Brain EEG Time-Series Clustering Using Maximum-Weight Clique

Chenglong Dai, Jia Wu, *Member, IEEE*, Dechang Pi, Stefanie I. Becker, Lin Cui, Qin Zhang, and Blake Johnson

***Abstract*—Brain electroencephalography (EEG), the complex, weak, multivariate, nonlinear, and nonstationary time series, has been recently widely applied in neurocognitive disorder diagnoses and brain–machine interface developments. With its specific features, unlabeled EEG is not well addressed by conventional unsupervised time-series learning methods. In this article, we handle the problem of unlabeled EEG time-series clustering and propose a novel EEG clustering algorithm, that we call mwcEEGc. The idea is to map the EEG clustering to the maximum-weight clique (MWC) searching in an improved Fréchet similarity-weighted EEG graph. The mwcEEGc considers the weights of both vertices and edges in the constructed EEG graph and clusters EEG based on their similarity weights instead of calculating the cluster centroids. To the best of our knowledge, it is the first attempt to cluster unlabeled EEG trials using MWC searching. The mwcEEGc achieves high-quality clusters with respect to intracluster compactness as well as intercluster scatter. We demonstrate the superiority of mwcEEGc over ten state-of-the-art unsupervised learning/clustering approaches by conducting detailed experimentations with the standard clustering validity criteria on 14 real-world brain EEG datasets. We also present that mwcEEGc satisfies the theoretical properties of clustering, such as richness, consistency, and order independence.**

***Index Terms*—Clustering, electroencephalography (EEG) time series, Fréchet distance (FD), maximum-weight clique (MWC), weighted EEG graph.**

Manuscript received April 10, 2019; revised November 6, 2019, January 8, 2020, and January 23, 2020; accepted February 4, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1433116 and Grant 61702355, in part by the Fundamental Research Funds for the Central Universities under Grant NP2017208, in part by the ARC DECRA Project under Grant DE200100964, and in part by the ARC Discovery Early Career Researcher Award (DECRA) under Grant DE200100964. This article was recommended by Associate Editor H. A. Abbass. *(Corresponding author: Dechang Pi.)*

Chenglong Dai and Dechang Pi are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: chenglongdai@nuaa.edu.cn; dc.pi@nuaa.edu.cn).

Jia Wu is with the Department of Computing, Macquarie University, Sydney, NSW 2109, Australia (e-mail: jia.wu@mq.edu.au).

Stefanie I. Becker is with the School of Psychology, University of Queensland, St. Lucia, QLD 4072, Australia (e-mail: s.becker@psy.uq.edu.au).

Lin Cui is with the Intelligent Information Processing Laboratory, Suzhou University, Suzhou 234000, China (e-mail: jsjxcuilin@nuaa.edu.cn).

Qin Zhang is with the Centre for Artificial Intelligence, University of Technology Sydney, Ultimo, NSW 2007, Australia, and also with the Data Centre of Social Network Group, Tencent, Shenzhen 518057, China (e-mail: amberqzhang@tencent.com).

Blake Johnson is with the Department of Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia (e-mail: blake.johnson@mq.edu.au).

This article has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. The supplementary material contains five parts that further elaborate on the idea of this article. (1) Section S-I: Proof of Theorem 1; (2) Section S-II: Proof of Proposition 1; (3) Section S-III: Proof of Theorem 2; (4) Section S-IV: Proof of Proposition 2; and (5) Section V: A figure that illustrates IFD-based EEG similarity distribution and similarity matrix of 14 EEG datasets. The total size of the supplementary material is 776 kB.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.2974776

I. Introduction

BRAIN electroencephalography (EEG), the electrical temporal signal generated by the cerebral cortex, is one specific type of time series, with features of high complexity, multivariate, nonlinearity, nonstationarity, and a low signal-to-noise ratio. It is reported that EEG throughout the entire life of a human reflects not only the particular brain functions but also the states of the entire body [1]. From the early 20th century, EEG, as a noninvasive technique, has been widely studied and used to research neurocognitive disorders, including Alzheimer's disease (AD) [2], [3]; epileptic seizures [4], [5]; stroke [6], [7]; etc. Meanwhile, it is also applied in brain–machine interface (BMI) [8] (or brain–computer interface (BCI) [9], [10]), including motor imagery detection classification [11], [12]; robotic arm control [13], [14]; wheelchair navigation [15], [16]; etc. As is known to us, the existing methods in the two most popular applications require labels of EEG signals. However, EEG signals that lack labels in these fields are increasing mainly due to: 1) the uncontrolled cerebral activities of subjects with unidentified EEG patterns, especially for those patients suffering from cerebral diseases; 2) the uncertainty of cerebral disease patterns in different stages for disease diagnosis; 3) the newly activating EEG with unidentified control commands to enrich multitask BCI applications close to real life; and 4) the label incompleteness or mislabeling of EEG signals when recording. As a result, manually labeling EEG becomes a time-consuming task and the absence of labels also precludes the conventional methods, for example, classification, from availably analyzing unlabeled EEG signals. Therefore, novel unsupervised techniques, for example, clustering, are required to solve the problems caused by ever-increasing unlabeled EEG.

To handle the challenging but valuable task, this article proposes a novel EEG clustering (i.e., mwcEEGc) inspired by maximum-weight clique (MWC), whose idea is to map unlabeled EEG clustering to repeating MWC searching in a complete-undirected Fréchet distance (FD)-weighted EEG

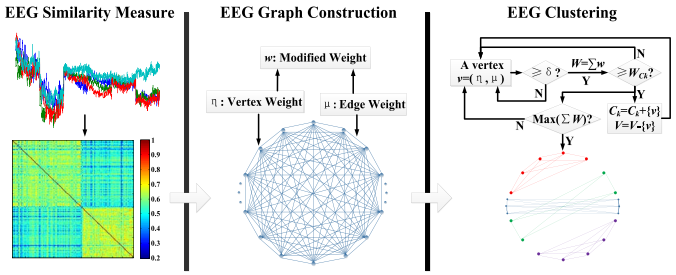

Fig. 1. Framework of the method, which mainly contains three parts: 1) EEG similarity measure; 2) EEG graph construction; and 3) EEG clustering.

graph. This method simultaneously considers vertex weights and edge weights of the FD-weighted EEG graph, and it concentrates on the compactness between any two EEG trials in the same cluster and the scatter in different clusters. Furthermore, unlike most conventional clustering methods, the proposed method is not required to calculate cluster centers. Consequently, the proposed method produces consistent and order-independent EEG clustering results. In detail, the contributions are highlighted as follows.

1) We formulate the problem of EEG clustering as MWC searching in a similarity-weighted EEG graph. To the best of our knowledge, it is the first attempt to cluster unlabeled EEG using MWC.
2) We propose a novel algorithm, mwcEEGc, simultaneously considering edge weight and vertex weight when using MWC to cluster EEG trials. The mwcEEGc provides high-quality EEG clustering with respect to intracluster compactness and intercluster scatter.
3) We present the efficacy of mwcEEGc with detailed experiments in a way that mwcEEGc is compared to ten state-of-the-art clustering approaches on 14 EEG datasets. The experimental results clearly demonstrate that mwcEEGc yields the best quality of unlabeled EEG clusters.
4) We also summarize the satisfiability of mwcEEGc with four theoretical clustering properties, and mwcEEG satisfies three: a) richness; b) consistency; and c) order independence.

The remainder of this article is organized as follows. An overall framework of our method is briefly introduced in Section II. The related works are reviewed in Section III. Preliminaries of the Fréchet similarity and MWC that we used in our method are introduced in Section IV. The proposed algorithm mwcEEGc for EEG clustering is presented in Section V. Then, detailed experimentation is carried out in Section VI. Subsequently, the satisfiability of four theoretical clustering properties, that is: 1) scale invariance; 2) richness; 3) consistency; and 4) order independence, for mwcEEGc is discussed in Section VII. Finally, we summarize this article and orientate future work in Section VIII.

II. Overall Framework of mwcEEGc

Fig. 1 presents the overall framework of the proposed MWC-based EEG clustering method, which includes the contents as follows.

1) *EEG Similarity:* Measures the correlations of pairwise EEG trials. Intuitively, EEG trials grouped into the same clusters are similar to each other, while those partitioned in different clusters have lower similarities. Moreover, the similarities weigh the correlations among EEG trials and construct the weighted graph. They also contribute to the importance of vertices (EEG trials) to clusters.
2) *EEG Graph Construction:* Forms a weighted undirected complete EEG graph, including vertex weight and edge weight, where EEG trials are transformed as vertices and any two of them are connected by an edge. Furthermore, a modified weight function is proposed based on vertex weight and edge weight simultaneously. Edge weight and vertex weight are two important parts for EEG clustering in this content, which will be introduced later.
3) *EEG Clustering:* Searches cliques from the weighted EEG graph with respect to similarity thresholds such that the total weight of all cliques is maximized. A vertex whose edge similarities to all the other vertices in the clique satisfy the similarity threshold is likely to join the clique based on whether the total weight of the new clique is larger or smaller than the former one. Then, this process is repeated until all the vertices are clustered while the total weight of all cliques is maximized. This step in the process leads to our algorithm, mwcEEGc, based on MWC.

III. Related Works

The early work using the graph theory is on the EEG community (electrode) clustering. Mammone *et al.* [17] proposed a hierarchical EEG electrode grouping method for graph connectivity density comparison, which applied the permutation Jaccard distance to measure the coupling strength between EEG signals from different electrodes and then provided a dissimilarity matrix of electrodes for partitioning EEG electrodes. Ozdemir *et al.* [18] also proposed a hierarchical consensus clustering method for partitioning EEG community structure, which constructed the connectivity matrices as the weighted undirected graphs on each subject and then partitioned EEG electrodes based on the spectral graph theory. These methods mainly focused on the relationships among EEG electrodes rather than the EEG trials combined with multichannel EEG signals, and they also ignored the correlations among EEG signals from the same electrode. Dai *et al.* [19] exploited MWC to select the valid EEG for classification. It conducted a different problem with our current work in this article. The work [19] was on EEG instance selection for EEG classification in a supervised way, where EEG labels are required and it is a common task. The present work in the article conducted EEG clustering in an unsupervised way that is seldom addressed and is a more challenging task because of its aim at unlabeled EEG. In detail, our work in this article focuses on EEG trial clustering through mapping it to searching MWCs such that the total weight is also maximized in a complete undirected similarity-weighted EEG graph, in a way that similar EEG trials are assigned into same clusters while dissimilar ones are separated into different clusters. In this section, we review

related works on EEG time-series clustering and similarity measures.

A. EEG Time-Series Clustering

With the continuous increase of unlabeled EEG signals, EEG clustering is becoming an important new technique for neurocognitive diagnoses and BCI applications. Unfortunately, there are few studies on clustering unlabeled EEG time series, such as k-means [20], [21] and the newest MTEEGC [22] which exploited an optimal objective function to search cluster centroid and then clustered EEG trials based on the cross-correlations between candidate EEG trials to the cluster centroid. But both of the two methods are influenced by the center/centroid initialization and they just consider the distance/similarity of EEG signals to the center/centroid, ignoring the correlations between other EEG signals in the same clusters. As a promising unsupervised analysis technique, time-series clustering is currently emerged out, such as 1) feature selection-based UDFS [23], NDFS [24], RUFS [25], and RSFS [26]: first extract/select features and then embed k-means strategy to cluster; 2) distance-based k-means++ [27], dynamic time warping (DTW) [28], k-DBA [29], and K-SC [30]: first randomly initialize or calculate cluster centers/centroid, and then cluster time series mainly based on their distances between candidates and the centers/centroid; 3) shape-based SCTS [31] and k-Shape [32]: first search time-series shapes and then cluster time series based on their shape similarities; and 4) shapelet-based USLM [34] and u-shapelet [33]: first calculate shapelets of time series and transform original time series to shapelet space, and then cluster time series with the shapelets. These methods have achieved good clustering results for conventional time-series data, but they are probably not applicable to cluster unlabeled EEG since compared to traditional time series, EEG has such characteristics as higher weakness, higher complexity, stronger oscillation, higher instability, higher dimension/multivariate, and lower signal-to-noise ratio. With the characteristics of EEG, it is probably: 1) difficult to learn optimal parameters to extract/select distinct features for feature selection-based methods; 2) hard to apply appropriate lengths to learn EEG shapes/shapelets for shape-/shapelet-based methods; and 3) not easy to exploit suitable distance measures to evaluate flexible similarities among EEG signals for distance-based methods. Furthermore, these methods need to calculate cluster centers based on an optimization function and they critically depend on the selection of initial cluster centers or the initial set of amount of clusters and selected features. Therefore, these conventional time-series clustering methods do not satisfy the following clustering properties: richness, consistency [35], and order independence [36], which will be discussed in Section VII. Besides, the experiments on 14 EEG datasets in Section VI also indicate that these conventional methods cannot achieve as good EEG clusters as our method.

B. Similarity Measures

Similarity weights contribute to representing the correlations among EEG signals in the EEG graph. Which similarity measure is the most appropriate for EEG? Several widely applied similarity measures are discussed in this section, including the Euclidean metric (ED) [37], [38]; DTW [29], [39]; the Hausdorff distance (HD) [40]; and FD [41]. ED does not correspond to the common notion of time series and it cannot capture flexible similarities of EEG time series since it requires the same length of sequences. DTW evaluates the similarity between time series by warping them in the time dimension, which outperforms ED [42], but it concentrates too much on minimizing the accumulation of all local distances among adjacent points while the FD emphasizes the overall distinction between series [43]. Meanwhile, FD is inherently independent of the sampling of curves, DTW does not work well when one of two curves is sampled less frequently [44]. Moreover, DTW does not satisfy the triangle inequality but the discrete FD does [45]. HD is sensitive to outliers [46], and it considers EEG as arbitrary point sets, which ignores point orders of EEG. Namely, HD measures distances just according to the nearest neighbor distances among points along the curves. It is likely to obtain a small HD but a large FD for two EEG signals [40], [41]. FD demands continuous and order-preserving assignments of points along the curves. It not only takes into account the location and points orders along the curves but also satisfies the triangle inequality, which theoretically makes it outperform ED, DTW, and HD in measuring EEG similarities, with respect to their intrinsic structure [19], [49]. Table I in detail shows their advantages and disadvantages, respectively, and Fig. 2 also illustrates their performance on EEG clustering. According to the suitability and superiority, we, therefore, applied FD as the similarity measure for EEG clustering in this article.

TABLE I
SIMILARITY MEASURES

Distance measure	Advantages	Disadvantages
ED	① efficient ② simple	① inflexible ② linear mapping ③ requires the same length of sequences
HD	① efficient ① non-linear mapping [47]	① treats sequences as sets, ignoring time orders of points [43] ② sensitive to outliers and noises [46], [47] ③ local alignment distance [46]
DTW	① considers time orders ② non-linear mapping [48]	① time consuming ② concentrates too much on minimizing the accumulation of all local distances between adjacent points [46] ③ regardless of phase difference between sequences [48] ④ violates the triangle inequality [45]
FD	① considers time orders ② non-linear mapping ③ emphasizes overall distinction between sequences [46] ④ satisfies triangle inequality [45]	① time consuming

IV. PRELIMINARIES

This section introduces the FD-based similarity measure and MWC, respectively, which are applied in the proposed EEG clustering approach.

A. Similarity Based on the Fréchet Distance

Originally, FD [50] is defined as the minimum length of a leash that a handler requires to walk a dog, where the dog

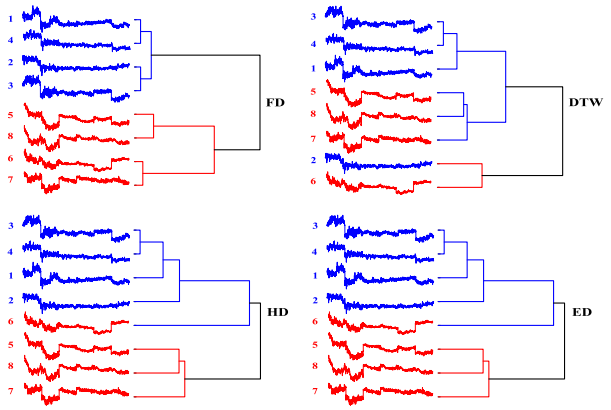

Fig. 2. EEG clustering with different distance measures (complete linkage and z-normalized). The blue and red colors, respectively, denote two-class EEG signals activated by two corresponding cerebral stimulations: moving a cursor up and moving a cursor down. Every class contains four trials with length of 5376.

walks monotonically along ϕ according to θ and the handler walks monotonically along ψ according to π. In this analogy, FD is the shortest possible leash admitting such a walk. Mathematically, given two reparameterizations θ and π such that $\theta, \pi : [0, 1] \to [0, 1]$ and $\theta(0) = 0, \theta(1) = 1$; $\pi(0) = 0$, $\pi(1) = 1$ for two curves (EEG signals) ϕ and ψ, respectively, define the width between ϕ and ψ as follows:

$$\text{width}_{\theta,\pi}(\phi, \psi) = \max_{q\in[0,1]} \|\phi(\theta(q)) - \psi(\pi(q))\| \tag{1}$$

where $||\cdot||$ is the underlying norm.

Formally, two curves (EEG trials) ϕ and ψ in $\mathbb{R}^d$, the FD between ϕ and ψ is defined as

$$\delta_{\mathcal{F}}(\phi, \psi) = \inf_{\theta,\pi\in[0,1]} \text{width}_{\theta,\pi}(\phi, \psi) \tag{2}$$

where θ and π are two orientation preserving reparameterizations of two curves ϕ and ψ.

B. Maximum-Weight Clique

The maximum-weight clique problem (MWCP) [51], without loss of generality, is a technique that searches the clique (i.e., a subgraph whose vertices are pairwise connected by a weighted edge) in a weighted graph, such that the weight of vertices and edges is maximized. In practice, MWCP is applied to search instances with specifically similar/same properties or structures, such as searching communities, networks, or protein structure analysis, etc. Given an undirected weighted graph $G = (V, E, \eta, \mu)$, where V and E define the vertex set and edge set, respectively; $\eta : V \to \{0\} \cup \mathbb{R}^+$ and $\mu : E \to \{0\} \cup \mathbb{R}^+$ denote correspondingly vertex weights and edge weights. $\sum_{v\in V} \eta_v + \sum_{e\in E} \mu_e$ is the weight of G.

Define $\mathbb{N}_n = \{1, \ldots, n\}, n = |V|$, and $e_{ij} = \{i, j\} \in E$, then MWCP aims to find a maximum-weight clique C in the given graph G [see (3)]. When the vertex v_i or edge e_{ij} (i and j of e_{ij} denotes vertex v_i and v_j, respectively) is selected into a clique, $v_i = 1, v_j = 1$, and $e_{ij} = 1$. Otherwise, $v_i = 0, v_j = 0$, and $e_{ij} = 0$. Therefore, $v_i, v_j, e_{ij} \in \{0, 1\}$, see the right of indicator

$$\sum_{i=1}^{n} \eta_i v_i + \sum_{1\le i<j\le n} \mu_{ij} e_{ij} \to \max_{v_i,e_{ij}\in\{0,1\}}. \tag{3}$$

In existing methods for solving MWCP [52], either the weight of vertex [53] or the weight of edge [54] is considered, but not both. Especially, when only considering the weights of vertices or the edge weights are transformed into vertex weights, the MWCP is correspondingly transformed to the maximum-clique problem (MCP) which searches a complete subgraph with maximum cardinality. In the case, the MWCP can be equivalently interpreted by a binary model which is formulated as (4) shows based on [55]

$$\begin{aligned} \max\ & f(x) = \sum_{i=1}^{n} w_i x_i \\ \text{s.t.}\ & x_i + x_j \le 1 \quad \forall\{v_i, v_j\} \in \overline{E} \\ & x_i \in \{0, 1\}, \quad i \in \mathbb{N}_n \end{aligned} \tag{4}$$

where $\mathbb{N}_n = \{1, \ldots, n\}, n = |V|$, w_i is the weight of v_i; x_i denotes the binary variable associated to vertex v_i; and $\overline{E}$ defines the edge set of complementary graph $\overline{G}$.

V. MWCEEGC FOR EEG CLUSTERING

The proposed method mwcEEGc clusters EEG trials via searching MWCs in an improved Fréchet similarity-weighted EEG graph. More important, mwcEEGc considers the weights of both vertices and edges to search MWCs.

A. Fundamental Support

Intuitively, the contour shapes of EEGs stimulated by the same specific cerebral activity are assumed to be highly similar to each other and different from those activated by different cerebral activities, which provides the fundamental support for mwcEEGc to cluster EEGs. According to the basic knowledge, the proposed method mwcEEGc clusters highly similar EEGs into a clique with tight intracluster compactness and large intercluster scatter, by searching MWCs in an improved Fréchet similarity-weighted EEG graph.

In detail, for activating a specific cerebral activity, the majority of EEG signals are similar to each other, only a few ones (invalid ones generated by nonspecific cerebral activities) are not. Namely, for stimulating a specific cerebral activity, the probability of getting the majority of similar EEG signals is larger than that of getting the majority of dissimilar ones (i.e., minority of similar ones).

Theorem 1: Stimulating a cerebral activity A, the probability of getting similar EEG is p, correspondingly getting dissimilar one is $1-p$. When $0 \le 1-p \le p$, activating A for $n_{\ge 2}$ times, let $P(E_{k\ge(n/2)})$ be the probability of getting equal or more than $k_{\ge(n/2)}$ similar EEG signals and $P(E_{k\le(n/2)})$ be the probability of getting equal or less than $k_{\le(n/2)}$ similar ones, then $P(E_{k\ge(n/2)}) \ge P(E_{k\le(n/2)})$. (The proof is shown in Section S-I of the supplementary file.)

B. Edge Weight of EEG Graph

The edge weight of EEG graph $G = (V, E, \eta, \mu)$ represented by EEG similarities determines to cut edges that do not satisfy threshold when partitioning EEG graph (i.e., EEG clustering). As a promising similarity measure, the FD is utilized in the work. In fact, the conventional FD (CFD) mainly measures the global trends but neglects the local structure [56]. In order to improve the similarity measure and balance the sensitivity of global similarity, we use local tendency to improve the CFD. Mathematically, let $tr_i = (a_1, \ldots, a_h)$, $tr_j = (b_1, \ldots, b_l)$ be two EEG signals in $\boldsymbol{Tri}$ (i.e., set of EEG trials), the local tendency of tr_i, tr_j is computed by

$$\mathrm{Loc}T\big(tr_i, tr_j\big) = \frac{\sum_{p=1}^{o-r}\big(a_{p+r} - a_p\big)\big(b_{p+r} - b_p\big)}{\sqrt{\sum_{p=1}^{o-r}\big(a_{p+r} - a_p\big)^2 \sum_{p=1}^{o-r}\big(b_{p+r} - b_p\big)^2}} \tag{5}$$

where r such that $1 \le r < o(o = \min\{h, l\})$ defines the length of the EEG subsequence. With $a_{p+r} - a_p$ in (5), r points are selected to measure the local tendency and commonly $r = 1$, since a larger $r \ge 2$ probably results in that more local tendencies of smaller EEG segments with lengths $\le r$ are ignored. Besides, $LocT(tr_i, tr_j) \in [-1, 1]$, where the positive values indicate two EEG have more similar local tendencies and the larger the value is, more similar local tendencies the two EEG have, otherwise the two EEG have more opposite local tendencies. Then, the improved similarity is calculated by

$$s_{ij} = \lambda \cdot \delta_{\mathcal{F}}\big(tr_i, tr_j\big) + (1 - \lambda)\frac{1 - \mathrm{Loc}T\big(tr_i, tr_j\big)}{2} \tag{6}$$

where $\delta_F(tr_i, tr_j)$ is the CFD similarity and $\lambda \in [0, 1]$. Commonly, $\lambda = 0.5$, to balance the global similarity and local tendency.

With the local and global trends, the normalized similarities of n EEG trials build the diagonal similarity matrix $\boldsymbol{S}^{n\times n}$, named edge weight matrix $\boldsymbol{\mu}$, as shown in

$$\boldsymbol{\mu} = \boldsymbol{S}^{n\times n} = \begin{pmatrix} 1 & \ldots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \ldots & 1 \end{pmatrix} \tag{7}$$

where $s_{ji} = s_{ij}$, and i, j denote EEG $tr_i, tr_j \in \boldsymbol{Tri}$.

Additionally, Fig. 3 establishes the superiority of the improved FD (IFD) over the conventional one (CFD) for measuring EEG similarities.

C. Vertex Weight of EEG Graph

The vertex weight of the EEG graph $G = (V, E, \eta, \mu)$ reflects the significance of vertices to MWC that are being searched. Namely, together with edge weights, it determines which vertices are assigned together into a same clique (cluster). In other words, EEG vertices with higher η_i are potentially clustered in a same clique. For n EEG trials in the EEG graph $G = (V, E, \eta, \mu)$, the importance to the being-searched clique is denoted by a partially ordered similarity matrix $\boldsymbol{\eta}^{n\times 1}$ such that $< \boldsymbol{\eta}, \preccurlyeq_\eta >= \{< \eta_i, \eta_j > | \ \eta_i \ge \eta_j; i \ne j; i, j \in \boldsymbol{Tri}\}$,

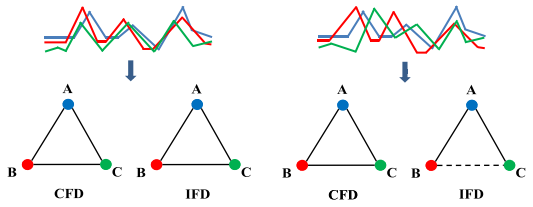

Fig. 3. Comparison of IFD and CFD. According to CFD, the contours of EEG B (red line) and C (green) are close to A (blue) in both (a) and (b) whose similarities satisfy the threshold, so a 3-vertex clique is built. Whereas with IFD, although B and C are, respectively, similar to A from the global closeness, the local tendency between B and C is low, so with the local and global measures, the IFD reveals that the similarity between B and C does not satisfy the threshold (dotted line), hence only a 2-vertex clique of B and A or C and A is formed, as (b) shows.

where η_i is defined as (8), which is also called vertex weight

$$\eta_i = \frac{1}{|\ \boldsymbol{Tri}\ | - 1}\sum_{j\in \boldsymbol{Tri}\backslash i} s_{ij}. \tag{8}$$

In particular, when $|\boldsymbol{Tri}| = 1$, then $\eta = 0$. η_i correspondingly indicates the rankings of objective EEG trial tr_i to the rest of EEG trials $tr_j \in \boldsymbol{Tri}\backslash tr_i$. In other words, when tr_i is selected into clique C, a vertex $tr_j \in \{tr_1, \ldots, tr_k\}$ with the highest rank to tr_i, according to vertex weight $\boldsymbol{\eta}$, is correspondingly highly likely clustered into the same clique C.

D. Formulation of EEG Clustering

Unlike the methods that only have edge weight or vertex weight, our method simultaneously considers edge weight and vertex weight to cluster EEG since the two weights jointly together: 1) measure the correlations between any two EEG trials, which contributes to intracluster compactness and intercluster scatter and 2) achieve same clusters without randomly initializing cluster centers/centroid, which satisfies order independence property [36]. Simultaneously considering the edge weight matrix $\boldsymbol{\mu}^{n\times n}$ and vertex weight matrix $\boldsymbol{\eta}^{n\times 1}$, the pairwise highly weighted EEG signals clustered into the same clique C can be represented by (9). Based on (9), the vertex with large value $c_{v_t} \in \boldsymbol{C}$ achieved by $\boldsymbol{\eta}^T\boldsymbol{\mu}$ is most likely selected into the clique to construct a larger-weight clique, see (10). With (9), the EEG clustering method can achieve excellent results with high intracluster compactness and intercluster scatter, without randomly initializing cluster centers or calculating centroid

$$\boldsymbol{C} = \boldsymbol{\eta}^T\boldsymbol{\mu}. \tag{9}$$

Correspondingly, the vertex v_t with largest $c_{v_t} \in \boldsymbol{C}$ can be chosen as the next potential candidate into the clique

$$v_t = \arg\max_{v_t}\big\{c_{v_t} \in \boldsymbol{C}\big\}. \tag{10}$$

E. mwcEEGc Algorithm

Given an improved Fréchet similarity-weighted EEG graph $G_W = (V, E, \eta, \mu)$, where $\eta : V \to \{0\} \cup \mathbb{R}^+$ and $\mu : E \to \{0\} \cup \mathbb{R}^+$ represent weights of the vertices (EEG trials) and edges, respectively, the proposed mwcEEGc algorithm aims

to find a family $\mathcal{C} = \{C_1, \ldots, C_m\}$ of $m(\geq 2)$ disjoint cliques that maximizes $\sum_{C_k \in \mathcal{C}} w_{C_k}$, whose cardinalities are denoted by positive integers $N_1, \ldots, N_m$ satisfy $\sum_{i=1}^{m} N_i \leq n, n \leq |V|$.

For any edge $(i, j) \in E(C_k)$, $i, j \in V(C_k)$, $k \in \{1, \ldots, m\}$ and the weight of edge in C_k : $\mu_{ij} \in \boldsymbol{\mu}, \boldsymbol{\mu} = \{s_{ij} | s_{ij} \geq \delta_{C_k}\}$, the weight function of the kth clique C_k simultaneously considering edge weight and vertex weight is modified as (11) defines, where similarity threshold δ_{C_k} determines the results of EEG clustering (the kth clique C_k searching) since vertices whose edge weights are larger than δ_{C_k} are potentially partitioned into C_k. The threshold $\boldsymbol{\delta}$ selection for EEG clustering will be discussed latter

$$w_{ij} = \begin{cases} \frac{\eta_i + \eta_j}{N_k - 1} + \mu_{ij}, & \text{if } \mu_{ij} \geq \delta_{C_k}, i \neq j \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Proposition 1: The mwcEEGc for EEG clustering with the modified weight function (11) as well as the descending thresholds $\boldsymbol{\delta} = \{\delta_0 = 1, \delta_1, \ldots, \delta_{m-1}, \delta_m = 0\}$ can be equivalently written as follows:

$$\mathcal{F}(\mathcal{C}) = \sum_{k=1}^{m} F(C_k) = \sum_{k=1}^{m} \sum_{\substack{e \in E(C_k) \\ \delta_k \leq \mu_e < \delta_{k-1}}} w_e \to \max_{\mathcal{C} \in \mathscr{C}}$$

or

$$\mathcal{F}(\mathcal{C}) = \frac{1}{2} \sum_{k=1}^{m} \sum_{i \in C_k} \sum_{\substack{j \in C_k \\ \delta_k \leq \mu_{ij} < \delta_{k-1}}} w_{ij} \to \max_{\mathcal{C} \in \mathscr{C}}$$

where $\mathscr{C} = \{\mathcal{C} = \{C_1, \ldots, C_m\} : C_i \cap C_j = \varnothing;\ |C_i| = N_i, |C_j| = N_j;\ \{i, j\} \in \mathbb{N}_m\}$.

The proof of Proposition 1 is shown in Section S-II of the supplementary file. $\delta_0 = 1, \delta_1, \ldots, \delta_k, \ldots, \delta_m = 0$ in a descending order correspondingly denote the similarity threshold of the default nonvertex clique (i.e., $C_{\delta_0} = \varnothing$), 1st, ..., mth clique, respectively, and especially $\delta_0 = 1$ to limit the other k thresholds smaller than the upper bound of 1. With thresholds $\boldsymbol{\delta} = \{\delta_0 = 1, \delta_1, \ldots, \delta_{m-1}, \delta_m = 0\}$, highly similar vertices, whose connected edges' weights are all higher than the corresponding δ_k, will be clustered into a same clique C_k. Searching m clusters requires $m-1$ thresholds, since obviously $m-1$ thresholds lead to $m-1$ cliques, and then the remaining EEG vertices whose similarities satisfy $0 \leq \mu_v \leq \delta_{m-1}$ are spontaneously regarded as the mth clique.

A vertex v_t joins the clique C when it satisfies two conditions simultaneously: 1) $\forall v_j \in C$, $\mu_{tj} \geq \delta$ and 2) $\sum_{v_i, v_j \in C \cup \{v_t\}} w_{ij} \geq \sum_{v_i, v_j \in C} w_{ij}$. Actually, once $\boldsymbol{\delta}$ is set, adding v_t to C is only required to satisfy condition 1).

Theorem 2: A vertex v_t is selected into the clique C to obtain the clique C^* whose weight is larger than C, if and only if $\forall v_i \in C, \mu_{it} \geq \delta_C$, where δ_C denotes the similarity threshold of clique C.

Based on Theorem 2 (see Section S-III of the supplementary file for the proof), once $\delta_{k=1,\ldots,m-1}$ is set, mwcEEGc searching cliques with $\max_{\mathcal{C} \in \mathscr{C}}$ is eventually transformed to $(m-1)$-searching MWCs. In other words, performing $(m_{\geq 1}, \delta_{k=1,\ldots,m})$-mwcEEGc is equivalently transformed to repeating $(1, \delta_k)$-mwcEEGc for $m-1$ times with $\delta_k, k = 1, \ldots, m$, especially $\delta_0 = 1$ and $\delta_m = 0$.

Algorithm 1 mwcEEGc

Input:
$\boldsymbol{\delta}$: Similarity threshold set, $\boldsymbol{\delta} = \{\delta_k | 0 \leq \delta_k < \delta_{k-1};\ \delta_0 = 1, \delta_m = 0, k = 1, \ldots, m_{\geq 1}\}$;
Output:
$\mathcal{C}$: Clique set such that $\max_{\mathcal{C} \in \mathscr{C}} \mathcal{F}(\mathcal{C})$;

```
Initialize G = (V, E, η, μ) with (7) and (8), k = 1,
C_1 = η_1^T μ_1 = η^T μ, C = ∅;
repeat
    δ_k ∈ δ;
    C_k = η_k^T μ_k;
    C_k = {v_1 | (v_1 with the maximum value in C_k) ∈ V};
    V = V\{v_1};
    for V ≠ ∅ do
        {v_t | v_t with the t_{≥2}^{th} largest value in C_k} ∈ V;
        if ∀{v_n} ∈ V_{C_k}, δ_k ≤ μ_{v_t v_n} < δ_{k-1} then
            V_{C_k} = V_{C_k} ∪ {v_t};
            N_k = |V_{C_k}| + 1;
            Update w_ij with (11);
            W_{C_k ∪ {v_t}} = Σ_{i,j ∈ V_{C_k} ∪ {v_t}} w_ij;
        end if
        if W_{C_k ∪ {v_t}} ≥ W_{C_k} then
            C_k = V_{C_k};
            V = V\{v_t};
            W_{C_k} = W_{C_k ∪ {v_t}};
        end if
        t = t + 1;
    end for
    C = C ∪ C_k;
    k = k + 1;
    η_k = {η_i | v_i ∉ V_{C_{k-1}}};
    μ_k = {μ_ij | v_i, v_j ∉ V_{C_{k-1}}};
until δ\{δ_0} = ∅;
```

Proposition 2: With Theorem 2, $(m_{\geq 1}, \delta_{k=1,\ldots,m})$-mwcEEGc is equivalently transformed to $(m-1)$-time repeating performing $(1, \delta_k)$-mwcEEGc with $0 \leq \delta_k < \delta_{k-1}, \delta_0 = 1$ and $\delta_m = 0$.

The proof of Proposition 2 is elaborated in Section S-IV of the supplementary file. Based on Proposition 2, the mwcEEGc is shown in Algorithm 1, which clusters EEG without randomly selecting cluster centers or computing cluster centroids. $\boldsymbol{C}_k = \boldsymbol{\eta}_k \boldsymbol{\mu}_k$ in Algorithm 1 also shows the importance of remaining vertices $v \in V \backslash V_{C_k}$ after the kth clique is identified, and it is also used to assign the next potential vertex with the largest value into the clique. With $\boldsymbol{C} = \boldsymbol{\eta}\boldsymbol{\mu}$, the ranking $\eta_i \in \boldsymbol{\eta}$ also reduces the time consumption mainly because cliques can be achieved without computing cluster centers/centroids. Besides, Fig. 4 also briefly illustrates the process of Algorithm 1.

F. Complexity Analysis of mwcEEGc

The mwcEEGc approach mainly contains two phases: 1) weighted EEG graph construction and 2) EEG clustering, both of which contribute to the time consumption of mwcEEGc.

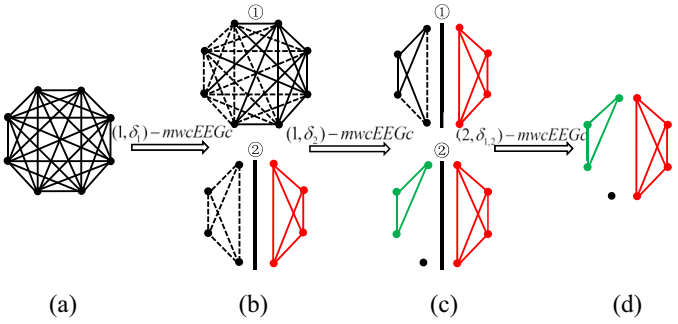

Fig. 4. Illustration of mwcEEGc to search three maximum weighted EEG cliques ($\delta_{\{1,2\}}$-mwcEEGc). (a) Original undirected weighted complete graph constructed by EEG signals and their similarities: $G = (V, E, \eta, \mu)$. (b) ① search the vertices and edges (solid black line) whose edge weights satisfy $\delta_1 \leq \mu_{ij} < 1$, ② run mwcEEGc with δ_1 and search the first maximum weighted clique (red) $(C_1 = (V_1, E_1, \eta, \mu))$ such that $\max \sum_{i,j\in C_1} w_{ij}$ according to ① and construct its difference graph (black): $G = (V_D, E_D, \eta, \mu) : V_D = \{x, y | x, y \notin V_1; x, y \in V\}, E_D = \{(x, y) | x, y \in V_D\}$. (c) ① similarly, search the edges and their adjacent vertices such that $\delta_2 \leq \mu_{xy} < \delta_1$ from the G_D, ② repeat running the mwcEEGc with δ_2 to search the second maximum weighted clique C_2 (green) such that $\max \sum_{x,y\in C_2} w_{xy}$. (d) Final three maximum weighted cliques with largest total weights satisfying $\delta_c, c = \{1, 2\}$ and $\max \sum_{c=\{1,2\}} \sum_{i,j\in C_c} w_{ij}$. Clearly, two similarity thresholds δ_1, δ_2 result in three EEG cliques, with the third clique being a single vertex (black) which is formed by the remaining EEG signals after green and red cliques are found.

TABLE II
EEG DATASET DETAILS

EEG Data	Descriptions of EEG data sets	# of EEGs and Length	# of Channels	# of Classes
II_Ia	SCPs of a healthy subject	268× 5377	6	2
II_Ib	SCPs of an ALS patient	200×8065	7	2
II_Ia_Ib	Mixed SCPs of a healthy subject and an ALS patient	468×5377	6	4
III_V_s1	Mental imagery with precomputed	3488×97	8	3
III_V_s2	features of 3 healthy subjects s1, s2, s3	3472×97		
III_V_s3	(left hand, right hand, word association)	3424×97		
IV_2a_s1	Multi-class motor imagery of	288×6887	22	4
IV_2a_s2	3 healthy subjects s1, s2, s3 (left	288×6887		
IV_2a_s3	hand, right hand, both feet, tongue)	288×6887		
IV_2b_s1	Motor imagery of 3	120×940	3	2
IV_2b_s2	healthy subjects s1, s2, and s3	120×940		
IV_2b_s3	(left hand, right hand)	120×940		
IV_3_s1	Hand movement of 2 healthy subjects	160×4001	10	4
IV_3_s2	s1 and s2 (left, right, forward, backward)	160×4001		

In weighted EEG graph construction, the time cost is primarily determined by the Fréchet-based similarity computation of EEG trials. The conventional algorithm proposed in [57] to compute the Fréchet similarity of n EEGs with length of l has a runtime of $O([(n(n-1))/2] \cdot l^2 \log l^2) = O(n(n-1) \cdot l^2 \log l)$ based on the symmetry of the Fréchet similarity of pairwise EEG trials, where $n \ll l$ commonly. With the development in recent decades, the computation for the Fréchet similarity is improved in a subquadratic time $O(l^2[(\log \log l)/(\log l)])$ as introduced in [41], which is used to compute the FD of EEGs in this article to measure EEG similarities and construct the weighted EEG graph. Hence, the real runtime for computing the Fréchet similarity for n EEG trials is $O([(n(n-1))/2] \cdot l^2[(\log \log l)/(\log l)]) \approx O(n^2 \cdot l^2[(\log \log l)/(\log l)])$.

In Algorithm 1, with the Fréchet-based EEG similarities, the mwcEEGc aims to group n EEG trials from the constructed weighted EEG graph into m clusters with respect to similarity thresholds $\boldsymbol{\delta} = \{\delta_1, \ldots, \delta_{m-1}\}$, such that the total weight of m clusters is maximized. In each iteration, computing $\boldsymbol{C}_k$ and seeking the vertex with the largest value in $\boldsymbol{C}_k$ requires $O((n - n_k)^2) < O(n^2)$ as lines 4 and 5 show, where n_k is the number of vertices of clique C_k with constraint of $\sum_{k=1}^{m} = n$. Then, searching clique mainly requires checking the edge weights with respect to $\boldsymbol{\delta}$, updating the modified weights, and summarizing the total weight of clique $C_k \cup \{v_t\}$ together cost $O((n - n_k + 1)(n - n_k)) < O(n^2)$ (see lines 7–21). Finally, to update the edge weight matrix and vertex weight matrix in lines 24 and 25 requires $O([(n - n_k)(n - n_k - 1)]/2) < O(n^2/2)$. In a result, the time complexity of mwcEEGc is $O(\sum_{k=1}^{m}((n - n_k)^2 + (n - n_k)(n - n_k + 1)(n - n_k) + [((n - n_k)(n - n_k - 1))/2])) < O(mn^2 + mn^3) \approx O(mn^3 \log n)$. Thus, the time complexity of mwcEEGc to cluster EEG trials is bounded from above by the polynomial time $O(mn^3)$.

In overall, considering the computation of the improved Fréchet-based similarities and EEG clustering, the total time consumption of mwcEEGc is $O(\max\{n^2 \cdot l^2[(\log \log l)/(\log l)], mn^3\})$.

VI. EXPERIMENTS

We first introduce the details of the EEG datasets, evaluation methodology, and baseline methods in the section. Then, we conduct the experiments and discuss the impact of key factors on mwcEEGc.

A. EEG Datasets

Fourteen EEG datasets shown in Table II are used to evaluate the efficacy of mwcEEGc, which includes the slow cortical potentials (SCPs), mental imagery EEG, motor imagery EEG, and hand movement EEG. Besides, all the original EEG data and their detailed descriptions are publicly available as online archives at https://github.com/Jackie-Day/EEG-data-and-descriptions. Importantly, as an essential preprocessing technique that assists data mining algorithms to concentrate on the structural similarities or dissimilarities rather than the amplitude-driven ones, z-normalization is first used to preprocess the EEG data before they are used to cluster in the paper. Moreover, we selected the specific target EEGs recordings activated by the specific cerebral activities by deleting such nontarget and unspecific EEGs in relaxing or in idle or in preparing period. Moreover, as these original EEG data were labeled, we first removed their labels for mwcEEGc, and then they are also adopted to evaluate the performance of mwcEEGc.

B. Evaluation Methodology

Six evaluation criteria of *intracluster compactness* (S_{In}), *intercluster scatter* (S_{Be}), *integrated ratio* (r), *rand index* (RI), *Fleiss's kappa* (κ), and *F-score* are utilized to measure the performance of EEG clustering algorithms.

1) *Intracluster Compactness (S_{In}):*

$$S_{In} = \frac{1}{\mid C \mid} \sum_{C_t \in C} \left(\frac{1}{\sqrt{\mid C_t \mid}} \left\| \boldsymbol{w}^t - \mu_t \right\|_2 \right) \tag{12}$$

where $C_t \in C$ denotes the t-th cluster, μ_t denotes the mean weight of C_t, and $\boldsymbol{w}^t$ indicates the weight matrix of C_t. S_{In} measures the similarities of pairwise EEG in the same cluster. A smaller S_{In} indicates higher holistic compactness within all the EEG trials in the same cluster.

2) *Intercluster Scatter (S_{Be}):*

$$S_{Be} = \frac{1}{\mid C \mid} \sum_{C_t, C_s \in C} \left(\frac{1}{\sqrt{\mid C_t \mid\mid C_s \mid}} \sum_{t \neq s} \times \left(\boldsymbol{w}^{ts} - \mu_t \right) \left(\boldsymbol{w}^{ts} - \mu_s \right)^T \right)^{\frac{1}{2}} \tag{13}$$

where $C_t, C_s \in C$ denotes two clusters such that $C_t \cap C_s = \varnothing$; and $\boldsymbol{w}^{ts}$ denotes the weight matrix between C_t and C_s. S_{Be} measures the scatter among clusters, that is, it evaluates the separation degree of clusters. Further, the larger the S_{Be} is, the higher is the scatter between clusters.

3) *Integrated Ratio (r):* An integrated evaluation ratio γ simultaneously considering S_{In} and S_{Be} is proposed. Mathematically, $\gamma = ([S_{Be}]/[S_{In}])$. A higher γ reflects a qualitatively better clustering performance.

4) *Rand index (RI) [58]* Evaluates clustering quality based on the number of correct classes that the method clusters. Namely, it denotes the percentage of correct assignments achieved by methods. In detail, $RI = [(TP + TN)/(TP + TN + FP + FN)]$, where TP stands for the amount of true positives; false positives (FPs); TN, true negatives; and FN defines the amount of false negatives.

5) *F-Score [59]* is a modified measure of RI that regards unequally FP and false negative (FN) by setting a scale factor $\beta \geq 0$ on *recall*, generally $\beta = 1$. In the end, *F-score*$= [((1 + \beta^2)pr)/(\beta^2 p + r)]$, where *precision*: $p = [TP/(TP + FP)]$ and *recall*: $r = [TP/(TP + FN)]$.

6) *Fleiss' kappa (κ) [60]* measures the coherence of decision ratings among different classes. $\kappa = [(\overline{P} - \overline{P}_e)/(1 - \overline{P}_e)]$, where $\overline{P} - \overline{P_e}$ reflects the agreement degree of actually achieved over chance; $1 - \overline{P}_e$ means the agreement degree of attainable above chance. In addition, $\overline{P} = [1/(Nn(n-1))](\sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij}^2 - Nn)$, $\overline{P}_e = \sum_{j=1}^{k} ([1/Nn] \sum_{i=1}^{N} n_{ij})^2$, where N, n, and k define the amount of subjects, the amount of ratings per subject, and the amount of classes, respectively.

In a word, the higher the RI, κ, and *F-score* are, the better the quality of clustering methods achieve.

C. Baseline Methods

Aim to verify the efficacy of mwcEEGc on unlabeled EEG clustering, we compare it to ten state-of-the-art EEG or time-series clustering algorithms. These baselines can be mainly categorized into four types: 1) *Classic Clustering: k*-means++; 2) *Feature Selection-Based Clustering Embedded With k-Means:* UDFS, NDFS, RUFS, and RSFS; 3) *Distance-Based Clustering:* K-SC and *k*-DBA; and 4) *Shape/Shapelet-Based Clustering: k*-Shape, USLM, and MTEEGC.

k-Means++ [27]: It specifies a procedure with the probability to initialize cluster centers and then it performs the standard *k*-means to cluster EEG.

UDFS [23]: Unsupervised discriminative time-series clustering that simultaneously explores local discriminative information and the manifold structure to cluster time series.

NDFS [24]: Non-negative discriminative time-series clustering that combines non-negative spectral analysis and $l_{2,1}$-norm minimization regularization as an integrated structure to select unsupervised time-series features and cluster.

RUFS [25]: Robust unsupervised time-series clustering that utilizes locally learning regularized robust orthogonal non-negative matrix factorization to perform robust label learning and $l_{2,1}$-norms minimization to jointly cluster time series.

RSFS [26]: Robust spectral learning for unsupervised time-series clustering that combines sparse spectral regression and the robustness of graph embedding.

k-DBA [29]: It defines clusters using *k*-means and DTW and then uses an averaging strategy: DTW barycenter averaging (DBA) for centroid computation.

K-SC [30]: Similar to *k*-means, it uses a scale-and-shift-invariance measure to compute similarities of time series and extract cluster centroid according to the spectral norm of the similarity matrix.

k-Shape [32]: It clusters time series by seeking time-series shapes with a normalized version of cross-correlation, and it also relies on a scalable iterative refinement procedure.

USLM [34]: Unsupervised shapelet learning model efficiently learns shapelets to cluster time series, which integrates pseudoclass label learning, spectral analysis, shapelet regularization, and regularized least-squares minimization.

MTEEGC [22]: Multitrial EEG clustering is recently proposed by Dai *et al.*, which utilizes an improved cross-correlation-based feature extraction function to seek EEG shapes as cluster centroid and then clusters EEG trials with the shape centroid.

D. Experimental Results and Discussion

We analyzed mwcEEGc by comparing it to ten state-of-the-art clustering algorithms introduced above. All algorithms are operated with MATLAB R2014b, on a Windows 10 machine with 4×3.30-GHz CPUs and 16-GB memory. The parameters for ten algorithms are set, respectively, as same as in their corresponding references. Besides, all the algorithms are operated ten times and the reported results are the mean of these ten runs.

1) Selection of Similarity Threshold δ: In mwcEEGc, $\boldsymbol{\delta}$ influences the quality and the balance of clusters. If $\boldsymbol{\delta}$ is too small, it is difficult for mwcEEGc to differentiate clusters. If $\boldsymbol{\delta}$ is too large, mwcEEGc partitions too unique EEG trials into clusters. Furthermore, too small or too large $\boldsymbol{\delta}$ seems to result in unbalanced clusters. In our method, we set optimal $\boldsymbol{\delta}$

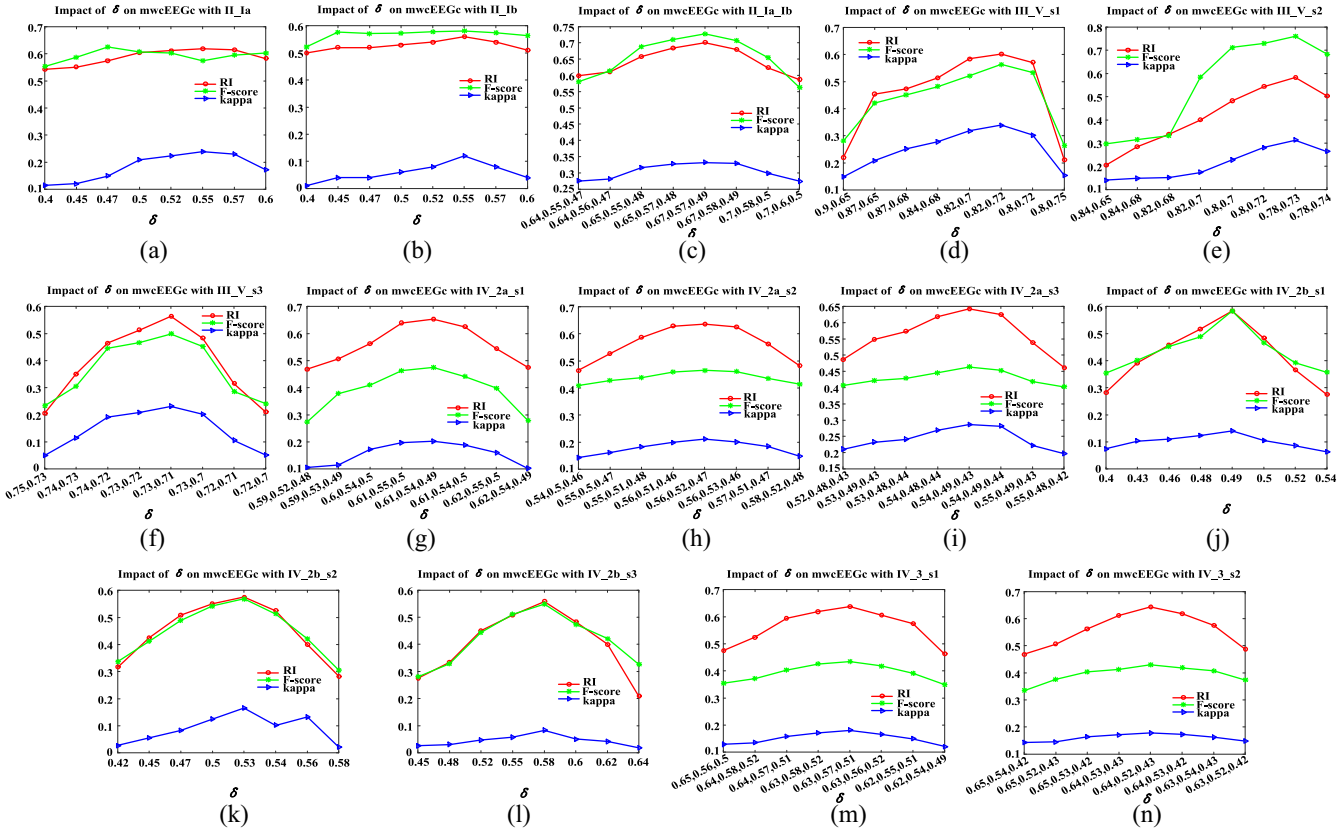

Fig. 5. Impact of $\boldsymbol{\delta}$ on mwcEEGc with 14 EEG datasets. Several groups of $\boldsymbol{\delta}$ are used to evaluate the clustering performance of mwcEEGc. (a) II_Ia. (b) II_Ib. (c) II_Ia_Ib. (d) III_V_s1. (e) III_V_s2. (f) III_V_s3. (g) IV_2a_s1. (h) IV_2a_s2. (i) IV_2a_s3. (j) IV_2b_s1. (k) IV_2b_s2. (l) IV_2b_s3. (m) IV_3_s1. (n) IV_3_s2.

according to the discrete probability distribution of EEG similarities. In detail, the discrete probability distribution $d(\boldsymbol{\delta})$ of EEG similarities is defined as follows:

$$d(\boldsymbol{\delta}) = \frac{n_{s\in[\delta_i,\delta_i+\text{Interval}]}}{N^2} \tag{14}$$

where N defines the amount of EEG trials in the Fréchet similarity-weighted EEG graph G, and n indicates the amount of similarities $s \in \boldsymbol{S}$ that stay in the interval $[\delta_i, \delta_i + \text{Interval}]$.

With (14), $\boldsymbol{\delta}$ can be chosen based on

$$\boldsymbol{\delta} = f^{-1}(D) \in \left\{ \boldsymbol{\delta} | f(\boldsymbol{\delta}): \sum_{i\in\left[0,\frac{1}{\text{Interval}}-1\right]}^{\frac{1}{\text{Interval}}} d_i(\boldsymbol{\delta}) = A \right\} \tag{15}$$

where $A \in (0, 1)$. To achieve balanced clusters, $\Delta A_i = |A_i - A_j|$, $i \neq j$ should be set in slight differences.

Figure S-1 in Section S-V of the supplementary file illustrates EEG similarity matrices of 14 EEG datasets and their discrete probability distributions, respectively, which shows that most EEG trials from the same class are similar to each other and dissimilar with those from different classes. Figure S-1 also indicates that the EEG trials stimulated by the same cerebral activity from the same subject are highly similar to each other while different cerebral activities or different subjects seem to produce dissimilar EEG trials. In addition, Fig. S-1 provides an illustration of the proof of Theorem 1 and it also presents the percentages of EEG amounts under different similarity thresholds δ, with which mwcEEGc can set proper similarity thresholds $\delta_{k=1,\ldots,m,m\geq 1}$ to achieve high-quality EEG clusters.

Since the number of clusters of the EEG datasets is fixed, we just analyzed mwcEEGc with respect to $\boldsymbol{\delta}$ that controls intracluster compactness, intercluster scatter, and the balance of clusters. The impact of similarity thresholds $\boldsymbol{\delta}$ on mwcEEGc quantified with *RI*, *F-score*, and κ is shown in Fig. 5. We can see that a moderate similarity threshold $\boldsymbol{\delta}$ obtains a better clustering result than a smaller or a larger one. In other words, moderate $\boldsymbol{\delta}$ that is set based on (15) achieves high-quality EEG clusters.

2) EEG Clustering Performance Analysis: We compared mwcEEGc to ten baselines for EEG clustering. Particularly, we, based on Fig. 5, set the optimal $\boldsymbol{\delta}$ for mwcEEGc. The experimental results on 14 EEG datasets are shown in Tables III–XVI, respectively. Accordingly, they clearly demonstrate that mwcEEGc yields the best clusters according to *RI*, *F-score*, and κ, indicating the superiority of mwcEEGc over the ten baselines for EEG clustering.

We also compared mwcEEGc to the ten state-of-the-art baselines with respect to intracluster compactness and intercluster scatter on the 14 EEG datasets. The comparisons are also presented in Tables III–XVI, which indicates that mwcEEGc achieves the best clusters with respect to intracluster compactness and intercluster scatter. Although for all the 14 EEG datasets, mwcEEGc does not achieve the best S_{In}

TABLE III
CLUSTERING RESULTS WITH THE EEG DATASET **II_IA**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.3694	0.4991	0.5324	0.5708	0.556	0.5	0.5336	0.5858	0.5182	0.5978	**0.6194**
F-score	0.4028	0.433	0.5118	0.5604	0.5316	0.4233	0.5614	0.5691	0.4852	0.5728	**0.575**
κ	0.2018	0.1878	0.2	0.1198	0.1239	0.074	0.0663	0.1711	0.1028	0.2158	**0.2388**
S_{In}	0.5901	0.5801	0.5801	0.5798	0.5811	0.5828	0.5816	0.5919	0.5845	0.5732	**0.5522**
S_{Be}	0.6149	0.6144	0.622	0.6212	0.6176	0.6184	0.6192	0.6308	0.6258	**0.6385**	0.6188
r	1.042	1.0591	1.0722	1.0714	1.0628	1.0611	1.0646	1.0657	1.0707	1.1139	**1.1206**

TABLE IV
CLUSTERING RESULTS WITH THE EEG DATASET **II_IB**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.52	0.4987	0.4983	0.5025	0.5028	0.485	0.475	0.505	0.4975	0.5397	**0.56**
F-score	0.5514	0.4352	0.5514	0.5652	0.5688	0.5381	0.4724	0.5395	0.4718	0.5732	**0.5815**
κ	0.04	0.09	0.04	0.02	0.06	0.03	0.05	0.01	0.02	0.1	**0.12**
In	0.5665	0.562	0.5826	0.5771	0.5702	0.5743	0.5691	0.5701	0.5736	0.5628	**0.5539**
Be	0.5889	0.602	0.6266	0.6185	0.6264	0.6291	0.6262	0.6344	0.6314	0.6358	**0.6745**
r	1.0395	1.0712	1.0755	1.0717	1.0986	1.0954	1.1003	1.1128	1.1008	1.1297	**1.2177**

TABLE V
CLUSTERING RESULTS WITH THE EEG DATASET **II_IA_IB**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.5413	0.6089	0.6342	0.6417	0.6549	0.5731	0.5868	0.635	0.5563	0.6855	**0.7009**
F-score	0.5304	0.676	0.681	0.6862	0.6814	0.5731	0.5923	0.6855	0.5912	0.6987	**0.7276**
κ	0.201	0.311	0.3122	0.3155	0.31	0.2762	0.2658	0.3154	0.2837	0.3716	**0.3917**
In	0.5733	0.5712	0.5789	0.5688	0.5714	0.5775	0.5651	0.5667	0.5698	**0.5642**	0.5651
Be	0.8674	0.8732	0.8791	0.8716	0.8772	0.8779	0.8596	0.8762	0.8731	0.8782	**0.8866**
r	1.513	1.5287	1.5186	1.5323	1.5352	1.5202	1.5211	1.5461	1.5323	1.5565	**1.5689**

TABLE VI
CLUSTERING RESULTS WITH THE EEG DATASET **III_V_S1**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4011	0.5729	0.5747	0.5510	0.5884	0.5437	0.3839	0.5355	0.3441	0.5892	**0.6018**
F-score	0.3720	0.5411	0.5358	0.5325	0.5527	0.5314	0.3642	0.5290	0.3118	0.5519	**0.5634**
κ	0.3305	0.3179	0.3177	0.3067	0.3258	0.3112	0.2812	0.3017	0.3206	0.3327	**0.3392**
In	0.5583	0.5471	0.5413	0.5509	0.5388	0.5468	0.5527	0.5457	0.5621	0.5341	**0.5262**
Be	0.6192	0.6395	0.6360	0.6202	0.6416	0.6532	0.6469	0.6487	0.6337	0.6511	**0.6724**
r	1.1091	1.1689	1.1749	1.1258	1.1908	1.1946	1.1704	1.1887	1.1274	1.2191	**1.2778**

TABLE VII
CLUSTERING RESULTS WITH THE EEG DATASET **III_V_S2**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4366	0.5353	0.5219	0.4088	0.5692	0.4554	0.4107	0.4836	0.3479	0.5712	**0.5824**
F-score	0.7165	0.7213	0.6855	0.5828	0.7528	0.7120	0.6046	0.7199	0.3234	0.7325	**0.7602**
κ	0.1541	0.2755	0.2310	0.1569	0.3082	0.1967	0.1645	0.2320	0.3015	0.2988	**0.3127**
In	0.5628	0.5528	0.5512	0.5495	0.5431	0.5478	0.5507	0.5446	0.5672	0.5403	**0.5342**
Be	0.6473	0.6572	0.6534	0.6487	0.6592	0.6581	0.6519	0.6606	0.6415	0.6610	**0.6679**
r	1.1501	1.8889	1.1854	1.1805	1.2138	1.2014	1.1838	1.2130	1.1310	1.2234	**1.2503**

TABLE VIII
CLUSTERING RESULTS WITH THE EEG DATASET **III_V_S3**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4416	0.5481	0.5475	0.5515	0.5487	0.4376	0.4600	0.3894	0.3334	0.5524	**0.5640**
F-score	0.4169	0.4814	0.4882	0.4835	0.4836	0.4235	0.4355	0.3395	0.3007	0.4876	**0.4988**
κ	0.1364	0.2155	0.2112	0.2172	0.2219	0.1435	0.1683	0.0565	0.2187	0.2207	**0.2313**
In	0.5688	0.5571	0.5622	0.5584	0.5603	0.5523	0.5572	0.5618	0.5675	0.5507	**0.5504**
Be	0.6417	0.6538	0.6569	0.6582	0.6594	0.6613	0.6558	0.6513	0.6502	0.6597	**0.6647**
r	1.1282	1.1736	1.1684	1.1787	1.1769	1.1974	1.1770	1.1593	1.1457	1.1979	**1.2077**

TABLE IX
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2A_S1**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4815	0.6295	0.6401	0.6215	0.6357	0.5622	0.5518	0.5788	0.5174	0.6376	**0.6528**
F-score	0.2692	0.4355	0.4686	0.4305	0.4613	0.4101	0.4007	0.4128	0.3813	0.4672	**0.4751**
κ	0.1019	0.1902	0.1857	0.1925	0.1958	0.1787	0.1532	0.1794	0.1122	0.1984	**0.2027**
In	0.5835	0.5674	0.5722	0.5697	0.5615	0.5668	0.5682	0.5634	0.5721	0.5618	**0.5612**
Be	**0.6352**	0.6117	0.6233	0.6172	0.6087	0.6216	0.6203	0.6168	0.6221	0.6305	0.6338
r	1.0886	1.0781	1.0893	1.0834	1.0841	1.0967	1.0917	1.0948	1.0874	1.1223	**1.1294**

TABLE X
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2A_S2**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.3308	0.6259	0.6272	0.6220	0.6202	0.5465	0.5453	0.5439	0.4482	0.6284	**0.6354**
F-score	0.3241	0.4601	0.4562	0.4588	0.4522	0.4254	0.4237	0.4217	0.3665	0.4577	**0.4653**
κ	0.1272	0.2012	0.1989	0.2033	0.2006	0.1602	0.1573	0.1648	0.1382	0.2038	**0.2113**
In	0.5713	0.5679	0.5628	0.5589	0.5688	0.5618	0.5611	0.5674	0.5652	0.5582	**0.5543**
Be	0.5887	0.5936	0.6137	0.6184	0.6203	0.6116	0.5988	0.6012	0.5931	0.6191	**0.6284**
r	1.0305	1.0453	1.0904	1.1065	1.0905	1.0886	1.0672	1.0596	1.0494	1.1091	**1.1337**

TABLE XI
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2A_S3**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.3735	0.6261	0.6284	0.6164	0.628	0.5758	0.5558	0.5586	0.4488	0.6302	**0.6424**
F-score	0.3155	0.4482	0.4537	0.4456	0.4511	0.4350	0.4273	0.4264	0.3852	0.4471	**0.4635**
κ	0.1804	0.2735	0.2711	0.2812	0.2574	0.2372	0.2357	0.2448	0.1831	0.2822	**0.2866**
In	0.5688	0.5598	0.5654	0.5610	0.5584	0.5572	0.5621	0.5587	0.5711	0.5551	**0.5512**
Be	0.5976	0.6125	0.6088	0.6143	0.6132	0.6083	0.6112	**0.6157**	0.6028	0.6129	0.6145
r	1.0506	1.0941	1.0768	1.0950	1.0981	1.0917	1.0874	1.1020	1.0555	1.1041	**1.1148**

TABLE XII
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2B_S1**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4417	0.5071	0.4980	0.5048	0.5222	0.4915	0.5597	0.5667	0.4958	0.5652	**0.5833**
F-score	0.4416	0.4933	0.4752	0.4985	0.4979	0.4888	0.5042	0.5667	0.4812	0.5648	**0.5834**
κ	0.1	0.0163	0.0287	0.0318	0.1087	0.0167	0.0833	0.1333	0.0204	0.1122	**0.1406**
In	0.5589	0.5596	0.5608	0.5572	0.5563	0.5512	0.5469	0.5487	0.5534	0.5473	**0.5432**
Be	0.5722	0.5818	0.5782	0.5756	0.5735	0.5867	0.5812	0.5788	0.5716	0.5854	**0.5895**
r	1.0238	1.0397	1.0310	1.0330	1.0309	1.0644	1.0627	1.0549	1.0329	1.0696	**1.0852**

TABLE XIII
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2B_S2**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4578	0.5098	0.4993	0.5097	0.5044	0.5505	0.5359	0.5522	0.4958	0.5541	**0.575**
F-score	0.4565	0.4899	0.4863	0.4902	0.4896	0.5489	0.5249	0.5172	0.4757	0.5512	**0.5684**
κ	0.0333	0.0403	0.0433	0.0482	0.0487	0.1	0.0667	0.0833	0.0369	0.1107	**0.1660**
In	0.5576	0.5527	0.5618	0.5588	0.5603	0.5571	0.5523	0.5489	0.5598	0.5519	**0.5472**
Be	0.5723	0.5776	0.5745	0.5782	0.5769	0.5821	0.5830	0.5815	0.5784	**0.5833**	0.5825
r	1.0264	1.0451	1.0226	1.0347	1.0296	1.0449	1.0556	1.0594	1.0332	1.0569	**1.0645**

TABLE XIV
CLUSTERING RESULTS WITH THE EEG DATASET **IV_2B_S3**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4807	0.4959	0.4959	0.5127	0.5017	0.5256	0.5333	0.5335	0.4958	0.5326	**0.5583**
F-score	0.4654	0.4810	0.4832	0.5082	0.5029	0.5223	0.4034	0.5328	0.4816	0.5338	**0.5486**
κ	0.0333	0.0318	0.040	0.0482	0.0438	0.05	0.0667	0.0667	0.0395	0.0693	**0.0824**
In	0.5617	0.5513	0.5534	0.5582	0.5501	**0.5412**	0.5461	0.5479	0.5582	0.5451	0.5426
Be	0.5778	0.5792	0.5813	0.5964	0.5918	0.5910	0.5928	0.5986	0.5853	0.5978	**0.6022**
r	1.0287	1.0506	1.0504	1.0684	1.0758	1.0920	1.0855	1.0925	1.0485	1.0967	**1.1098**

TABLE XV
CLUSTERING RESULTS WITH THE EEG DATASET **IV_3_S1**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4511	0.6082	0.6195	0.6287	0.6064	0.5572	0.5436	0.5625	0.4453	0.6241	**0.6375**
F-score	0.3479	0.4177	0.4213	0.4256	0.4158	0.3934	0.3842	0.3990	0.3473	0.4241	**0.4347**
κ	0.1238	0.1612	0.1758	0.1735	0.1707	0.1556	0.1428	0.1603	0.1372	0.1765	**0.1806**
In	0.5886	0.5724	0.5698	0.5722	0.5711	0.5866	0.5837	0.5845	0.5823	0.5706	**0.5682**
Be	0.8628	0.8715	0.8687	**0.8738**	0.8692	0.8636	0.8576	0.8588	0.8607	0.8704	0.8735
r	1.4659	1.5225	1.5246	1.5271	1.5220	1.4722	1.4692	1.4693	1.4781	1.5254	**1.5373**

TABLE XVI
CLUSTERING RESULTS WITH THE EEG DATASET **IV_3_S2**

Measure	k-means++	UDFS	NDFS	RUFS	RSFS	K-SC	k-DBA	k-Shape	USLM	MTEEGC	mwcEEGc
RI	0.4510	0.6195	0.6006	0.6322	0.5929	0.5450	0.5451	0.5524	0.4653	0.6248	**0.6438**
F-score	0.3202	0.4154	0.4125	0.4236	0.4113	0.3801	0.3816	0.3869	0.3312	0.4167	**0.4293**
κ	0.1422	0.1725	0.1688	0.1758	0.1702	0.1683	0.1585	0.1623	0.1338	0.1752	**0.1786**
In	0.5792	0.5753	0.5676	0.5721	0.5694	0.5777	0.5812	0.5764	0.5789	0.5686	**0.5657**
Be	0.8710	0.8716	0.8698	0.8723	0.8745	0.8637	0.8654	0.8622	0.8576	0.8723	**0.8782**
r	1.5038	1.5150	1.5324	1.5247	1.5358	1.4951	1.4890	1.4958	1.4814	1.5341	**1.5524**

and S_{Be} simultaneously, it at least produces either the best S_{In} or S_{Be}. Consequently, when evaluating with integrated ratio r ($r = [S_{Be}/S_{In}]$) that simultaneously considers intracluster compactness and intercluster scatter of clusters, mwcEEGc outperforms all the ten state-of-the-art baselines.

3) Impact Analysis of Similarity Measures: As we introduced in Section V-B, we improved the conventional FD to measure similarities of EEG trials via bringing in local tendency. To demonstrate the improvement of the modified FD on mwcEEGc for EEG clustering, we further compared the IFD with the CFD, along with several other advanced and widely used similarity measures, such as ED, HD, and DTW. The clustering results of mwcEEGc with such similarity measures under optimal $\boldsymbol{\delta}$ (as illustrated in Fig. 5) are shown in Fig. 6. It clearly indicates that the IFD-based mwcEEGc outperforms ED-based, HD-based, DTW-based, and even CFD-based ones, since the IFD-based mwcEEGc achieves the highest *RI*, *F-score*, and *kappa* (κ) on 14 EEG datasets.

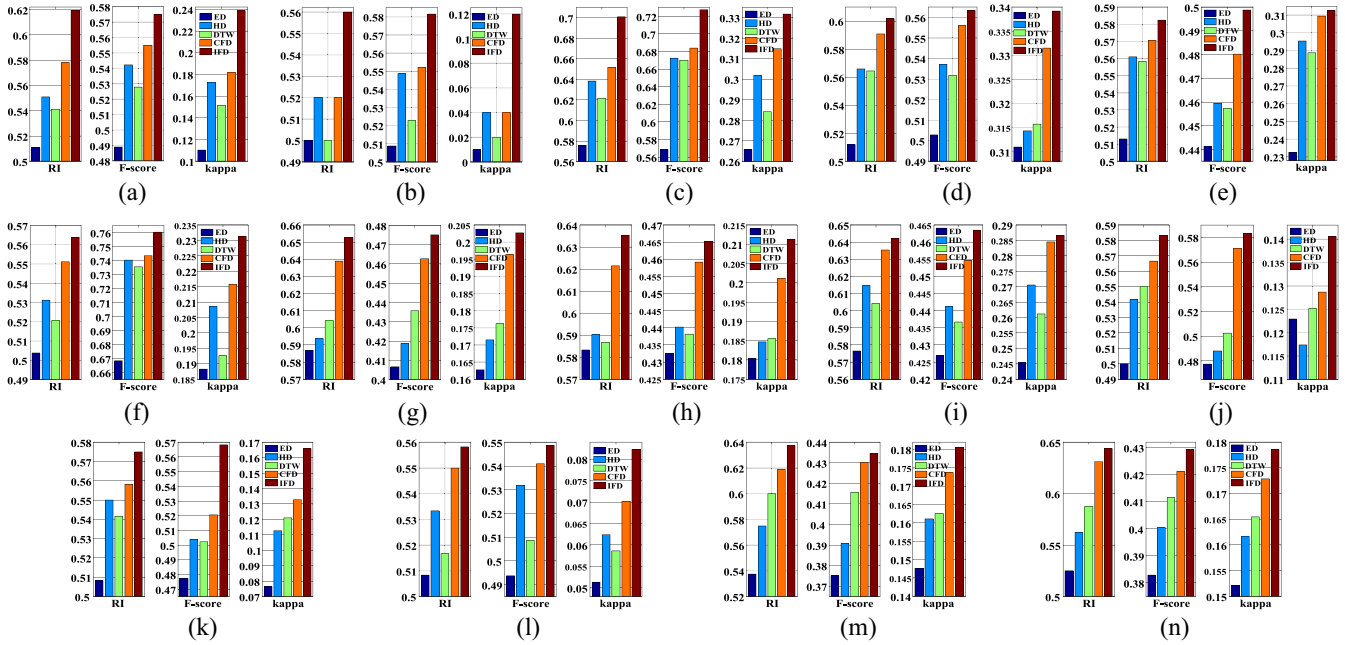

Fig. 6. Impact of distance (similarity) metrics on mwcEEGc with 14 EEG datasets. The results of all the distance metrics are achieved with the optimal similarity threshold δ as shown in Fig. 5. (ED: Euclidean distance; HD: Hausdorff distance; DTW: dynamic time warping; CFD: conventional FrÂt'echet distance without modification; and IFD: improved FrÂt'echet distance). (a) II_Ia. (b) II_Ib. (c) II_Ia_Ib. (d) III_V_s1. (e) III_V_s2. (f) III_V_s3. (g) IV_2a_s1. (h) IV_2a_s2. (i) IV_2a_s3. (j) IV_2b_s1. (k) IV_2b_s2. (l) IV_2b_s3. (m) IV_3_s1. (n) IV_3_s2.

4) Execution Time: As introduced in Section V-F, the time complexity of mwcEEGc is $O(\max\{(n^2 \cdot l^2[(\log\log l)/(\log l)]), mn^3\})$. But in practice, the real time is mostly determined by the improved FD-based similarity computation rather than by clique searching. The reasons are: 1) the min–max searching strategy of the Fréchet similarity requires lots of time, especially with a large amount of EEG trials and 2) for clique searching in weighted EEG graph, the similarity threshold $\boldsymbol{\delta}$ reduces the number of EEG trials whose similarities are smaller than $\boldsymbol{\delta}$ and it likely leads to a sparsely weighted EEG graph and lowering time for searching clusters in the sparse EEG graph. We also compared the time consumption of all clustering methods. For 14 EEG datasets, time consumption of clustering algorithms is illustrated in Fig. 7. As the results indicated, although mwcEEGc runs slower than *k*-means++, UDFS, RSFS, K-SC, *k*-Shape, and MTEEGC, it is more efficient than *k*-DBA, USLM on most EEG datasets as well as having competitive efficiency to NDFS and RUFS. Furthermore, mwcEEGc is not the most efficient algorithm for EEG clustering, but it achieves the best clustering results of *RI*, *F-score*, and κ on 14 EEG datasets.

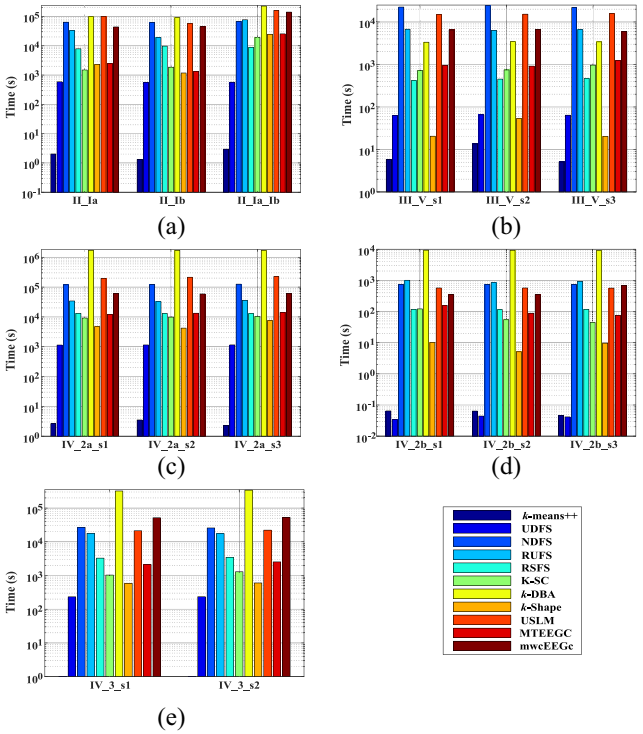

Fig. 7. Time consumption comparisons of EEG time-series clustering algorithms on 14 EEG datasets.

VII. Properties Satisfied by mwcEEGc

In Kleinberg's work [35], it discussed three clustering properties, that is, *scale invariance*, *richness*, and *consistency*. Kleinberg also proved that no clustering method satisfies the three properties simultaneously. A similar consequence has been proved on unification of clustering in [61]. *Order independence*, similar to Ackerman's isomorphism invariance [36], is another desirable clustering property. A clustering algorithm is order independent if it returns same clusters with different runs, without being influenced by the input sequences of objects [62]. Here, we discuss the four theoretical clustering properties for mwcEEGc: 1) *scale invariance*; 2) *richness*; 3) *consistency*; and 4) *order independence*.

A. Scale Invariance

Scale invariance requires that clustering algorithms do not build in length scale. mwcEEGc partitions EEG trials based

on similarity threshold $\boldsymbol{\delta}$, which builds in fundamental similarity scales. Therefore, mwcEEGc does not satisfy scale invariance. k-means++, UDFS, NDFS, RUFS, and RSFS satisfy scale invariance since they all utilize k-means strategy to clusters, relying on their relative distance to centers, instead of the absolute distance. K-SC also uses this property since it utilizes a scale-invariant distance measure $\hat{d}(x, y) = \min_{\alpha,q}[(\|x - \alpha y_{(q)}\|)/\|x\|]$ to cluster. k-DBA utilizes DBA distance measure to find the centroid sequence while minimizing the sum of squared distance. The DBA is deterministic once the initial average sequence (i.e., center) is selected, which means that it is not sensitive to changes in units of distance measurements. Namely, no "length scale" is built in, so k-DBA satisfies scale invariance. k-Shape is a shape-based clustering method with normalized cross-correlation based on the input number of clusters k, rather than a built-in distance scale, so k-Shape satisfies scale invariance. The USLM uses a joint optimization function of spectral regularization, shapelet similarity regularization, and regularized least-squares minimization to learn unsupervised shapelet and cluster, which does not build in distance scales, thus USLM also satisfies scale invariance. Similarly, MTEEGC is a shape centroid-based method that clusters EEG trials also based on the initialized number of clusters, rather than the built-in distance scale, so it also satisfies scale invariance.

B. Richness

Richness imposes that all possible partitions (clusters) can be produced. mwcEEGc clusters EEG trials according to $\boldsymbol{\delta}$, so that the thresholds can be adjusted to make the similarities of desired EEG trials satisfy the similarity thresholds, which means that it can produce any desired partitions. Therefore, mwcEEGc guarantees richness. k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, k-DBA, k-Shape, USLM, and MTEEGC fail to guarantee richness since their k is a constant initialization, which achieves only k clusters and precludes them generating any desired partitions.

C. Consistency

Consistency demands no change in clusters when shrinking intracluster distances and expanding intercluster distances. The mwcEEGc aims to search a family of cliques $\mathcal{C} = \{C_1, \ldots, C_m\}$ such that $\max_{C_k \in \mathcal{C}}$, which requires that any pairwise vertices (i, j) in the same clique are connected by an edge while any two vertices from different cliques are disjointed. Once $\boldsymbol{\delta}$ is fixed, cliques such that $\max_{C_k \in \mathcal{C}}$ are correspondingly fixed, otherwise the connections within cliques and separation between cliques would be broken, which would result in the total weight not being maximized. Therefore, mwcEEGc satisfies consistency. For k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, k-DBA, k-Shape, and MTEEGC, their clustering process is similar to k-means that utilizes the (k, g)-centroid function to cluster, that is, the objective function of these algorithms can be presented as $g(d) = d^2$ [35], [61]. As Kleinberg concluded in [35] that "for a fairly general class of centroid-based clustering functions, including k-means and k-median, none of the functions in the class satisfies the consistency property," therefore k-means++, UDFS, NDFS, RUFS, RSFS, K-SC, k-DBA, k-Shape, and MTEEGC do not satisfy consistency. Similarly, the USLM clusters EEG using a joint optimization objective function, and it probably returns local optima. When shrinking intracluster distances and expanding intercluster distances, the probability of EEG belonging to the cluster may change and the solution of the optimization function correspondingly changes. That is, the clusters change along with the change of distances. Therefore, USLM does not satisfy consistency.

D. Order Independence

Order independence requires no changes in clusters with the presentation order of objects. The mwcEEGc chooses first vertex v_0 with largest value based on $\boldsymbol{C} = \boldsymbol{\eta}^T \boldsymbol{\mu}$ into the potential clique C. Furthermore, adding a new vertex v_t with $\boldsymbol{C}_{-v_i} = \boldsymbol{\eta}^T_{-v_i} \boldsymbol{\mu}_{-v_i}$ such that $\forall v_i \in C, \mu_{ti} \geq \delta_C$ into the clique C to construct a new clique C_{+1}, according to Theorem 2, the weight of C_{+1} satisfies $\sum_{e \in E(C_{+1})} w_e = \sum_{e \in E(C)} w_e + \eta_t + \sum_{i \in C} \mu_{ti}$, and after checking all the vertices and their similarities, the final clique C^* is obtained with weight $\sum_{e \in E(C^*)} w_e = \sum_{i \in C^*} \eta_i + \sum_{e \in E(C^*)} \mu_e$. It indicates that the order of vertices that are added into the clique just affects the temporary weight of the cluster, rather than the final one. Therefore, the mwcEEGc satisfies order independence. For k-means++, UDFS, NDFS, RUFS, and RSFS, their cluster centers are randomly initialized, which results in different clusters for different runs. So they are not order independent. The K-SC randomly initializes cluster center μ, although it is updated in a way that the new center μ_k^* is the minimizer of the sum of $\hat{d}(x_i, \mu_k)^2$ over $\forall x_i \in C_k$. That is, $\mu_k^* = \arg\min_\mu \sum_{x_i \in C_k} \hat{d}(x_i, \mu)^2$. Due to its random initialization of center, K-SC does not satisfy order independence. The k-DBA utilizes DBA to minimize DTW squared distance to averaged sequence (center) which avoids applying iterative pairwise averaging to cluster. Although it is insensitive to ordering effects, the k-DBA is still not independent from the initialization of the first average sequence, so it returns different clusters along with different initializations. Consequently, the k-DBA does not satisfy order independence. The k-Shape computes cluster centroid by transforming to an optimization problem that aims to obtain the minimizer of the sum of the squared distances between EEG trials to the centroid. But in iterative clustering, the previously computed centroid is used as the reference sequence (centroid). According to this, the k-Shape likely returns different clusters during different runs. Therefore, the k-Shape is not order independent. The USLM requires many initializations, such as the length of shapelets, learning rate, etc., to the cluster. However, the leading influence is the initialization of the first shapelet. As a result, with different initializations for the fist shapelet, the USLM obtains different clusters. So it does not satisfy order independence. Although the MTEEGC updates cluster centroid during clustering, it also relies on the centroid sequence initialization, so it produces different clusters with multiple runs, which precludes it satisfying order independence.

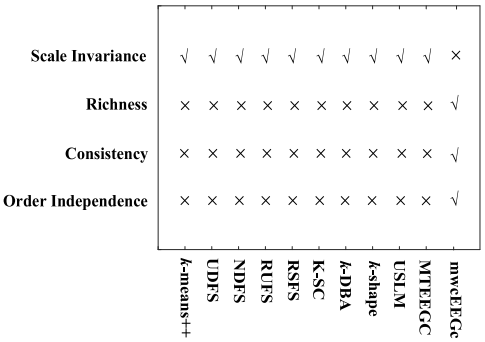

Fig. 8. Comparison of clustering properties that clustering algorithms satisfy. ($\surd$ and $\times$, respectively, denote the "satisfied" versus "not satisfied" clustering property.)

Fig. 8 summarizes the foregoing discussion and clearly indicates that mwcEEGc outperforms the ten state-of-the-art clustering algorithms since it satisfies three of four properties while the ten baselines only satisfy one. In accord with the impossibility theorem [35], [61] that all four properties cannot be satisfied simultaneously, we, therefore, underline theoretically that no other EEG clustering algorithms perform better than mwcEEGc for now.

VIII. Conclusion

With ever-increasing unlabeled EEG, unsupervised EEG clustering is a challenging but valuable task. This article explored the problem and proposed an MWC inspired approach, that is, mwcEEGc. The mwcEEGc clusters EEG trials by mapping it to searching cliques of maximum weights in an improved Fréchet similarity-weighted EEG graph with respect to intracluster compactness as well as intercluster scatter. The method simultaneously considers vertex weights and edge weights, and it is not required to compute cluster centroid during the process. The experimental comparisons with ten state-of-the-art clustering approaches on 14 EEG datasets with respect to six cluster evaluation criteria demonstrated the superiority of mwcEEGc. Besides, the mwcEEGc theoretically satisfies three of four key clustering properties: 1) richness; 2) consistency; and 3) order independence while those ten baselines only satisfy one: scale invariance.

In this article, we mainly focused on within-subject EEG clustering, but the cross-subject analysis is also important in EEG applications, so we also plan to handle it in our future work. The proposed method mwcEEGc exploits an FD-based metric to measure EEG similarities, and it would be an interesting direction to evaluate more similarity or dissimilarity measures for mwcEEGc. Moreover, mwcEEGc in this article treats all EEG channels equally, and it is worthy of weighting more for event-related channels or EEG channel selection for mwcEEGc clustering in future work. In addition, as a specific of time series with such characteristics as complexity, nonlinearity, nonstationarity, high dimension, and low signal-to-noise ratio, EEG signals are probably handled by mwcEEGc, so it could be also applied on other types of biological signals, such as electrocardiogram (ECG), electromyogram (EMG), and may be fMRI signals as well. It is another orientation to apply mwcEEGc for these signal clustering in the future.

References

[1] M. F. Glasser *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[2] E. Barzegaran, B. van Damme, R. A. Meuli, and M. G. Knyazeva, "Perception-related EEG is more sensitive to Alzheimer's disease effects than resting EEG," *Neurobiol. Aging*, vol. 43, pp. 129–139, Jul. 2016.

[3] J. S. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clin. Neurophysiol.*, vol. 115, no. 7, pp. 1490–1505, 2004.

[4] J. G. Bogaarts, E. D. Gommer, D. M. Hilkman, V. H. van Kranen-Mastenbroek, and J. P. H. Reulen, "EEG feature pre-processing for neonatal epileptic seizure detection," *Ann. Biomed. Eng.*, vol. 42, no. 11, pp. 2360–2368, 2014.

[5] K. Samiee, P. Kovacs, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015.

[6] B. Hordacre, N. C. Rogasch, and M. R. Goldsworthy, "Commentary: Utility of EEG measures of brain function in patients with acute stroke," *Front. Hum. Neurosci.*, vol. 10, Dec. 2016, Art. no. 621.

[7] M. R. Nuwer, S. E. Jordan, and S. S. Ahn, "Evaluation of stroke using EEG frequency analysis and topographic mapping," *Neurology*, vol. 37, no. 7, pp. 1153–1159, 1987.

[8] W. He, Y. Zhao, H. Tang, C. Sun, and W. Fu, "A wireless BCI and BMI system for wearable robots," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 936–946, Jul. 2016.

[9] H.-H. Kim and J. Jeong, "Representations of directions in EEG-BMI using winner-take-all readouts," in *Proc. 5th Int. Win. Conf. Brain Comput. Interface (BCI)*, 2017, pp. 121–122.

[10] H. Cecotti and A. Graeser, "Convolutional neural networks for P300 detection with application to brain–computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[11] A. K. Das, S. Suresh, and N. Sundararajan, "A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification," *Expert Syst. Appl.*, vol. 64, pp. 375–384, Dec. 2016.

[12] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.

[13] R. Roy, M. Mahadevappa, and C. S. Kumar, "Trajectory path planning of EEG controlled robotic arm using GA," in *Proc. 7th Int. Conf. Intell. Human Comput. Interact. (IHCI)*, 2016, pp. 147–151.

[14] H. A. Shedeed and M. F. Issa, "Brain-EEG signal classification based on data normalization for controlling a robotic arm," *Int. J. Tomograghy Simul.*, vol. 29, no. 1, pp. 72–85, 2016.

[15] F. Velasco-Álvarez, A. Fernandez-Rodríguez, and R. Ron-Angevin, "Switch mode to control a wheelchair through EEG signals," in *Proc. 3rd Int. Conf. NeuroRehabil. (ICNR)*, 2017, pp. 801–805.

[16] S. K. Swee, K. D. T. Kiang, and L. A. You, "EEG controlled wheelchair," in *Proc. Int. Conf. Mech. Manuf. Model. Mechatron. (IC4M)*, vol. 51, 2016, Art. no. 02011.

[17] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, and F. C. Morabito, "Permutation Jaccard distance-based hierarchical clustering to estimate EEG network density modifications in MCI subjects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 5122–5135, Oct. 2018.

[18] A. Ozdemir, M. Bolaños, E. Bernat, and S. Aviyente, "Hierarchical spectral consensus clustering for group analysis of functional brain networks," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2158–2169, Sep. 2015.

[19] C. Dai, J. Wu, D. Pi, and L. Cui, "Brain EEG time series selection: A novel graph-based approach for classification," in *Proc. SDM*, 2018, pp. 558–566.

[20] P. A. Bizopoulos, D. G. Tsalikakis, A. T. Tzallas, D. D. Koutsouris, and D. I. Fotiadis, "EEG epileptic seizure detection using k-means clustering and marginal spectrum based on ensemble empirical mode decomposition," in *Proc. 13th IEEE Int. Conf. BioInform. BioEng.*, 2013, pp. 1–4.

[21] P. Wahlberg and G. Lantz, "Methods for robust clustering of epileptic EEG spikes," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 857–868, Jul. 2000.

[22] C. Dai, D. Pi, L. Cui, and Y. Zhu, "MTEEGC: A novel approach for multi-trial EEG clustering," *Appl. Soft Comput.*, vol. 71, pp. 255–267, Oct. 2018.

[23] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "$l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning," in *Proc. IJCAI*, 2011, pp. 1589–1594.

[24] Z. Li, Y. Yang, J. Liu, X. Zhou, and H. Lu, "Unsupervised feature selection using nonnegative spectral analysis," in *Proc. 26th AAAI Conf. Artif. Intell. (AAAI)*, 2012, pp. 1026–1032.

[25] M. Qian and C. Zhai, "Robust unsupervised feature selection," in *Proc. 23rd Int. Joint Conf. Artif. Intell. (IJCAI)*, 2013, pp. 1621–1627.

[26] L. Shi, L. Du, and Y.-D. Shen, "Robust spectral learning for unsupervised feature selection," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, 2014, pp. 977–982.

[27] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. SODA*, 2007, pp. 1027–1035.

[28] T. Rakthanmanon *et al.*, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2012, pp. 262–270.

[29] F. Petitjean, A. Ketterlin, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011.

[30] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proc. WSDM*, 2011, pp. 177–186.

[31] W. Meesrikamolkul, V. Niennattrakul, and C. A. Ratanamahatana, "Shape-based clustering for time series data," in *Proc. Pac.–Asia Conf. Knowl. Disc. Data Min. (PAKDD)*, 2012, pp. 530–541.

[32] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proc. SIGMOD*, 2015, pp. 1855–1870.

[33] L. Ulanova, N. Begum, and E. J. Keogh, "Scalable clustering of time series with U-shapelets," in *Proc. SDM*, 2015, pp. 900–908.

[34] Q. Zhang, J. Wu, H. Yang, Y. Tian, and C. Zhang, "Unsupervised feature learning from time series," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 2322–2328.

[35] J. Kleinberg, "An impossibility theorem for clustering," in *Proc. 15th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2002, pp. 463–470.

[36] M. Ackerman and S. Ben-David, "Measures of clustering quality: A working set of axioms for clustering," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2008, pp. 121–128.

[37] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *Proc. SIGMOD*, 1994, pp. 419–429.

[38] M. R. Berthold and F. Höppner, "On clustering time series using Euclidean distance and Pearson correlation," 2016. [Online]. Available: arXiv:1601.02213v1.

[39] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005.

[40] A. A. Taha and A. Hanbury, "An efficient algorithm for calculating the exact Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2153–2163, Nov. 2015.

[41] P. K. Agarwal, R. B. Avraham, H. Kaplan, and M. Sharir, "Computing the discrete Fréchet distance in subquadratic time," *SIAM J. Comput.*, vol. 43, no. 2, pp. 429–449, 2014.

[42] U. Mori, A. Mendiburu, and J. A. Lozano, "Similarity measure selection for clustering time series databases," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 1, pp. 181–195, Jan. 2016.

[43] P. Chen, K. Xu, G. Li, J. Wan, and Y. Chen, "Local Fréchet distance in specific emitter identification," in *Proc. IEEE Int. Conf. Commun. Softw. Netw.*, 2017, pp. 842–845.

[44] A. Driemel, A. Krivosija, and C. Sohler, "Clustering time series under the Fréchet distance," in *Proc. SODA*, 2016, pp. 766–785.

[45] A. Driemel and F. Silvestri, "Locality-sensitive hashing of curves," 2017. [Online]. Available: arXiv:1703.04040v1.

[46] E. Billet, A. Fedorov, and N. Chrisochoides, "The use of robust local Hausdorff distances in accuracy assessment for image alignment of brain MRI," *Insight J.*, Jun. 2008. [Online]. Available: http://hdl.handle.net/1926/1354

[47] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[48] Y. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognit.*, vol. 44, no. 9, pp. 2231–2240, 2011.

[49] M. Schlesinger, E. Vodolazskiy, and V. Yakovenko, "Fréchet similarity of closed polygonal curves," *Int. J. Comput. Geom. Appl.*, vol. 26, no. 1, pp. 53–66, 2016.

[50] A. Driemel, S. Har-Peled, and C. Wenk, "Approximating the Fréchet distance for realistic curves in near linear time," *Discr. Comput. Geom.*, vol. 48, no. 1, pp. 97–127, 2012.

[51] K. T. Malladi, S. Mitrovic-Minic, and A. P. Punnen, "Clustered maximum-weight clique problem: Algorithms and empirical analysis," *Comput. Oper. Res.*, vol. 85, pp. 113–128, Sep. 2017.

[52] Q. H. Wu and J. Hao, "A review on algorithms for maximum-clique problems," *Eur. J. Oper. Res.*, vol. 242, no. 3, pp. 693–709, 2015.

[53] Y. Wang, J.-K. Hao, F. Glover, Z. Lü, and Q. Wu, "Solving the maximum vertex weight clique problem via binary quadratic programming," *J. Comb. Optim.*, vol. 32, no. 2, pp. 531–549, 2016.

[54] S. Cai and J. Lin, "Fast solving maximum-weight clique problem in massive graphs," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 568–574.

[55] N. S. Sengör, Y. Cakir, C. Guzelis, F. Pekergin, and Ö. Morgul, "An analysis of maximum-clique formulations and saturated linear dynamical network," *ARI Int. J. Phys. Eng. Sci.*, vol. 51, no. 4, pp. 268–276, 1998.

[56] A. Chouakria-Douzal and P. N. Nagabhushan, "Improved Fréchet distance for time series," *Data Sci. Classification*, pp. 13–20, 2006.

[57] H. Alt and M. Godau, "Computing the Fréchet distance between two polygonal curves," *Int. J. Comput. Geom. Appl.*, vol. 5, pp. 75–91, 1995.

[58] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[59] C. J. van Rijsbergen, *Information Retriveal*, 2nd ed. London, U.K.: Butterworths, 1979.

[60] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971.

[61] R. B. Zadeh and S. Ben-David, "A uniqueness theorem for clustering," in *Proc. Conf. Uncertainty Artif. Intell.*, 2009, pp. 639–646.

[62] V. K. Garg, Y. Narahari, and M. N. Murty, "Novel biojective clustering (BiGC) based on cooperative game theory," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1070–1082, May 2013.

Chenglong Dai received the master's degree in computer science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014, where he is currently pursuing the Ph.D. degree in computer science and technology.

He has published several related high-quality papers in journals and top conferences like SDM (awarded the Best Paper Award in Data Science Track). His research interests include EEG processing, EEG analyzing, and data mining.

Mr. Dai has served as a Reviewer for IJCNN'18 and IJCNN'19.

Jia Wu (Member, IEEE) received the Ph.D. degree in computer science from the University of Technology Sydney, Ultimo, NSW, Australia.

He is currently a Lecturer with the Department of Computing, Macquarie University, Sydney, NSW, Australia, and a Chair Professor with the School of Computer Science, Wuhan University, Wuhan, China. Since 2009, he has been published in over 100 refereed journal and conference papers, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the *ACM Transactions on Knowledge Discovery from Data*, IJCAI, AAAI, ICDM, and SIAM SDM. His current research interests include data mining and machine learning.

Dr. Wu was a recipient of the SDM'18 Best Paper Award in Data Science Track, the IJCNN'17 Best Student Paper Award, and the ICDM'14 Best Paper Candidate Award. He is an Associate Editor of the *ACM Transactions on Knowledge Discovery From Data*, the *Journal of Network and Computer Applications*, and *Neural Networks*.

Dechang Pi received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2002.

He is currently a Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. He has published seven research books and more than 100 papers in journals and conferences, won over 20 grants, including the National Science Foundation of China, the Technology Foundation of National Defence, and the Aviation Science Foundation of China. His main research interests lie in the area of big data, including data mining and EEG data preprocessing.

Stefanie I. Becker received the Ph.D. degree in cognitive psychology/experimental psychology from the University of Bielefeld, Bielefeld, Germany, in 2007.

She is currently an Associate Professor and an ARC Future Fellow with the University of Queensland, St. Lucia, QLD, Australia. She is an internationally recognized expert on visual attention. She has authored or coauthored over 60 papers in high-ranking journals.

Ms. Becker received several awards for her work involving eye tracking, EEG, and fMRI. She was as an Associate Editor for the *Journal of Experimental Psychology: Human Perception and Performance*.

Lin Cui received the master's degree from the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China, in 2008.

She is currently a Professor with the Intelligent Information Processing Laboratory, Suzhou University, Anhui, China. Her research interests include data mining, POI recommendation, and social network analysis.

Qin Zhang received the master's degree from the University of Chinese Academy of Sciences, Beijing, China, in 2014, and the Ph.D. degree from the Center for Artificial Intelligence, University of Technology Sydney, Ultimo, NSW, Australia, in 2017.

She has published several papers in top journals and conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IJCAI, and ICDM. Her main research interests include data mining and online learning.

Dr. Zhang has served as a Reviewer (Subreviewer) for KDD-15, ICDM-15, IJCAI-15, AAAI-15, and NIPS-15.

Blake Johnson received the M.A. and Ph.D. degrees from Fraser University, Burnaby, BC, Canada, in 1988 and 1993, respectively.

He is currently an Associate Professor with the Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia, where he was a Chief Investigator with the Australian Research Council Centre of Excellence for Cognition and its Disorders and a Project Leader with the Australian Government's Hearing Cooperative Research Centre (Hearing CRC). Current work on the development of speech motor control in children is funded by the Australian Research Council. His neuroimaging research investigates typical and a typical human–brain function using a variety of brain measurement techniques, including EEG and MEG (magnetoencephalography).