that if an algorithm is not feasible (due to the out-of-memory error) for some large ensemble sizes on a dataset, then its curve will stop earlier than other algorithms.

on the benchmark datasets in Table V. The experiments are conducted in MATLAB R2016a 64-bit on a workstation with Intel i9-7940X CPU and 64 GB memory.

In general, larger dimensions and larger sample sizes lead to greater computational costs for the ensemble clustering algorithms. As shown in Table V, the proposed MDEC-HC, MDEC-SC, and MDEC-BG algorithms each consume less than 1 second of time on 14 out of the totally 18 cancer gene expression datasets. On the 12 image/speech datasets, the proposed algorithms also show comparable time efficiency with the other ensemble clustering algorithms.

To summarize, as can be seen in Tables III to V and Figs. 2 to 4, the proposed three ensemble clustering algorithms have shown considerable advantages in clustering effectiveness while exhibiting competitive time efficiency when compared with the state-of-the-art ensemble clustering algorithms.

## V. Conclusion

In this paper, we present a new ensemble clustering approach termed MDEC, which is capable of jointly exploiting large populations of diversified metrics, random subspaces, and weighted clusters in a unified ensemble clustering framework. Specifically, a large number of diversified metrics are generated by randomizing a scaled exponential similarity kernel. The diversified metrics are then coupled with the random subspaces to form a large set of metric-subspace pairs. Upon the similarity matrices derived from the metric-subspace pairs, the spectral clustering algorithm is performed to construct an ensemble of diversified base clusterings. With the base clusterings generated, an entropy-based cluster validity strategy is utilized to evaluate and weight the clusters with consideration to the distribution of the cluster labels in the entire ensemble. Finally, based on diversified metrics, random subspaces, and weighted clusters, three specific ensemble clus-

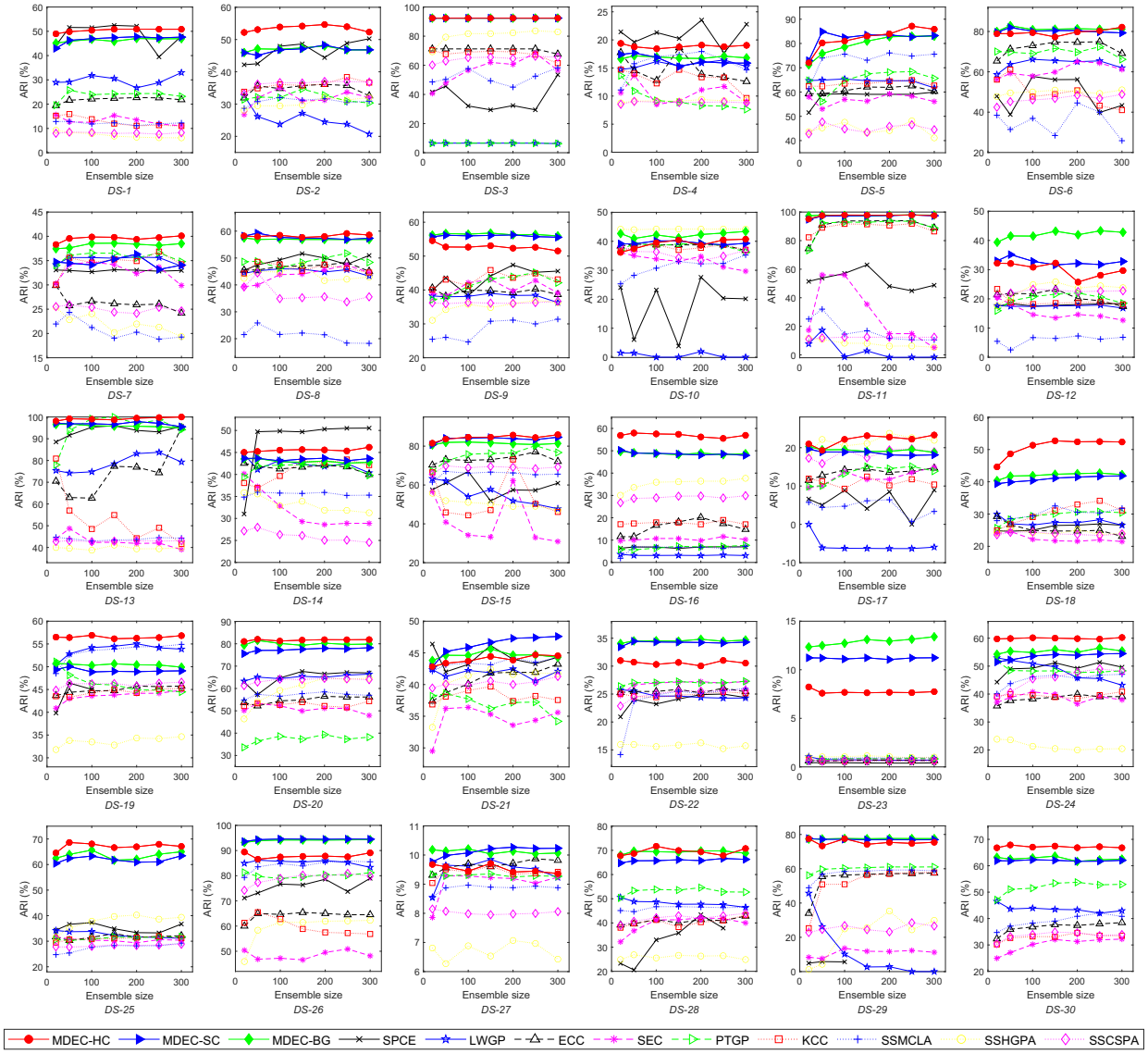
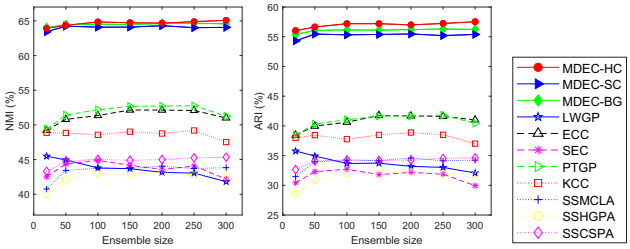



Fig. 3. The average performances (w.r.t. ARI) over 20 runs by different ensemble clustering algorithms with varying ensemble sizes $M$. Note that if an algorithm is not feasible (due to the out-of-memory error) for some large ensemble sizes on a dataset, then its curve will stop earlier than other algorithms.

tering algorithms are devised by incorporating three types of consensus functions. Extensive experiments are conducted on 30 real-world high-dimensional datasets (including 18 cancer gene expression datasets and 12 image/speech datasets), which have demonstrated the advantages of the proposed algorithms over the state-of-the-art ensemble clustering algorithms.

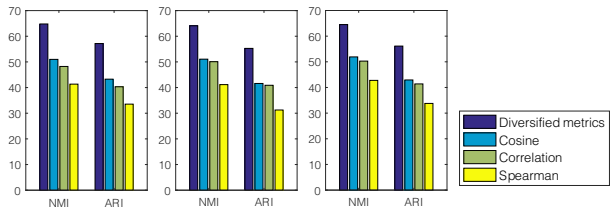
Note that in the practical application process of the proposed algorithms (as well as other ensemble clustering algorithms), two types of balance should be taken into consideration, i.e., the balance between diversity and quality (which involves enhancing different levels of diversity while maintaining the quality of the individuals) and the balance between effectiveness and efficiency (which involves the choice of base clusterers and the size of the ensembles).

In the future work, there are several possible research directions. First, starting from this work, more metric diversification strategies can be studied to seek more opportunities in metric spaces for ensemble clustering of high-dimensional data. Second, different consensus methods for exploiting multi-level diversity in ensembles can be investigated to deal with different



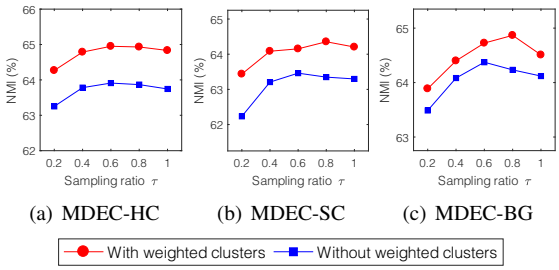

(a) NMI (b) ARI

Fig. 4. Average curves (across 30 datasets) by different algorithms with varying ensemble size $M$. Note that Fig. 4(a) is obtained by averaging the 30 sub-figures in Figs. 2, and Fig. 4(b) the 30 sub-figures in Figs. 3.



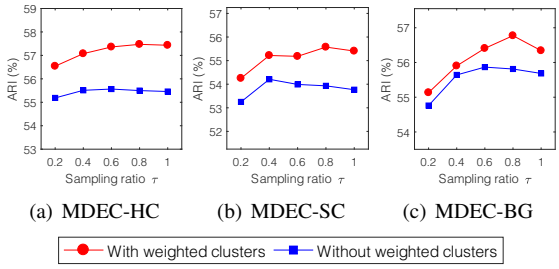

(a) MDEC-HC (b) MDEC-SC (c) MDEC-BG

Fig. 5. Average performance (across 30 datasets) of the proposed algorithms using diversified metrics against using other similarity metrics.




(a) MDEC-HC (b) MDEC-SC (c) MDEC-BG

Fig. 6. Average NMI scores (across 30 datasets) of the proposed algorithms *with* and *without* weighted clusters using varying sampling ratios $\tau$.




(a) MDEC-HC (b) MDEC-SC (c) MDEC-BG

Fig. 7. Average ARI scores (across 30 datasets) of the proposed algorithms *with* and *without* weighted clusters using varying sampling ratios $\tau$.

types of complex data. Last but not least, the diversification-and-fusion strategy may be generalized to other unsupervised learning tasks (such as unsupervised feature selection and unsupervised outlier detection) to alleviate the difficulties in parameter selection, metric selection, and model selection.

## Acknowledgment

This project was supported by NSFC (61976097, 61876193 & 61876104) and A*STAR-NTU-SUTD AI Partnership Grant (No. RGANS1905).

## References

[1] T. Li and C. Ding, "Weighted consensus clustering," in *Proc. of SIAM International Conference on Data Mining (SDM)*, 2008, pp. 798–809.

[2] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2396–2409, 2011.

[3] T. Wang, "CA-Tree: A hierarchical structure for efficient and scalable coassociation-based cluster ensembles," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 3, pp. 686–698, 2011.

[4] N. Li and L. J. Latecki, "Clustering aggregation as maximum-weight independent set," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 782–790.

[5] L. Zheng, T. Li, and C. Ding, "A framework for hierarchical ensemble clustering," *ACM Transactions on Knowledge Discovery from Data*, vol. 9, no. 2, pp. 9:1–9:23, 2014.

[6] L. Jing, K. Tian, and J. Z. Huang, "Stratified feature sampling method for ensemble clustering of high dimensional data," *Pattern Recognition*, vol. 48, no. 11, pp. 3688–3702, 2015.

[7] J. Wu, H. Liu, H. Xiong, J. Cao, and J. Chen, "K-means-based consensus clustering: A unified view," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 1, pp. 155–169, 2015.

[8] D. Huang, J.-H. Lai, and C.-D. Wang, "Robust ensemble clustering using probability trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1312–1326, 2016.

[9] D. Huang, J. Lai, and C.-D. Wang, "Ensemble clustering using factor graph," *Pattern Recognition*, vol. 50, pp. 131–142, 2016.

[10] H. Liu, M. Shao, S. Li, and Y. Fu, "Infinite ensemble clustering," *Data Mining and Knowledge Discovery*, vol. 32, pp. 385–416, 2018.

[11] H. Liu, J. Wu, T. Liu, D. Tao, and Y. Fu, "Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1129–1143, 2017.

[12] H. Liu, R. Zhao, H. Fang, F. Cheng, Y. Fu, and Y.-Y. Liu, "Entropy-based consensus clustering for patient stratification," *Bioinformatics*, vol. 33, no. 17, pp. 2691–2698, 2017.

[13] Z. Yu, P. Luo, J. Liu, H. Wong, J. You, G. Han, and J. Zhang, "Semi-supervised ensemble clustering based on selected constraint projection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2394–2407, 2018.

[14] Y. Shi, Z. Yu, C. L. P. Chen, J. You, H. Wong, Y. Wang, and J. Zhang, "Transfer clustering ensemble selection," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2872–2885, 2020.

[15] D. Huang, C. D. Wang, and J. H. Lai, "Locally weighted ensemble clustering," *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1460–1473, 2018.

[16] D. Huang, C.-D. Wang, H. Peng, J.-H. Lai, and C.-K. Kwoh, "Enhanced ensemble clustering via fast propagation of cluster-wise similarities," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[17] D. Huang, C.-D. Wang, J.-S. Wu, J.-H. Lai, and C.-K. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1212–1226, 2020.

[18] L. Bai, J. Liang, and Y. Guo, "An ensemble clusterer of multiple fuzzy $k$ -means clusterings to recognize arbitrarily shaped clusters," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3524–3533, 2018.

[19] L. Bai, J. Liang, H. Du, and Y. Guo, "An information-theoretical framework for cluster ensemble," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 8, pp. 1464–1477, 2019.

[20] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "Marginalized multiview ensemble clustering," *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

[21] X. Yu, G. Yu, J. Wang, and C. Domeniconi, "Co-clustering ensembles based on multiple relevance measures," *IEEE Transactions on Knowledge and Data Engineering, in press*, 2019.

[22] F. Li, Y. Qian, J. Wang, C. Dang, and L. Jing, "Clustering ensemble based on sample's stability," *Artificial Intelligence*, vol. 273, pp. 37–55, 2019.

[23] Z. Yu, H.-S. Wong, and H. Wang, "Graph-based consensus clustering for class discovery from gene expression data," *Bioinformatics*, vol. 23, no. 21, pp. 2888–2896, 2007.

[24] Z. Yu, L. Li, J. Liu, J. Zhang, and G. Han, "Adaptive noise immune cluster ensemble using affinity propagation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 12, pp. 3176–3189, 2015.

[25] Z. Yu, P. Luo, J. You, H. S. Wong, H. Leung, S. Wu, J. Zhang, and G. Han, "Incremental semi-supervised clustering ensemble for high dimensional data clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 701–714, 2016.

[26] Z. Yu, Z. Kuang, J. Liu, H. Chen, J. Zhang, J. You, H. S. Wong, and G. Han, "Adaptive ensembling of semi-supervised clustering solutions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 8, pp. 1577–1590, 2017.

[27] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach," in *Proc. of International Conference on Machine Learning (ICML)*, 2003, pp. 186–193.

TABLE V
THE EXECUTION TIMES (IN SECONDS) OF DIFFERENT ENSEMBLE CLUSTERING ALGORITHMS (WITH BOTH ENSEMBLE GENERATION AND CONSENSUS FUNCTION INCLUDED).

| Dataset | SSCSPA | SSHGPA | SSMCLA | KCC | PTGP | SEC | ECC | LWGP | SPCE | MDEC-HC | MDEC-SC | MDEC-BG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DS-1* | 1.12 | 1.51 | 1.38 | 0.87 | 1.25 | 0.80 | 0.95 | 1.65 | 1.80 | 1.00 | 1.01 | 1.07 |
| *DS-2* | 0.82 | 1.02 | 1.01 | 0.44 | 0.62 | 0.42 | 0.45 | 1.38 | 0.53 | 0.75 | 0.76 | 0.77 |
| *DS-3* | 0.91 | 1.04 | 1.08 | 0.45 | 0.69 | 0.42 | 0.46 | 1.10 | 0.79 | 0.77 | 0.78 | 0.80 |
| *DS-4* | 1.12 | 1.31 | 1.31 | 0.64 | 0.81 | 0.61 | 0.68 | 1.24 | 1.05 | 0.87 | 0.88 | 0.90 |
| *DS-5* | 0.86 | 1.02 | 1.06 | 0.52 | 0.73 | 0.49 | 0.52 | 1.14 | 0.71 | 0.78 | 0.79 | 0.80 |
| *DS-6* | 0.86 | 1.08 | 1.06 | 0.54 | 0.71 | 0.50 | 0.56 | 1.44 | 0.79 | 0.78 | 0.78 | 0.80 |
| *DS-7* | 1.13 | 1.00 | 0.99 | 0.46 | 0.64 | 0.43 | 0.48 | 1.04 | 0.58 | 0.73 | 0.74 | 0.75 |
| *DS-8* | 0.80 | 0.97 | 0.97 | 0.46 | 0.61 | 0.43 | 0.48 | 1.01 | 0.56 | 0.71 | 0.72 | 0.72 |
| *DS-9* | 1.06 | 1.52 | 1.28 | 0.79 | 1.04 | 0.68 | 0.90 | 1.47 | 1.72 | 0.90 | 0.91 | 0.95 |
| *DS-10* | 0.80 | 0.93 | 0.96 | 0.45 | 0.61 | 0.43 | 0.45 | 1.01 | 0.51 | 0.72 | 0.73 | 0.74 |
| *DS-11* | 1.68 | 1.78 | 1.89 | 1.31 | 1.74 | 1.23 | 1.37 | 2.15 | 4.01 | 1.14 | 1.15 | 1.21 |
| *DS-12* | 1.52 | 1.81 | 1.75 | 1.34 | 1.68 | 1.20 | 1.50 | 2.10 | 3.94 | 1.15 | 1.17 | 1.23 |
| *DS-13* | 0.85 | 1.09 | 1.03 | 0.49 | 0.67 | 0.47 | 0.50 | 1.07 | 0.63 | 0.74 | 0.74 | 0.76 |
| *DS-14* | 0.86 | 1.09 | 1.03 | 0.52 | 0.69 | 0.49 | 0.54 | 1.09 | 0.64 | 0.74 | 0.75 | 0.76 |
| *DS-15* | 0.89 | 1.12 | 1.09 | 0.52 | 0.71 | 0.49 | 0.54 | 1.14 | 0.76 | 1.16 | 1.17 | 1.19 |
| *DS-16* | 1.07 | 1.48 | 1.40 | 0.91 | 1.75 | 0.75 | 1.06 | 2.26 | 1.97 | 0.87 | 0.89 | 0.93 |
| *DS-17* | 0.82 | 0.98 | 0.99 | 0.44 | 0.58 | 0.41 | 0.45 | 0.99 | 0.55 | 0.73 | 0.74 | 0.75 |
| *DS-18* | 1.02 | 1.33 | 1.22 | 0.66 | 0.89 | 0.62 | 0.70 | 1.31 | 1.21 | 0.83 | 0.84 | 0.87 |
| *DS-19* | 41.32 | 90.94 | 26.43 | 61.71 | 156.60 | 43.51 | 87.18 | 59.94 | 4032.33 | 137.40 | 144.16 | 139.01 |
| *DS-20* | 3.43 | 4.22 | 3.09 | 5.32 | 4.89 | 4.15 | 6.58 | 7.41 | 114.84 | 5.44 | 5.71 | 5.84 |
| *DS-21* | 3.43 | 3.58 | 2.91 | 2.82 | 2.72 | 1.48 | 4.26 | 5.39 | 178.47 | 6.89 | 7.18 | 7.37 |
| *DS-22* | 2.98 | 43.10 | 3.37 | 3.52 | 3.53 | 2.04 | 5.78 | 5.72 | 106.74 | 6.09 | 6.56 | 6.48 |
| *DS-23* | 9.55 | 16.17 | 6.84 | 17.40 | 19.09 | 13.82 | 20.82 | 18.93 | 354.36 | 15.66 | 16.92 | 16.32 |
| *DS-24* | 3.42 | 5.85 | 3.54 | 3.78 | 4.54 | 2.37 | 5.80 | 6.33 | 90.75 | 7.23 | 7.47 | 7.64 |
| *DS-25* | 5.92 | 7.04 | 6.18 | 13.46 | 13.43 | 13.11 | 13.90 | 15.21 | 29.72 | 4.72 | 4.79 | 4.89 |
| *DS-26* | 5.58 | 5.36 | 4.33 | 6.80 | 7.66 | 4.73 | 8.87 | 9.37 | 140.78 | 9.49 | 9.82 | 10.06 |
| *DS-27* | 44.27 | 46.46 | 44.12 | 204.92 | 204.92 | 203.34 | 206.45 | 206.91 | 299.38 | 31.26 | 31.51 | 31.64 |
| *DS-28* | 33.05 | 31.17 | 18.75 | 43.00 | 69.10 | 33.34 | 53.65 | 44.77 | 1115.70 | 66.18 | 68.29 | 67.67 |
| *DS-29* | 117.37 | 95.22 | 97.91 | 450.27 | 602.59 | 441.79 | 464.34 | 459.80 | 3590.84 | 238.26 | 240.38 | 239.67 |
| *DS-30* | 74.94 | 79.57 | 37.49 | 65.89 | 590.31 | 43.67 | 109.65 | 70.75 | OM | 302.52 | 312.08 | 305.36 |


[28] J. H. Lee, K. T. McDonnell, A. Zelenyuk, D. Imre, and K. Mueller, "A structure-based distance metric for high-dimensional space exploration with multidimensional scaling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 351–364, 2014.

[29] C. M. Hsu and M. S. Chen, "On the design and applicability of distance functions in high-dimensional data space," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 4, pp. 523–536, 2009.

[30] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835–850, 2005.

[31] J. Wu, H. Liu, H. Xiong, and J. Cao, "A theoretic framework of k-means-based consensus clustering," in *Proc. of International Joint Conference on Artificial Intelligence*, 2013.

[32] C. Zhong, X. Yue, Z. Zhang, and J. Lei, "A clustering ensemble: Two-level-refined co-association matrix with path-based transformation," *Pattern Recognition*, vol. 48, no. 8, pp. 2699–2709, 2015.

[33] D. Huang, J.-H. Lai, and C.-D. Wang, "Combining multiple clusterings via crowd agreement estimation and multi-granularity link analysis," *Neurocomputing*, vol. 170, pp. 240–250, 2015.

[34] Y. Fan, N. Li, C. Li, Z. Ma, L. J. Latecki, and K. Su, "Restart and random walk in local search for maximum vertex weight cliques with evaluations in clustering aggregation," in *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, 2017, pp. 622–630.

[35] M. Yousefnezhad, S. J. Huang, and D. Zhang, "WoCE: A framework for clustering ensemble by exploiting the wisdom of crowds theory," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 486–499, 2018.

[36] D. Huang, J.-H. Lai, C.-D. Wang, and P. C. Yuen, "Ensembling over-segmentations: From weak evidence to strong segmentation," *Neurocomputing*, vol. 207, pp. 416–427, 2016.

[37] X. Wang, C. Yang, and J. Zhou, "Clustering aggregation by probability accumulation," *Pattern Recognition*, vol. 42, no. 5, pp. 668–675, 2009.

[38] J. Yi, T. Yang, R. Jin, and A. K. Jain, "Robust ensemble clustering by matrix completion," in *Proc. of IEEE International Conference on Data Mining (ICDM)*, 2012.

[39] A. Strehl and J. Ghosh, "Cluster ensembles: A knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003.

[40] X. Z. Fern and C. E. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning," in *Proc. of International Conference on Machine Learning (ICML)*, 2004.

[41] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: models of consensus and weak partitions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1866–1881, 2005.

[42] L. Franek and X. Jiang, "Ensemble clustering by means of clustering embedding in vector spaces," *Pattern Recognition*, vol. 47, no. 2, pp. 833–842, 2014.

[43] A. K. Jain, "Data clustering: 50 years beyond $k$-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[44] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.

[45] E. Weiszfeld and F. Plastria, "On the point for which the sum of the distances to n given points is minimum," *Annals of Operations Research*, vol. 167, no. 1, pp. 7–41, 2009.

[46] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[47] J. Chu, H. Wang, J. Liu, Z. Gong, and T. Li, "Unsupervised feature learning architecture with multi-clustering integration RBM," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[48] Q. Wang, P. C. Yuen, and G. Feng, "Semi-supervised metric learning via topology preserving multiple semi-supervised assumptions," *Pattern Recognition*, vol. 46, no. 9, pp. 2576–2587, 2013.

[49] F. Xiong, M. Kam, L. Hrebien, B. Wang, and Y. Qi, "Kernelized information-theoretic metric learning for cancer diagnosis using high-dimensional molecular profiling data," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 4, 2016.

[50] J. Li, A. J. Ma, and P. C. Yuen, "Semi-supervised region metric learning for person re-identification," *International Journal of Computer Vision*, vol. 126, pp. 855–874, 2018.

[51] X. Liu, L. Wang, X. Zhu, M. Li, E. Zhu, T. Liu, L. Liu, Y. Dou, and J. Yin, "Absent multiple kernel learning algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence, in press*, 2019.

[52] X. Chen, W. Sun, B. Wang, Z. Li, X. Wang, and Y. Ye, "Spectral clustering of customer transaction data with a two-level subspace weighting method," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3230–3241, 2019.

[53] H. Liu, Z. Han, Y.-S. Liu, and M. Gu, "Fast low-rank metric learning for large-scale and high-dimensional data," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 819–829.

[54] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, pp. 333–337, 2014.

[55] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[56] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[57] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: A comparative study," *BMC bioinformatics*, vol. 9, no. 1, p. 497, 2008.

[58] K. Bache and M. Lichman, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[59] S. A. Nene, S. K. Nayar, H. Murase *et al.*, "Columbia object image library (COIL-20)," *Technical report CUCS-005-96*, 1996.

[60] F. Nie, Z. Wang, R. Wang, and X. Li, "Submanifold-preserving discriminant analysis with an auto-optimized graph," *IEEE Transactions on Cybernetics, in press*, 2019.

[61] S. Roweis, http://www.cs.nyu.edu/\%7eroweis/data.html.

[62] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignatures for general purpose face recognition," in *Face Recognition*. Springer, 1998, pp. 446–456.

[63] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 1447–1454.

[64] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *Journal of Machine Learning Research*, vol. 11, no. 11, pp. 2837–2854, 2010.

[65] P. Zhou, L. Du, X. Liu, Y.-D. Shen, M. Fan, and X. Li, "Self-paced clustering ensemble," *IEEE Transactions on Neural Networks and Learning Systems, in press*, 2020.

**Dong Huang** received the B.S. degree in computer science from South China University of Technology, Guangzhou, China, in 2009, and the M.Sc. and Ph.D. degrees in computer science from Sun Yat-sen University, Guangzhou, in 2011 and 2015, respectively. He joined South China Agricultural University, Guangzhou, in 2015, where he is currently working as an Associate Professor and Deputy Head in the Department of Computer Science. From 2017 to 2018, he was a visiting fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His current research interests include data mining and machine learning, and more specifically focus on ensemble clustering, multi-view clustering, and large-scale clustering. He has published more than 50 papers in refereed journals and conferences, including IEEE TKDE, IEEE TCYB, IEEE TSMC-S, ACM TKDD, Pattern Recognition, KBS, AAAI, and ICDM. He is a member of the IEEE.

**Chang-Dong Wang** received the B.S. degree in applied mathematics in 2008, the M.Sc. degree in computer science in 2010, and the Ph.D. degree in computer science in 2013, all from Sun Yat-sen University, Guangzhou, China. He was a visiting student at the University of Illinois at Chicago from January 2012 to November 2012. He is currently an Associate Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His current research interests include machine learning and data mining. He has published more than 100 scientific papers in international journals and conferences such as IEEE TPAMI, IEEE TKDE, IEEE TCYB, ACM TKDD, Pattern Recognition, KAIS, IJCAI, AAAI, KDD, ICDM and SDM. His ICDM 2010 paper won the Honorable Mention for Best Research Paper Awards. He was awarded 2015 Chinese Association for Artificial Intelligence (CAAI) Outstanding Dissertation. He is a member of the IEEE.

**Jian-Huang Lai** received the M.Sc. degree in applied mathematics in 1989 and the Ph.D. degree in mathematics in 1999 from Sun Yat-sen University, China. He joined Sun Yat-sen University in 1989 as an Assistant Professor, where he is currently a Professor with the School of Data and Computer Science. His current research interests include the areas of digital image processing, pattern recognition, multimedia communication, wavelet and its applications. He has published more than 200 scientific papers in the international journals and conferences on image processing and pattern recognition, such as IEEE TPAMI, IEEE TKDE, IEEE TNN, IEEE TIP, Pattern Recognition, ICCV, CVPR, IJCAI, ICDM and SDM. Prof. Lai serves as a Standing Member of the Image and Graphics Association of China, and also serves as a Standing Director of the Image and Graphics Association of Guangdong. He is a senior member of the IEEE.

**Chee-Keong Kwoh** received the bachelor's degree in electrical engineering (first class) and the master's degree in industrial system engineering from the National University of Singapore in 1987 and 1991, respectively. He received the PhD degree from the Imperial College of Science, Technology and Medicine, University of London, in 1995. He has been with the School of Computer Science and Engineering, Nanyang Technological University (NTU) since 1993. He is the programme director of the MSc in Bioinformatics programme at NTU. His research interests include data mining, soft computing and graph-based inference; applications areas include bioinformatics and biomedical engineering. He has done significant research work in his research areas and has published many quality international conferences and journal papers. He is an editorial board member of the International Journal of Data Mining and Bioinformatics; the Scientific World Journal; Network Modeling and Analysis in Health Informatics and Bioinformatics; Theoretical Biology Insights; and Bioinformation. He has been a guest editor for many journals such as Journal of Mechanics in Medicine and Biology, the International Journal on Biomedical and Pharmaceutical Engineering and others. He is a member of the Association for Medical and Bio-Informatics, Imperial College Alumni Association of Singapore. He has provided many services to professional bodies in Singapore and was conferred the Public Service Medal by the President of Singapore in 2008. He is a senior member of the IEEE.