Sparse tensor dimensionality reduction with application to clustering of functional connectivity

Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves

Sparse tensor dimensionality reduction with application to clustering of functional connectivity

Gaëtan Frusque *Student Member, IEEE*, Julien Jung,
Pierre Borgnat *Member, IEEE*, Paulo Gonçalves *Member, IEEE*.

***Abstract*—Functional connectivity (FC) is a graph-like data structure commonly used by neuroscientists to study the dynamic behaviour of the brain activity. However, these analyses rapidly become complex and time-consuming, as the number of connectivity components to be studied is quadratic with the number of electrodes. In this work, we address the problem of clustering FC into relevant ensembles of simultaneously activated components that reveal characteristic patterns of the epileptic seizures of a given patient. While $k-$means is certainly the most popular method for data clustering, it is known to perform badly on large dimensional data sets, and to be highly sensitive to noise. To overcome the co-called curse of dimensionality, we propose a new tensor decomposition to reduce the size of the data set formed by FC time series recorded for several seizures, before applying k-means. The contribution of this paper is twofold: First, we derive a method that we compare to the state of the art, emphasizing one variant that imposes sparsity constraints. Second, we conduct a real case study, applying the proposed sparse tensor decomposition to epileptic data in order to infer the functional connectivity graph dynamics corresponding to the different stages of an epileptic seizure.**

***Index Terms*—dynamic networks, graph decomposition, clustering, dimensionality reduction, sparsity, tensor decompositions, HOSVD, HOOI, functional connectivity, iEEG.**

I. Introduction

Epilepsy is one of the most common neurological disorders in the world population. About 40% of the patients are drug-resistant and a surgical operation can be considered to extract the epileptogenic area. To locate this area and understand the evolution of epilepsy, practitioners often use intracranial electroencephalography (iEEG) recordings [1], [2]. The patient makes a stay in the hospital for several days with electrodes implemented in the brain to record multiple epileptic seizures. The stages of a seizure are distinguished by similar evolutions of the recorded iEEG signals in different areas of the brain. Functional Connectivities (FC) that quantify along time these similarities are calculated between all pairs of signals, usually by means of the spectral coherence or the Phase Locking Value [3]. Considering electrodes as nodes and FCs as weights on the edges, neuroscientists try to find in the data, graph-like structures evolving through time (see Fig. 1). The study of the FCs is a complex and expensive task, with a typical experiment consisting of about 100 electrodes, hence ≈ 5000 FC times series. Also, since it is generally assumed that the FC dynamics are comparable from one seizure to another for a same patient, another asset of iEEG monitoring is the possibility to record several seizures. The joint analysis of these repetitions should ease the identification of a dynamical FC pattern, common to all seizures, and characteristic of the patient's health disorder. All this calls for a method able to extract relevant ensembles of simultaneously activated components, viewed as the edges of sub-graphs of a FC network inferred from the iEEG signals.

More concretely, for each seizure, the recorded iEEG signals yield L pairwise Functional Connectivity measures (FC) as time series of T samples. The corresponding data matrix $\mathbf{X} \in \mathbb{R}^{L\times T}$ is referred to as an Epoch. As we observe S different seizures for a same patient, the resulting epochs are stacked in a 3-modes tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L\times T\times S}$. Our objective is then, to identify characteristic clusters of dynamical FCs, using unsupervised k-means applied to the underlying dynamic graph.

The current generalisation of k-means to find relevant clusters in a tensor $\boldsymbol{\mathcal{X}}$, first imposes to unfold it as a matrix. Here, since it is the L time series that are to be grouped, this leads to cluster FC components that lie in a high dimensional $T \times S$ space, a situation where k-means is known to perform poorly [4]. In addition, such generalisation does not explicitly take advantage of the natural tensor structure of the data. Then, a first step will be to reduce the tensor $\boldsymbol{\mathcal{X}}$ into a factor matrix $\mathbf{F} \in \mathbb{R}^{L\times K}$ with $K \ll T \times S$, and to apply k-means on $\mathbf{F}$. We propose to perform this dimension reduction directly on $\boldsymbol{\mathcal{X}}$ using tensor decompositions and imposing relevant structural constraints to the solution, such as sparsity.

Treating dynamic graphs as tensors is not new and previous works have already explored this idea, e.g. for sampling [5], dynamic graph inference [6] or community detection [7]. But the method we are developing here can be seen as a decomposition of dynamical graphs, where the

Manuscript submitted May 28, 2019. This work was supported by the ANR-14-CE27-0001 GRAPHSIP grant, and by the ACADEMICS grant of the IDEXLYON, project of the Université de Lyon, PIA operated by ANR-16-IDEX-0005.

G. Frusque and P. Gonçalves are with Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, F-69342, Lyon, France (e-mail: gaetan.frusque@ens-lyon.fr and paulo.goncalves@inria.fr).

J. Jung is with HCL, Neuro. Hosp., Functional Neurology and Epileptology Dept & Lyon Neurosc. Res. Cent., INSERM, CNRS, Lyon, France (e-mail: julien.jung@chu-lyon.fr).

P. Borgnat is with Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France (e-mail: Pierre.Borgnat@ens-lyon.fr).

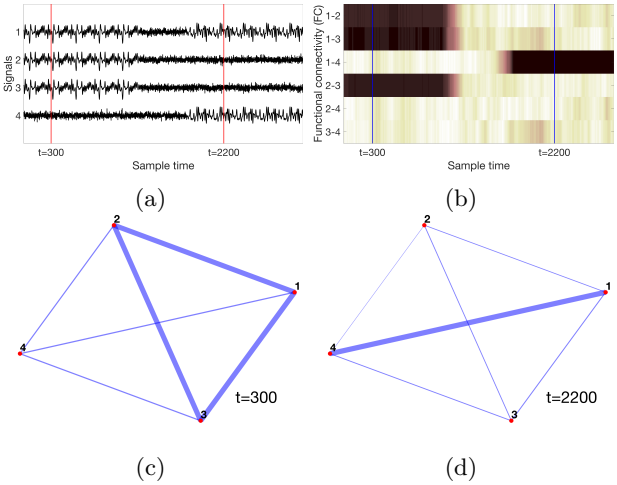

Fig. 1: (a) 4 simulated iEEG signals, (b) FC computed between the 6 pairs of signals by PLV, (c) & (d) Graphs with electrodes 1,2,3 and 4 as nodes and PLV measures as the weight of edges, at t=300 and t=2200 (red/blue bars on Fig. (a) and (b)).

data is decomposed into a structural (e.g., sub-graphs of functional connectivity) and temporal (e.g., time varying activation) signature of individuals (e.g. epileptic seizure). This procedure could be used in other contexts involving dynamic graphs, such as social face-to-face interactions [8], [9] or transportation networks [10].

Besides the necessary survey on the state-of-the-art methods for tensor reductions, our contribution in this paper is manyfold. Firstly, we propose new algorithms for reducing data dimension, that impose pertinent structural constraints on the solutions. Then, based on these latters, we design an original data processing workflow, depicted in Fig. 2, to infer spatio-temporal patterns in non-stationary relational data. To assess the ability of the proposed methods at compressing data while enhancing structural features, we apply k-means to cluster the reduced factor matrix $\mathbf{X}$ obtained from our algorithm and compare it with other dimensionality reduction approaches (including tensor decompositions). Finally, we apply our spatio-temporal decomposition of dynamic graphs to real iEEG recordings, to characterise the prime lineament of functional connectivity during epileptic seizures of a patient.

The article is organized as follows. Section II recalls the general notations, background and state-of-the-art. Section III describes the methods (from k-means to tensor reduction methods) mobilized in the article. Then, Section IV describes the new proposed tensor decomposition. In Section V, the different possible approaches are compared on a simple, yet flexible model of iEEG FCs, highlighting the performance and the limits of our method. Section VI applies the method on real data, exhibiting its capacity to infer a dynamic graph characterizing the evolution of the seizure. We conclude the article in Section VII.

II. Notations and state-of-the-art

A. Notations

To keep the presentation as clear as possible we introduce the following notation coming from [11], [12]. We also refer to these works as good introductions to tensor decomposition theory, and to [13] for a more recent and deep introduction. A D-modes array (where D correspond to the number of dimensions used to write the data) is called a vector if $D = 1$, a matrix if $D = 2$ and a tensor if $D = 3$ or more. In this work, we use only 3-modes tensors, but the presented theory can be generalized to higher dimensions. Tensors are denoted with bold case calligraphic letters $\boldsymbol{\mathcal{X}}$, matrices and vector are denoted respectively in bold upper-case and lower-case $\mathbf{X}$, $\mathbf{x}$, and scalars by lower-case letters x. Notice that : l, t, s, k, n will be used as indices, and L, T, S, K, N will be reserved to denote their index upper bounds. Here, they correspond to the FC mode, time mode, epoch (or trial) mode, number of factors and number of clusters, respectively. Then, $\mathbf{x}_{:t}$, resp. $\mathbf{x}_{l:}$, corresponds to the column t, resp. to the row l, of the matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$. The matrices $\mathbf{X}_{l::}$, $\mathbf{X}_{:t:}$ or $\mathbf{X}_{::s}$ correspond to the slices of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$. The slice for each trial $\mathbf{X}_{::\mathbf{s}}$ is referred to as an epoch.

The matricization corresponds to the matrix representation of a tensor. It can be made for each mode of a tensor. For a 3-mode tensor they are noted $\mathbf{X}_{(L)} \in \mathbb{R}^{L \times TS}$ (mode-1 matricization), $\mathbf{X}_{(T)} \in \mathbb{R}^{T \times LS}$ (mode-2 matricization) and $\mathbf{X}_{(S)} \in \mathbb{R}^{S \times LT}$ (mode-3 matricization) [12]. Writing $[\mathbf{A}, \mathbf{B}]$ the concatenation of two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same number of rows, the mode-1 matricization reads:

$$\mathbf{X}_{(L)} = [\mathbf{X}_{::1}, \mathbf{X}_{::2}, ..., \mathbf{X}_{::S}]. \tag{1}$$

Similarly, the vectorisation of a matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$, denoted by $vec(\mathbf{X}) \in \mathbb{R}^{1 \times LT}$ converts a matrix to a vector [12]:

$$vec(\mathbf{X}) = [\mathbf{x}_{1:}, \mathbf{x}_{2:}, ..., \mathbf{x}_{L:}]. \tag{2}$$

Considering $\mathbf{x}, \mathbf{y} \in \mathbb{R}^L$, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{l=1}^{L} x_l y_l$ is the scalar product between two vectors and $|| \mathbf{x} ||_F = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ the Frobenius norm. These notations can be extended to D-mode arrays [12], [14], e.g. $\langle \mathbf{X}, \mathbf{Y} \rangle = \langle vec(\mathbf{X}), vec(\mathbf{Y}) \rangle$ or $|| \boldsymbol{\mathcal{X}} ||_F = || vec(\mathbf{X}_{(L)}) ||_F$. The L_1-norm for a matrix corresponds to:

$$|| \mathbf{X} ||_1 = \sum_{l=1}^{L} \sum_{t=1}^{T} | x_{lt} |. \tag{3}$$

Finally we define the Kronecker product between two matrices, noted $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ with $\mathbf{A} \in \mathbb{R}^{L \times T}$, $\mathbf{B} \in \mathbb{R}^{K \times N}$ and $\mathbf{C} \in \mathbb{R}^{LK \times TN}$, as:

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1T}\mathbf{B} \\ \vdots & & \vdots \\ a_{L1}\mathbf{B} & \dots & a_{LT}\mathbf{B} \end{pmatrix}. \tag{4}$$

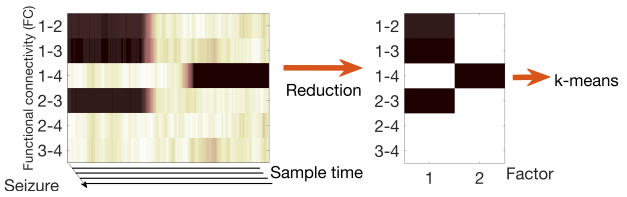

Fig. 2: Model of the data processing workflow: tensor $\boldsymbol{\mathcal{X}}$ is reduced to a factor matrix $\mathbf{F}$, and k-means is applied on $\mathbf{F}$

B. State-of-the-art

Several approaches to cluster functional connectivity have been proposed, using various techniques: community detection [15] to discriminate FC characterizing different states of an epileptic seizure; spectral clustering [16] to combine electrodes so as to cluster FC graphs; non-negative matrix factorisation [17] to discover subgraphs of FC with evolutionary activation over time; k-means for the clustering of FC or time states [17], [18], [19]. The clustering method used here is the k-means algorithm [4], as it is simple to use, yet performs well against other more recent clustering method [20] like dbscan [21] or spectral clustering [22]. Moreover, its limitations are well studied. In particular, the performance of k-means drastically decrease on high dimensional data [4], and each iteration of this algorithm has a computational cost of $o(LTK)$ if one want to partition the matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$ into K clusters. Clearly, this hampers its use with large dataset, notably because of the number of repetitions - with different initial conditions - needed to warrant a satisfactory local minima. Moreover, since FC measurements are often noisy due to spurious similarities between iEEG signals, many local minima exist, that corresponds to irrelevant clustering solutions. Finally, k-means can be seen as a constrained matrix factorisation [23], [24]. This last observation motivates the need for dimensionality reduction, before factorizing the matrix with k-means: the dataset is reduced into a matrix $\mathbf{F} \in \mathbb{R}^{L \times K}$ on which k-means is applied. Actually, this pre-processing has a double advantage: it filters the data by removing spurious content and it avoids to apply k-means on high dimensional data sets.

The most used dimensionality reduction for k-means is the singular value decomposition (SVD) [25]–[27]. The tandem of SVD and k-means clustering was applied to network applications [25], [28], in medical imaging [29] or FC analysis [17] and it was thoroughly studied in gene expression [26]. The denoising properties of the SVD are well understood [30] and recent theoretical results points to the similarity between the best clustering results that we can have from a dataset and its reduction via SVD [31], [32]. However, as it will be detailed in Section III-B, this tandem has its own limitations [33], [34] and existing extensions, like the sparse SVD, are promising alternative to overcome these. The use of sparse methods for FC is recent. The authors of [35] show an interesting link between k-means, SVD and sparse decompositions, and they also comment on their use for FC grouping. In [36], a sparse non-negative matrix factorisation is proposed, that leverages the non-negativity of the data, to produce sparse FC graphs and their associated temporal evolution.

In addition, we would like to take advantage of the tensor structure of the data. Tensor factorisations methods were developed, which avoid concatenating epochs of different seizures [36], or, having to use the averaged epochs [19] to perform matrix factorisation. The work in [37] reports an interesting survey on the use of tensor decomposition methods in the context of brain data, including FC. Different tensor decompositions applied on FC data can also be highlighted: the article [38] proposes to use a canonycal polyadic decomposition (or CP, PARAFAC) to exhibit the principal patterns of data. However, this decomposition does not necessarily identify the pattern common to all seizures, which is a severe limitation for our application. In [39], the high order SVD (HOSVD) is used to extract a representative epoch from a 4-modes dataset . In this setting, the functional connectivities of one epoch at a given time step are represented by a $N \times N$ matrix. At the end of the procedure, the data is chopped into several steps, each one being associated to a particular FC graph. The authors of [40] use a higher order robust principal component analysis applied to the same dataset as in [39] and following the same goal, but the graph they obtain is more regular. However, both methods suffer from a lack of sparsity in their resulting graphs, entailing results that are difficult to analyse as they do not single out any critical functional connectivity that would characterise the seizure onset.

III. Problem and background

Combining dimensionality reduction (with tensor or matrix factorisation) and clustering (k-means here) has given promising results to cluster dynamic FCs in structural and temporal components. Our problem is to develop a tensor factorisation which both extracts temporal patterns common to epoch of each seizure, and constrains $\mathbf{F}$ to be sparse. The sparsity constraint appears to be important because, in comparison to the case without this constraint, $\mathbf{F}$ becomes closer to a cluster assignment matrix. Moreover, it limits the complexity of clusters by reducing the number of FC they contain. This is important in the context of epileptic data where a large number of FC measurements can be passively implied in a neurological process (during the discharge of the seizure for example).

Then, our proposed approach seeks to perform the following decomposition:

$$\mathbf{X}_{(L)} \approx \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t \tag{5}$$

where $\mathbf{w} \in \mathbb{R}^S$ is a vector performing a weighted average of the S epochs in order to extract common patterns, $\mathbf{F} \in \mathbb{R}^{L \times K}$ is sparse and scaled such that $|| \mathbf{w} ||_F^2 = || \mathbf{v}_{:k} ||_F^2 = 1$ (to remove scaling indeterminacy and transfer all the energy in our factor matrix). Finally, $\mathbf{V} \in \mathbb{R}^{T \times K}$ will also be constrained to be sparse so as to select specific temporal steps of FC activation, eliminating

thus, periods where there is no common activation of FC clusters. This decomposition is close to a Block term decomposition [41] with one component and with an additional sparsity constraint imposed on the two first modes. It is also a variant of a parsimonious Tucker decomposition, presented in [42] or [43], with the difference that factorisation with a core tensor is not used in our model (it would be difficult to use if we want for each sparse FC patterns a sparse temporal activation profile).

After reduction, the k-means step enforces the graphs to be binary, sparse and simple to study. In order to understand the proposed tensor decomposition, we show how to rewrite k-means algorithm so as to highlight its link with the SVD and its extensions. Then, two useful extensions of SVD are presented, which will serve as the building blocks of our proposed tensor decomposition.

A. k-means clustering and dimensionality reduction

Applying k-means algorithm [4] on the tensor $\boldsymbol{\mathcal{X}}$ amounts to find the N matrices $\boldsymbol{\Theta}_{n::}$, centroids of the N clusters that best characterise the data. The l-th slice $\mathbf{X}_{l::}$ of the tensor, belongs to the cluster n (noted $l \in C_n$), if its nearest centroid is $\boldsymbol{\Theta}_{n::}$. As shown in Appendix A-A, this is equivalent to applying k-means clustering on the rows of the matrix $\mathbf{X}_{(L)}$, where the centroids of the clusters are now vectors $\theta_{n:} \in \mathbb{R}^{TS}$, solutions of the optimization problem :

$$\underset{\theta_{1:},\theta_{2:},\dots,\theta_{N:}}{\operatorname{argmin}} \sum_{n=1}^{N} \sum_{l \in C_n} || \mathbf{X}_{(L)l:} - \theta_{n:} ||_F^2 \tag{6}$$

Moreover, as we show in appendix A-B, the solution to (6) amounts to find the matrix $\mathbf{A} \in \mathbb{R}^{L \times N}$ that maximises the following norm (the optimal solution of the following problem is called $\mathbf{A}^*$):

$$\underset{\mathbf{A}}{\operatorname{argmax}} \quad || \mathbf{A}^t \mathbf{X}_{(L)} ||_F^2 \tag{7}$$

and such that the columns $\mathbf{a}_{:i}$ of $\mathbf{A}$ form a standard basis of a subspace of $\mathbb{R}^L$, i.e. $\mathbf{A}$ is unitary, sparse and non-negative. Now, in order to both highlight relevant and filtered FC factors (e.g. sparse and common to every seizure) and compensate for the curse of dimensionality, we consider performing dimensionality reduction on $\boldsymbol{\mathcal{X}}$ before applying k-means. The goal is to run the k-means procedure on a reduced data set, called factor matrix and noted $\mathbf{F} \in \mathbb{R}^{L \times K}$. For that we seek to replace, in Eq. (7), the raw data matrix $\mathbf{X}_{(L)}$ by this lower dimension factor matrix $\mathbf{F}$, leading to the following optimal solution:

$$\underset{\mathbf{A}}{\operatorname{argmax}} \quad || \mathbf{A}^t \mathbf{F} ||_F^2 \tag{8}$$

where $\mathbf{A}$ should have the same structure as in (7). It is important to remark that, in a general case, the optimal solutions of Eqs (7) and (8) are not the same. Indeed, the dimensionality reduction is not only performed to compensate the curse of dimensionality, but it also helps clustering as it denoises the data into more relevant components. In addition, if the reduction is able to impose to $\mathbf{F}$ some constraints that $\mathbf{A}$ must verify in Eq. (7), then the clustering solution can be improved [24].

B. The singular value decomposition

This last point can be done partially using the SVD [24]. By imposing only the orthogonality constraint in Eq. (7), the solution $\mathbf{A}^*$ corresponds to the projection of $\mathbf{X}_{(L)}$ on its N first right singular vectors. As a result, the truncated SVD of $\mathbf{X}_{(L)}$ can be seen as a relaxation of the k-means clustering problem [24]. Indeed, the SVD of $\mathbf{X}_{(L)}$ can be written as [27]:

$$\mathbf{X}_{(L)} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{Z}^t, \tag{9}$$

where $\mathbf{U} \in \mathbb{R}^{L \times L}$ and $\mathbf{Z} \in \mathbb{R}^{TS \times TS}$ are unitary matrices, and $\boldsymbol{\Lambda} \in \mathbb{R}^{L \times TS}$ is a positive diagonal matrix such that, if $L \leq TS$, $\lambda_{11} \geq \lambda_{ll} \geq \lambda_{LL}$. The best low order approximation according to the Frobenius norm of $\mathbf{X}_{(L)}$ is $\mathbf{U}^{(K)}\boldsymbol{\Lambda}^{(K)}\mathbf{Z}^{(K)t}$ with $\mathbf{U}^{(K)} = \mathbf{U}_{:[1,\dots,K]}$, $\mathbf{Z}^{(K)} = \mathbf{Z}_{:[1,\dots,K]}$ and $\boldsymbol{\Lambda}^{(K)} = \mathbf{U}^{(K)t}\mathbf{X}\mathbf{Z}^{(K)}$. This dimensionality reduction method is in fact the application of k-means clustering on $\mathbf{F}_{\text{SVD}} = \mathbf{U}^{(K)}\boldsymbol{\Lambda}^{(K)}$, as in Eq. (8). The columns of $\mathbf{F}_{\text{SVD}}$ are also referred to as the principal components of $\mathbf{X}_{(L)}$.

The SVD reduction is also used for data denoising – limiting the noise to the reduced signal subspace only [30]. Hence it should yield a clustering that is less sensitive to spurious FC. According to [31], another property is that, under some hypothesis, $\mathbf{F}_{SVD}$ is a good approximation of $\mathbf{X}_{(L)}$ to obtain the best clustering solution $\mathbf{A}^*$ of Eq. (7). However, this property is limited for our problem since it does not lead to clusters of components that are sparse and common to each seizure. Also, in high dimensions, relevant information is more likely to be masked by misleading dimensions (in our case, corresponding to periods of time where there is no relevant activation of FC). The euclidean distance between FCs becomes less discriminatory, exposing thus the SVD to the curse of dimensionality too. To overcome these weaknesses of the SVD, we propose in the following to use sparsity to favour more relevant dimensionality reductions.

C. The sparse singular value decomposition

Finding a sparse SVD (sSVD) (or sparse PCA) was declined in different manners, e.g. [44], [45]. The ideal constraints would be both at sparsity and orthogonality. Yet, we will use the relaxed version [44] to find a low-rank approximation of $\mathbf{X}_{(L)}$ under sparse constraints. Mathematically speaking, this sparse version of the SVD looks for the low-rank matrices $\mathbf{U} \in \mathbb{R}^{L \times K}$ and $\mathbf{Z} \in \mathbb{R}^{TS \times K}$, solutions of the following constrained optimization problem:

$$\underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} \; || \mathbf{X}_{(L)} - \mathbf{U}\mathbf{Z}^t ||_F^2 + \gamma_1 || \mathbf{U} ||_1 + \gamma_2 || \mathbf{Z} ||_1 \tag{10}$$

where the meta-parameters γ_1 and γ_2 allow for tuning the trade-off between accuracy and sparsity of the approximation. Finally, we choose to normalize the columns of $\mathbf{U}$ and of $\mathbf{Z}$ such that $|| \mathbf{z}_{:\mathbf{k}} ||_F^2 = 1 \; \forall k \in 1, \dots, K$ in order to avoid

scaling indeterminacies. To solve problem (10), we use the SPAMS library on matlab [46].

Combined with k-means, the low dimension factor matrix $\mathbf{F}$ to be plugged in Eq. (8) would simply read here $\mathbf{F}_{sSVD} = \mathbf{U}$. The optimization problem in (10) can be seen as a dictionary learning task, with parsimony imposed both to the dictionary $\mathbf{Z}$ and to the coefficients $\mathbf{U}$. Notice that we could impose sparsity uniquely on $\mathbf{U}$ or on $\mathbf{Z}$, but empirically both options give poorer results. Let us remark also that instead of sparsity, we could impose a non-negativity constraint to the optimization problem (10), leading thus to the so-called non-negative matrix factorisation (NMF) [47]. NMF is also viewed as a soft clustering method [48], but its use for dimensionality reduction prior to k-means clustering is less common. Still, we will compare the use of the corresponding low dimension factor matrix $\mathbf{F}_{NMF} = \mathbf{U}$ in Eq. (8) and check its performance against other alternatives.

D. Higher order SVD

In our problem, since data takes on the form of a tensor $\boldsymbol{\mathcal{X}}$, it is logical to use a higher-order version of the SVD rather than applying SVD to the concatenation of the seizures. Actually, there are several extensions of SVD to tensors [49]. The most popular one is certainly the canonical polyadic decomposition [50], which extends the SVD in the sense that it decomposes the data into a sum of rank one tensors, like SVD decomposes the data in a sum of rank one matrices. It amounts to find the matrices $\mathbf{U} \in \mathbb{R}^{L\times K}$, $\mathbf{V} \in \mathbb{R}^{T\times K}$ and $\mathbf{W} \in \mathbb{R}^{S\times K}$ that minimize the norm:

$$\underset{\mathbf{U},\mathbf{V},\mathbf{W}}{\operatorname{argmin}} \quad || \boldsymbol{\mathcal{X}} - \sum_{k=1}^{K} \mathbf{u}_{:k} \times \mathbf{v}_{:k} \times \mathbf{w}_{:k} ||_F^2, \qquad (11)$$

where $\times$ corresponds to the tensor product (let $\mathbf{u} \in \mathbb{R}^L$ and $\mathbf{v} \in \mathbb{R}^T$, the tensor product $\mathbf{u}\times\mathbf{v} = \mathbf{M} \in \mathbb{R}^{L\times T}$ is such that $m_{lt} = u_l v_t$ [11]). To warrant the existence of a global solution for this decomposition, it is recommended to add non-negativity constraints to all factors [51], leading to to the non-negative canonycal polyadic decompostion [52] (nnCP). After rescaling $\mathbf{U}$ such that $|| \mathbf{v}_{:\mathbf{k}} ||_F^2 = || \mathbf{w}_{:\mathbf{k}} ||_F^2 = 1$ $\forall k \in 1, ..., K$ in order to avoid scaling indeterminacies, we get $\mathbf{F}_{nnCP} = \mathbf{U}$ to be substituted in Eq. (8). By construction, this decomposition is not intended to identify common patterns in one particular mode (e.g. for all epochs); each factor $\mathbf{u}_{:k}$ has a proper activation $\mathbf{w}_{:k}$ alongside the epoch mode, and the odds to get seizure specific patterns increase with K. This is a limiting for our application.

Interestingly though, the canonical polyadic decomposition is a particular case of a more general SVD in high dimensions, for which the core tensor is hyperdiagonal, with the same size K for all modes. Relaxing these two constraints and imposing on each matrix to be unitary yields the high order singular value decomposition (HOSVD) [14], another possible extension of the SVD. It amounts to find three unitary matrices $\mathbf{U} \in \mathbb{R}^{L\times L}$, $\mathbf{V} \in \mathbb{R}^{T\times T}$ and $\mathbf{W} \in \mathbb{R}^{S\times S}$ such that the following decomposition holds:

$$\mathbf{X}_{(L)} = \mathbf{U}\mathbf{G}_{(L)}(\mathbf{W} \otimes \mathbf{V})^t. \qquad (12)$$

Like it is the case with SVD, HOSVD seeks for orthogonal matrices that best characterise the data with respect to each of its modes. In contrast to SVD though, the matrix $\mathbf{G}_{(L)} \in \mathbb{R}^{L\times TS}$ is no longer diagonal and is dense in general. To compute the decomposition (12), we use the HOSVD algorithm proposed in [14] and recalled in Algorithm 1.

Algorithm 1 Estimate $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ via HOSVD.

Require: $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L\times T\times S}$
$\mathbf{U} = left\_L\_SVD(\mathbf{X}_{(L)})$ ▷ [1]
$\mathbf{V} = left\_T\_SVD(\mathbf{X}_{(T)})$
$\mathbf{W} = left\_S\_SVD(\mathbf{X}_{(S)})$

[1]: $\mathbf{U}^{(K)} = left\_K\_SVD(\mathbf{X})$ denotes the algorithm computing the first K singular components of X.

Now HOSVD can be seen, as SVD, as a dimension reduction of the data before clustering. In other words, we are looking for lower rank matrices $\mathbf{U}^{(K_L)} \in \mathbb{R}^{L\times K_L}$, $\mathbf{V}^{(K_T)} \in \mathbb{R}^{T\times K_T}$ and $\mathbf{W}^{(K_S)} \in \mathbb{R}^{S\times K_S}$ that are now solutions of the optimal approximation:

$$\underset{\mathbf{U}^{(K_L)},\mathbf{V}^{(K_T)},\mathbf{W}^{(K_S)}}{\operatorname{argmin}} \quad || \mathbf{X}_{(L)} - \mathbf{U}^{(K_L)}\mathbf{G}_{(L)}(\mathbf{W}^{(K_S)} \otimes \mathbf{V}^{(K_T)})^t ||_F^2, \qquad (13)$$

with $\mathbf{G}_{(L)} \in \mathbb{R}^{K_L\times K_T K_S}$, a dense matrix. However, in contrast to the SVD, the sought low rank matrices do not simply stem from truncating the solution $\mathbf{U}$ (resp. $\mathbf{V}$, $\mathbf{W}$) to its first K_L (resp. K_T, K_S) columns. Instead, we need to resort to an iterative optimization algorithm as the Higher Order Orthogonal Iteration of Tensors (HOOI) proposed in [43], [53] and depicted in Algorithm 2. The work in [54] shows that in most cases, HOOI reaches the optimal solution of (13), with good convergence performance.

Algorithm 2 Estimate $\mathbf{U}^{(K_L)}$, $\mathbf{V}^{(K_T)}$ and $\mathbf{W}^{(K_S)}$ via HOOI.

Require: $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L\times T\times S}$, The parameter for reduction (K_L, K_T, K_S), increment tolerance $\epsilon > 0$, i_{max}.
$i = 0$
$[\mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_0] = HOSVD(\boldsymbol{\mathcal{X}})$
while $i < i_{max}$ or $|| \mathbf{G}_{(L)i} ||_F^2 - || \mathbf{G}_{(L)i-1} ||_F^2 > \epsilon$ **do**
$\mathbf{U}_{i+1}^{(K_L)} = left\_K_L\_SVD(\mathbf{X}_{(L)}(\mathbf{W}_i^{(K_S)} \otimes \mathbf{V}_i^{(K_T)}))$
$\mathbf{V}_{i+1}^{(K_T)} = left\_K_T\_SVD(\mathbf{X}_{(T)}(\mathbf{U}_{i+1}^{(K_L)} \otimes \mathbf{W}_i^{(K_S)}))$
$\mathbf{W}_{i+1}^{(K_S)} = left\_K_S\_SVD(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K_T)} \otimes \mathbf{U}_{i+1}^{(K_L)}))$
$\mathbf{G}_{(L)i+1} = \mathbf{U}_{i+1}^{(K_L)t}\mathbf{X}_{(L)}(\mathbf{W}_{i+1}^{(K_S)} \otimes \mathbf{V}_{i+1}^{(K_T)})$
i = i+1
end while

IV. New tensor reductions for clustering

Our aim here is to highlight the component of FC dynamics that is common to all seizures. So, setting $K_S = 1$ in (13), leads to a vector $\mathbf{W}^{(1)} \in \mathbb{R}^{S\times 1}$ that measures the contribution rate of each seizure to the common pattern. For instance, a constant vector ($\mathbf{W}_s^{(1)} = \frac{1}{\sqrt{S}}$ $\forall s \in 1, ..., S$)

means that the common pattern is simply the average of all seizures. Whereas, if the seizures share no common features, the retained component will match the seizure s for which $|| \mathbf{X}_{::s} ||_F^2$ is maximum. In general though, the common component is a linear combination of all seizures' patterns and it is always good practice to have a close look at $\mathbf{W}^{(1)}$ to see how heterogeneous the seizures are.

For the sake of simplicity and without loss of generality, we will consider in the following that $K_L = K_T = K$, and propose a modification of the HOOI procedure for this specific case. This new procedure has the advantage to converge more rapidly to the optimal solution, moreover some simple modification can be done in order to add constraints on the reduction.

A. modified HOOI (mHOOI)

In Algorithm 2, matrices $\mathbf{U}_i^{(K)}$ and $\mathbf{V}_i^{(K)}$ are computed by extracting the left singular components of two different matrices. By noting that the common epoch (associated to $\mathbf{W}^{(1)}$) is a matrix, the idea behind the modification consists in extracting the left and the right singular components of this one matrix in order to obtain $\mathbf{U}_i^{(K)}$ and $\mathbf{V}_i^{(K)}$. Then, the three steps of the original algorithm are reduced to two. Notice that if $K_S > 1$ (we take $K_S = 2$ for the example), the tensor $\boldsymbol{\mathcal{X}}$ will be compressed in two Epoch forming a tensor, and the respective mode-L and mode-T matricization of this tensor will be two matrices with different dimensions, implying the impossibility to extract $\mathbf{U}_i^{(K)}$ and $\mathbf{V}_i^{(K)}$ from the same matrix factorisation.

The proposed modified HOOI (noted mHOOI) is as follows, iterating 2 steps in alternance:

(A) At iteration $(i+1)$ we assume $\mathbf{W}_i^{(1)}$ to be known. From this vector we compute the common Epoch corresponding to $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}) \in \mathbb{R}^{L \times T}$, with $\mathbf{I} \in \mathbb{R}^{T \times T}$ the identity matrix (it corresponds to the mode-L matricization of the contraction product between the tensor $\boldsymbol{\mathcal{X}}$ and the vector $\mathbf{W}^{(1)t}$ [12]). We then compute the low rank approximations $\mathbf{U}_{i+1}^{(K)}$ and $\mathbf{V}_{i+1}^{(K)}$ as the K first components of the singular value decomposition of $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}) = \mathbf{U}_{i+1}\boldsymbol{\Lambda}_{i+1}\mathbf{V}_{i+1}^t$.

(B) In order to update $\mathbf{W}_{i+1}^{(1)}$ we first filter the tensor $\boldsymbol{\mathcal{X}}$ by projecting it on the subspaces spanned by the matrices $\mathbf{U}_{i+1}^{(K)}$ and $\mathbf{V}_{i+1}^{(K)}$. $\mathbf{W}_{i+1}^{(1)}$ is then obtained by a special HOOI decomposition on the filtered tensor with $K_L = L$, $K_T = T$ and $K_S = 1$. As shown in Appendix B-A, because only one mode is reduced (the Epoch mode here), the optimal decomposition can be found analytically. Both the filtering and the decomposition, in order to get $\mathbf{W}_{i+1}^{(1)}$, reduce to the problem of computing the dominant left singular vector of $\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K)} \otimes \mathbf{U}_{i+1}^{(K)})$.

We summarize in Algorithm 3, this modified HOOI method (mHOOI) and how to obtain the corresponding lower dimension factor matrix $\mathbf{F}_{mHOOI}$. Worth noticing too, we use an angular metric between two successive estimates of $\mathbf{W}^{(1)}$ as the stopping criterion of the iterative procedure, sparing thus the unnecessary computations of $\mathbf{G}_{(L)i}$. As the two steps **(A)** and **(B)** yield unique and optimal solutions, and provided the calculation of SVD does not degenerate [54], we expect the procedure to converge. In fact, it empirically converges more rapidly than HOOI to an optimal solution, as shown in Appendix B-B.

Algorithm 3 Estimation of $\mathbf{F}_{mHOOI}$

Require: $\boldsymbol{\mathcal{X}}$, K, angular tolerance $\epsilon > 0$, et i_{max}.
  $i = 0$
  $\mathbf{W}_0^{(1)} = left\_1\_SVD(\mathbf{X}_{(3)})$
  while $i < i_{max}$ or $acos(< \mathbf{W}_i^{(1)}, \mathbf{W}_{i-1}^{(1)} >) > \epsilon$ **do**
    (A). $[\mathbf{U}_{i+1}^{(K)}, \boldsymbol{\Lambda}_{i+1}, \mathbf{V}_{i+1}^{(K)}] = K\_SVD(\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}))$
    (B). $\mathbf{W}_{i+1}^{(1)} = left\_1\_SVD(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K)} \otimes \mathbf{U}_{i+1}^{(K)}))$
    $i = i + 1$
  end while
  $\mathbf{F}_{mHOOI} = \mathbf{U}_i^{(K)}\boldsymbol{\Lambda}_i$

B. High Order sparse SVD (HOsSVD)

The previous algorithm requires several SVD and the final matrix F_{mHOOI} is not sparse in general. As we already noticed, sparsity constraints is wanted to get FC graphs with few edges, and because the SVD is impacted by the curse of dimensionality it is desirable to limit its use. Following the same rationale, and in order to perform a decomposition as proposed in Eq. (5), we develop a sparse version of mHOOI, where we replace the computation of the SVD in step **(A)**, by the sparse SVD of Eq. (10). We call this decomposition a High Order sparse SVD (HOsSVD). Since the matrix $\mathbf{U}_{i+1}$ is not necessarily orthogonal anymore, we compute the subspace spanned by this matrix by performing its QR decomposition, and we retain only the first K column vectors of the $\mathbf{Q}$ part. The same goes for $\mathbf{V}_{i+1}$. Writing $qr(\mathbf{V}_{i+1}, K)$ this operation, Algorithm 4 presents the computation of $\mathbf{F}_{HOsSVD}$, the lower dimension factor matrix stemming from this high order sparse SVD reduction (HOsSVD).

As a final remark, let us stress that in this work, we focused on the tensor extension of the sparse SVD. But similarly, other matrix factorisations could be used instead (NMF [47], sparse NMF [36] , or a k-SVD dimensionality reduction [35]). However, the approach proposed here empirically performs the best on synthetic and real data.

Algorithm 4 Estimation of $\mathbf{F}_{HOsSVD}$

Require: $\boldsymbol{\mathcal{X}}$, K, γ_1, γ_2, angular tolerance $\epsilon > 0$, et i_{max}.
  $i = 0$
  $\mathbf{W}_0^{(1)} = left\_1\_SVD(\mathbf{X}_{(3)})$
  while $i < i_{max}$ or $acos(< \mathbf{W}_i^{(1)}, \mathbf{W}_{i-1}^{(1)} >) > \epsilon$ **do**
    (A). $\mathbf{U}_{i+1}$, $\mathbf{V}_{i+1}$ minimising (10) with $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I})$.
    $\mathbf{A} = qr(\mathbf{U}_{i+1}, K)$, $\mathbf{B} = qr(\mathbf{V}_{i+1}, K)$
    (B). $\mathbf{W}_{i+1}^{(1)} = left\_1\_SVD(\mathbf{X}_{(S)}(\mathbf{B} \otimes \mathbf{A}))$
    $i = i + 1$
  end while
  $\mathbf{F}_{HOsSVD} = \mathbf{U}_i^{(K)}$

V. Comparison on a model

In this Section, we evaluate the performance of the proposed methods and compare them to the state of the art methods for dimensionality reduction of data, prior to their k-means clustering. To this end, and because we are primarily interested in iEEG signals, we first present an original and oversimplified model of functional connectivity that, still, integrates four sources of uncertainty able to reproduce most experimental variabilities.

A. A dynamic graph model for FC

Epileptic seizure implies a pathological FC that starts in a focal brain onset, then spreads to the other connected regions, and sometimes split to give rise to new FC components. Fig. 3 displays a characteristic example of actual FC time series measured by the phase lock value (PLV) [3]. The model we propose is aimed at reproducing the global structured pattern of FCs' activation, and the uncertainties of the measures. More precisely, we consider a matrix $\mathbf{X} \in \mathbb{R}^{L\times T}$, where x_{lt} is set to a high value if the FC of index $l \in \{1,\dots,L\}$ is active at time $t \in \{1,\dots,T\}$, and to a low value otherwise (to account for the non ON-OFF discrepancy of the PLV measurement, we choose values equal to 0.7 and 0.2, respectively). A cluster C_n, $n \in \{1,\dots,N\}$, is composed of all FCs that are activated over the same period of time T_n. This is for the deterministic part of the model, defining the structural and temporal pattern of FCs activation, common to all seizures. Superimposed to it, we add a seizure dependant random component, composed of four uncertainty sources:

i Random duration: each activation period T_n is uniformly distributed between a minimum duration (here, 5 time steps) and $\frac{T}{N}$ time steps (the binary variable α allows to able (1) or disable (0) this random mode).
ii Activation error: with probability $\beta \in [0,1]$, each FC of a given group C_n incurs the risk to be replaced by any other randomly chosen FC.
iii Connectivity noise: we add to x_{lt} a white centered Gaussian noise of power σ^2 (SNR $= \sigma^{-2}$).
iv Jitter: All FCs of the same group start activating with independent jitters, uniformly distributed in $\left[-\frac{\delta}{2},\frac{\delta}{2}\right]$.

Figure 3 displays one realization of these synthetic FC time series. Compared to real data, the model succeeds in reproducing a realistic global pattern. More importantly, as it allows a control of the nature and the intensity of variability between epochs, it will serve to evaluate the sensitivity and the robustness of the different dimension reduction methods for clustering, with respect to each source of uncertainty.

To this end, we simulate different seizures of a same patient as i.i.d. realizations of our model with the same set of parameters $b = [\alpha,\beta,\sigma,\delta]$. As for real data, the S modeled seizures are then stacked in a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L\times T\times S}$. Figure 4 displays the unfolded matrix $\mathbf{X}_{(L)}$ corresponding to the particular choice $[\alpha = 1, \beta = 0.2, \sigma = 0.05, \delta = 0]$, $L = 66$, $T = 1000$ and $S = 4$.

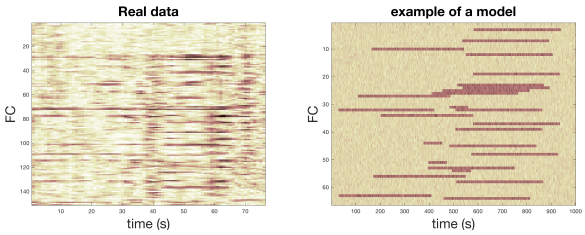

Fig. 3: Real data vs example of model we can get for one seizure with noise parameters $b = [1, 0.2, 0.1, 0.1]$.

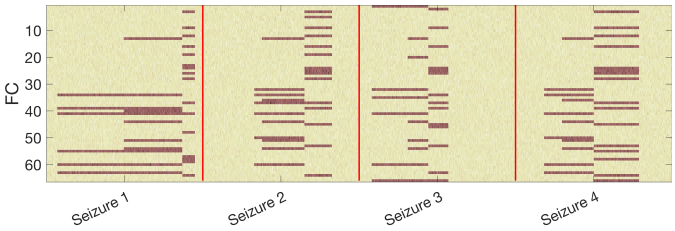

Fig. 4: realization of $\mathbf{X}_{(L)}$ corresponding to 4 simulated seizures (or epochs) with noise parameters $b = [1, 0.2, 0.05, 0]$.

B. Experimental setup

Using the FC model introduced above, we can now study and compare the different methods for reducing data dimensionality, derived in Section IV, to the state-of-the-art methods recalled before. Each method yields a lower dimension factor matrix $\mathbf{F}$ that serves as an input to k-means. Performance refer here to the ability at retrieving the FC clusters of the global pattern. Table I summarizes the 8 methods we considered, indicating for each, which implementation was used and the best (empirically determined) hyper-parameters values. As benchmarks, we also consider the two straightforward approaches that consist in applying k-means directly on the unfolded tensor $\mathbf{F}_{Direct} = X_{(L)}$, or on the seizures average $\mathbf{F}_{mean} = \sum_{s=1}^{S} X_{::s}$.

Method	Section	Implementation	Parameters
1 - Direct	V-B	(1)	-
2 - Mean	V-B	(1)	-
3 - NMF	III-C	(1)	$\epsilon = 10^{-4}$
4 - nnCP	III-D	(2)	$\epsilon = 10^{-6}$
5 - SVD	III-B	(1)	-
6 - mHOOI	IV-A	(1)-(3)	$\epsilon = 10^{-3}$
7 - sSVD	III-C	(1)-(4)	$\lambda_1 = 0.1$, $\lambda_2 = 1$, $\epsilon = 10^{-3}$
8 - HOsSVD	IV-B	(1)-(3)-(4)	$\lambda_1 = 0.1$, $\lambda_2 = 4$, $\epsilon = 10^{-3}$

TABLE I: Comparative method to convert the tensor $\boldsymbol{\mathcal{X}}$ to a factor matrix $\mathbf{F}$

The links for the different toolbox for implementation and our code. (in Matlab) :

- (1) Algorithms used in this article: FCTensDec
- (2) N-way toolbox version 3.30 [55]
- (3) MATLAB Tensor Toolbox Version 2.6 [56]
- (4) SPAMS toolbox version 2.6 [46]

The reduced dimension K is varied from 1 to 4, and we retain, in our comparisons, the value yielding the best grouping score for each method. Regarding k-means

algorithm, we used k-means++ version [57] that we stopped after 1000 iterations and repeated 120 times with different seeds. The number of sought groups is set to $N = 4$. Finally, to assess the clustering performance of each method, we use the Adjusted Rand Index (ARI) score [58], computed between the resulting grouping and the ground truth: the score equals 1 for a perfect match and 0 if the correspondence does not outperform a random grouping.

C. Results and discussion

Under these experimental conditions, we evaluate all methods for various configurations of uncertainty (different configurations are resumed in table II). Fig. 5 displays the ARI scores (mean and variance estimated out of 120 independent realizations of $\boldsymbol{\mathcal{X}}$) for 6 different combinations of (i) random duration, (ii) activation error and (iii) connectivity noise. Here, the jitter uncertainty is disabled. Fig. 6, displays the same but with (iv) the jitter activated. We do not show the results with only random duration as all methods perform equally in this case.

From Fig. 5, we observe that the methods with dimensional reduction have better performances than the standard methods. Tensor methods perform better than their matrix counterparts, and sparse methods generally outperform conventional methods. On the opposite, the non-negativity constraint does not seem to be helpful here. Globally, HOsSVD distinguishes itself systematically, and especially for weak SNRs. Focusing on the case with only connectivity noise (experiment b_1), we can see that HOsSVD clearly outperforms other methods.

We isolated in Fig. 6 the impact of jitter since it is the sole uncertainty source for which HOsSVD does not systematically reach the best performance. From that perspective, the worst case corresponds to the combination of jitter and random duration and jitter alone (case b_2) where our method performs just better than the direct methods, but worse than matrix dimensionality reduction. For the case b_{10} the results are similar to that of the other dimensionality reduction methods and for the case b_7 both of our proposed approach (mHOOI, HOsSVD) get similar results, slightly better than those of the other methods. Experimentally it corresponds to situations where we could get better results if we would perform tensor reduction with $K_S > 1$. Indeed the complexity generated by the jitter impact the common activation of clusters and can mix different groups of FC. A solution to this problem with a similar method as ours, including $K_S > 1$ scenario, could be to extract only the left factor matrix on both temporal and FC mode since it is impossible to do it simultaneously. This is not pursued further here. Finally, HOsSVD dominates again in the presence of connectivity noise.

Conclusion Globally HOsSVD performs generally better (or equal) than all other reduction methods except for two case, it outperforms other method for models with connectivity noise. Methods mHOOI and sSVD get good performances overall. The non-negativity constraint does not seem advantageous in any scenarios.

$b_i = [\alpha, \beta, \sigma, \delta]$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
α			✓		✓	✓			✓			✓	✓	✓
β		✓	✓	✓		✓		✓			✓	✓		✓
σ	✓			✓	✓	✓				✓	✓		✓	✓
δ							✓	✓	✓	✓	✓	✓	✓	✓

TABLE II: Checkmark when the uncertainty in configuration b_i is present

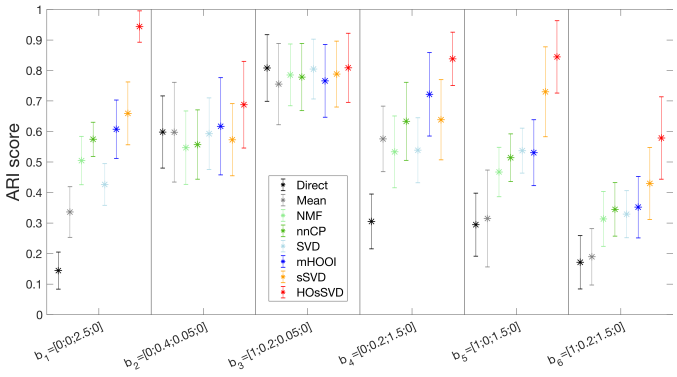

Fig. 5: ARI score for each method, for 6 models associated to vector b_i without jitter uncertainty

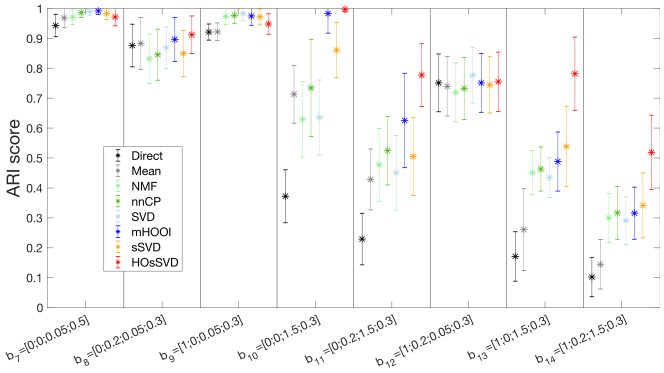

Fig. 6: ARI score for each method, for 7 models associated to vector b_i with jitter uncertainty

VI. Application on real data

Data: We consider real iEEG data from a patient with focal epilepsy [1], [2]. The electrodes used are distributed on stems implanted in the brain. The activity of the brain is recorded via 5 to 10 electrodes per stem. The space between two consecutive electrodes is 3.5 mm. Over a recording time of 15 days, 4 seizures were selected. Each seizure is delimited in time by a window of 100 seconds centered on the beginning of the seizures. The signal is sampled at 256 Hz. 33 equitably distributed contacts over the initial 108 are selected to avoid too strong spatial correlations. The functional connectivity metric used is PLV [3]. A strong PLV between two signals means that their phases are similar. The 528 FC (corresponding to each pair of electrodes) were calculated over a sliding rectangular window of 4 seconds duration, with a time step of one second. After eliminating the points that suffer from border effects, the data is formatted as a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{528\times96\times4}$.
Application: From the tensor $\boldsymbol{\mathcal{X}}$, we obtain the matrix

$\mathbf{F}_{HOsSVD}$ (using Algorithm 4), where the parameters, $K = 4$, $\gamma_1 = 1$, $\gamma_2 = 1$ and $\epsilon = 10^{-3}$ are empirically fixed (It is current to consider between 3 and 5 different steps in an epileptic seizure justifying $K = 4$; for γ_1 and γ_2 we search for the highest value of parameters where results seem coherent). Fig. 7 (top) shows the temporal activation profiles of the components of $\mathbf{F}_{HOsSVD}$ (corresponding to the matrix $\mathbf{V}^{(K)}$ from Algorithm 4). There are 4 activation periods that can be easily associated with 4 steps of the seizures: before seizure, start, propagation and end of the seizure. The time interval around time 50s is particularly interesting as it shows no activated FC. This is likely to correspond to the functional decoupling at the early start of the seizure, a short period when iEEG activities in different areas of the brain are suddenly decorrelated [59]. We apply k-means to $\mathbf{F}_{HOsSVD}$ to identify the $N = 5$ corresponding FC groups. Since $\mathbf{F}_{HOsSVD}$ is close to an assignment matrix (a solution $\mathbf{A}^*$ of Eq. (8)), we retain the 4 groups of smaller sizes, which can be associated with the 4 activation periods while the the 5th group corresponds to the unsynchronized FCs. Figure 7 (bottom) materializes the positions of the 33 electrodes projected on the transverse plane (according to the lair of Tailarach). An FC is represented by a link between the pair of electrodes that are in phase. The four FC groups can be associated to four snapshots of a time-varying graph: before the seizure, only two electrodes interact in what could correspond to the epileptogenic zone. At the beginning of the seizure: spreading of FC activation with the appearance of a cluster of FC localized around the epileptogenic zone. During the crisis, other FC appears spontaneously in the other hemisphere, at the same time the FC are diffused in the left hemisphere. At the end of the crisis, the two hemispheres interact with the appearance of common FC.

The graphs obtained are in agreement with the clinical results, the focus of the beginning of the seizure being close to the graphs "before-seizure" and "seizure start". The propagation of seizures in the right hemisphere of the brain is well represented by the graphs "Propagation" and "Seizure end".

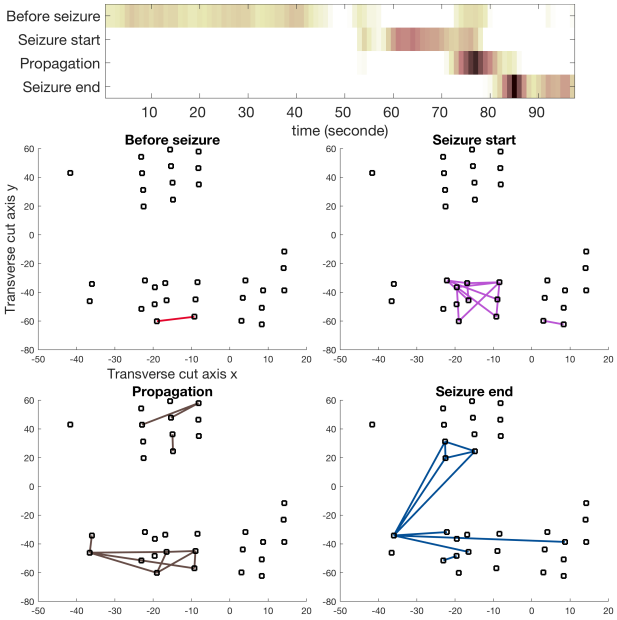

Fig. 7: (above) activation profile of HOsSVD (variable $\mathbf{V}$ of algorithm 4 (under) Cluster of FC corresponding to the 4 activation steps of the seizures of the considered patient)

VII. Conclusion

In this work, we presented different dimensionality reduction for k-means clustering with the objective of clustering functional connectivity data over time and different epochs. As we expect each trial (or epoch) to have the same sorting of FC activation, but with different temporal duration, we developed a tensor decomposition, based on the existing HOOI algorithm, to select groups of FC activation that are representative of all trials. We compared the methods on a model of epileptic functional connectivity, with 4 kinds of uncertainty. Globally, methods using tensor decomposition and sparse constraints get the best performance, in particular the proposed HOsSVD reduction which consists in a sparse tensor decomposition. This HOsSVD method is an interesting alternative for the reduction of tensor data. Applied as pre-processing in raw data, it significantly improves the clustering performance of k-means. Applied to real iEEG data recorded during an epileptic seizure, our method allowed us to identify the 4 FC activation groups corresponding to the 4 significant time periods of the evolution of the seizure. As a perspective of this work, a more automated choice of the hyper-parameters would lead for a more systematic and exhaustive analysis of comparative performance. This would allow us to further analyse the dynamics of functional connectivity in a richer clinical dataset.

Appendix A
More about k-means clustering

A. K-means on tensor

The goal here is to perform k-means to cluster the first mode of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$. This lead to find N

matrix centroids $\mathbf{\Theta}_{n::} \in \mathbb{R}^{T\times S}$ to cluster the L slices $\mathbf{X}_{l::}$.

$$\underset{\mathbf{\Theta}_{1::},\mathbf{\Theta}_{2::},...,\mathbf{\Theta}_{N::}}{\operatorname{argmin}} \sum_{n=1}^{N}\sum_{l\in C_n} || \mathbf{X}_{l::} - \mathbf{\Theta}_{n::} ||_F^2 \quad (14)$$

moreover, the L_2-norm is invariant the linear function $vec()$ ($|| \mathbf{X} ||_F^2$=$|| vec(\mathbf{X}) ||_F^2$), then :

$$\underset{\mathbf{\Theta}_{1::},\mathbf{\Theta}_{2::},...,\mathbf{\Theta}_{N::}}{\operatorname{argmin}} \sum_{n=1}^{N}\sum_{l\in C_n} || vec(\mathbf{X}_{l::}) - vec(\mathbf{\Theta}_{n::}) ||_F^2 \quad (15)$$

By considering mode-1 matricization of $\boldsymbol{\mathcal{X}}$ with $\mathbf{X}_{(1)l:} = vec(\mathbf{X}_{l::})$ and noting $\theta_{n:} = vec(\mathbf{\Theta}_{n::})$ we get Eq. (6).

B. K-means is matrix factorisation

We start from Eq. (6) and set $\mathbf{X} = \mathbf{X}_{(1)}$ to simplify the notations. By calling $\mathbf{S} \in \mathbb{R}^{L\times N}$ the matrix of binary indicator variables such that :

$$s_{ln} = \begin{cases} 1 & \text{if } l \in C_n, \\ 0 & \text{else} \end{cases} \quad (16)$$

we can rewrite the cost function, Eq. (6) [23], as :

$$\sum_{n=1}^{N}\sum_{l\in C_n} || \mathbf{x}_{l:} - \theta_{n:} ||_F^2 = \sum_{n=1}^{N}\sum_{l}^{L} s_{ln} || \mathbf{x}_{l:} - \theta_{n:} ||_F^2, \quad (17)$$

$$= || \mathbf{X} - \mathbf{S\Theta} ||_F^2, \quad (18)$$

$$= Tr(\mathbf{XX}^t) - 2Tr(\mathbf{S\Theta X^t}) + Tr(\mathbf{S\Theta\Theta^t S^t}), \quad (19)$$

$$= Tr(\mathbf{XX}^t) - Tr(\mathbf{S\Theta X^t}), \quad (20)$$

The passage from (19) to (20) is proved in the following equations. When $\mathbf{S}$ is fixed, the centroids are found by least square regression, using the Moore-penrose Pseudoinverse of $\mathbf{S}$, noted $\mathbf{S}^\dagger \in \mathbb{R}^{N\times L}$, (we recall that $\mathbf{S}^\dagger\mathbf{S} = \mathbf{I}$):

$$\mathbf{\Theta} = (\mathbf{S^t S})^{-1}\mathbf{S}^t\mathbf{X} = \mathbf{S}^\dagger\mathbf{X}, \quad (21)$$

Note this result have a physical meaning, since $\mathbf{S}$ is column orthogonal we have

$$\theta_{n:} = \frac{\sum_{l=1}^{L} s_{ln}\mathbf{x}_{l:}}{\sum_{l=1}^{L} s_{ln}} = \frac{1}{c_n}\sum_{l\in C_n} x_{l:}. \quad (22)$$

with c_n the cardinal of cluster C_n, so that it corresponds to the average of all FC belonging to the cluster C_n. Then the first term of Eq. (19) is:

$$Tr(\mathbf{XX}^t) = || \mathbf{X} ||_F^2 = Cste \quad (23)$$

By noticing that $\mathbf{SS}^\dagger$ is a projection matrix (and $(\mathbf{SS}^\dagger)^t\mathbf{SS}^\dagger = \mathbf{SS}^\dagger$), the third term is:

$$Tr(\mathbf{S\Theta\Theta^t S^t}) = Tr((\mathbf{SS}^\dagger)^t\mathbf{SS}^\dagger\mathbf{XX^t}) \quad (24)$$

$$= Tr(\mathbf{S\Theta X^t}), \quad (25)$$

This proves the passage from Eq. (19) to (20). Finally:

$$Tr(\mathbf{S\Theta X^t}) = Tr(\mathbf{S}(\mathbf{S^t S})^{-1}\mathbf{SXX^t}), \quad (26)$$

$$= Tr((\mathbf{S^t S})^{-\frac{1}{2}}\mathbf{S^t XX^t S}(\mathbf{S^t S})^{-\frac{1}{2}}), \quad (27)$$

$$= || \mathbf{A^t X} ||_F^2, \quad (28)$$

where $\mathbf{A} \in \mathbb{R}^{L\times N}$ is to the normalized indication matrix, $\mathbf{A} = \mathbf{S}(\mathbf{S^t S})^{-\frac{1}{2}}$, such that:

$$a_{ln} = \begin{cases} \frac{1}{\sqrt{c_n}} & \text{if } l \in C_n, \\ 0 & \text{else} \end{cases}. \quad (29)$$

As $|| \mathbf{X} ||_F^2$ is fixed, we can see that minimizing Eq. (6) is equivalent to maximizing the second element of Eq. (20). By noticing that columns of $\mathbf{A}$ form a standard basis of a subspace of dimension K of $\mathbb{R}^{L\times L}$, the optimization problem (6) becomes equivalent to [60]:

$$\underset{\mathbf{A}}{\operatorname{argmax}} \quad || \mathbf{A}^t\mathbf{X} ||_F^2 \quad (30)$$

$$s.t. \quad \mathbf{a}_{:i} \text{ standard basis vector} \quad (31)$$

Appendix B
More about the proposed tensor decomposition

A. On the optimal result of (L,T,1) HOSVD

We want to find unitary matrix matrix $\bar{\mathbf{U}} \in \mathbb{R}^{L\times L}$ and $\bar{\mathbf{V}} \in \mathbb{R}^{T\times T}$ and a column orthogonal matrix $\bar{\mathbf{W}}^{(K_S)} \in \mathbb{R}^{S\times K_S}$ minimizing the criteria ($\forall K_S \in 1, ..., S$):

$$|| \mathbf{X}_{(L)} - \bar{\mathbf{U}}\bar{\mathbf{G}}(\bar{\mathbf{W}}^{(K_S)} \otimes \bar{\mathbf{V}})^t ||_F^2, \quad (32)$$

Where $\bar{\mathbf{G}} \in \mathbb{R}^{L\times T\times K}$ correspond to $\bar{\mathbf{G}} = \bar{\mathbf{U}}^t\mathbf{X}_{(1)}(\bar{\mathbf{W}}^{(K_S)} \otimes \bar{\mathbf{V}})$ The optimal solution is obtained by Algorithm 1 and considering the matrix $\mathbf{W}^{(K)} = \mathbf{W}_{:[1,...,K]}$. We prove it by noting that the Frobenius norm of a tensor is the same as the Frobenius norm of his D-mode matricization (including the third):

$$J = || \mathbf{X}_{(S)} - \mathbf{W}^{(K_S)}\mathbf{G}_{(S)}^{(L,T,K_S)}(\mathbf{V} \otimes \mathbf{U})^t ||_F^2, \quad (33)$$

Also we wave:

$$\mathbf{G}_{(S)}^{L,T,K} = \mathbf{I}_{\mathbf{S}\to\mathbf{K_S}}\mathbf{G}_{(S)}, \quad (34)$$

Where $\mathbf{I}_{\mathbf{S}\to\mathbf{K_S}} \in \mathbb{R}^{K_S\times S}$ is a rectangular diagonal matrix selecting only the first K rows of the matrix $\mathbf{G}_{(S)}$. Moreover by using the third mode matricization of the full rank HOSVD decomposition we can easily show that:

$$\mathbf{G}_{(S)}(\mathbf{V} \otimes \mathbf{U})^t = \mathbf{W^t X}_{(S)}, \quad (35)$$

Eq. (33) becomes (with $\mathbf{X}_{(S)} = \mathbf{W\Sigma Z}^t$ the SVD of $\mathbf{X}_{(S)}$):

$$J = || \mathbf{X}_{(S)} - \mathbf{W}^{(K_S)}\mathbf{I}_{\mathbf{S}\to\mathbf{K}}\mathbf{W}^t\mathbf{X}_{(S)} ||_F^2, \quad (36)$$

$$= || \mathbf{X}_{(S)} - \mathbf{W}^{(K_S)}\mathbf{W}^{(K_S)t}\mathbf{X}_{(S)} ||_F^2, \quad (37)$$

$$= || \mathbf{W\Sigma Z}^t - \mathbf{W}^{(K_S)}\mathbf{\Sigma}^{(K_S)}\mathbf{Z}^{(K_S)t} ||_F^2, \quad (38)$$

$$= \sum_{s=K_S+1}^{S} \sigma_s^2. \quad (39)$$

This corresponds, using the Eckart-Young-Mirsky theorem, to:

$$\underset{rk(\mathbf{X})=K_S}{min} || \mathbf{X}_{(S)} - \mathbf{X} ||_F^2 \quad (40)$$

Then, we can conclude that the (L, T, K_S)-core tensor reduction via HOSVD minimize the criteria (32) because it is the best low rank approximation of the mode-3 matricization of $\boldsymbol{\mathcal{X}}$.

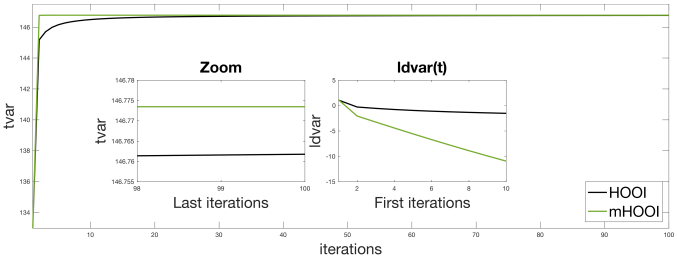

Fig. 8: $tvar(t)$ for the first 100 iterations using the tensor $\boldsymbol{\mathcal{X}}_{model}$ as input of the algorithm with a zoom of the last iteration and the first value of $ldvar(t)$

B. Note on empirical convergence of mHOOI

To empirically observe the convergence towards the optimal solution, with faster performance than HOOI algorithm we propose the following experiment. We consider a tensor $\boldsymbol{\mathcal{X}}_{model} \in \mathbb{R}^{66\times1000\times4}$ which corresponds to the model of Section V-A, with a noise vector of $b = [0.2, 1, 1, 0.3]$. We fix $K_L = K_T = 4$ and $K_S = 1$. We compare the performance of both HOOI and mHOOI algorithm by computing the total variance at each iteration $tvar(t) = || \mathbf{G}_i ||_F^2$ for $HOOI$, $tvar(t) = || \mathbf{F}_i ||_F^2$ (since they are matrices containing all the variance) and the log differential of the total variance between two consecutive iterates $ldvar(t) = log(tvar(t) - tvar(t-1)))$. Fig. 8 shows the mean of $tvar(t)$ for the first 100 iterations using 100 realisations of the tensor $\boldsymbol{\mathcal{X}}_{model}$ as input of the algorithm. A zoom of the last iteration is provided as well as the first value of $ldvar(t)$. The new proposed algorithm has always better performances than HOOI, with exponential convergence, on this scenario. This experiment, tested for other configurations, gave similar results.

References

[1] M. Guenot, J. Isnard, P. Ryvlin, C. Fischer, K. Ostrowsky, F. Mauguiere, and M. Sindou, "Neurophysiological monitoring for epilepsy surgery: the talairach seeg method," *Stereotactic and functional neurosurgery*, vol. 77, no. 1-4, pp. 29–32, 2001.

[2] P. Chauvel, S. Rheims, A. McGonigal, and P. Kahane, "French guidelines on stereoelectroencephalography (seeg): Editorial comment." *Neurophysiologie clinique= Clinical neurophysiology*, vol. 48, no. 1, p. 1, 2018.

[3] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, and D. Marinazzo, "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization," *Progress in neurobiology*, vol. 121, pp. 19–35, 2014.

[4] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[5] G. Ortiz-Jiménez, M. Coutino, S. P. Chepuri, and G. Leus, "Sampling and reconstruction of signals on product graphs," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 713–717.

[6] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3675–3687, July 2017.

[7] F. Sheikholeslami and G. B. Giannakis, "Overlapping community detection via constrained parafac: A divide and conquer approach," in *2017 IEEE International Conference on Data Mining (ICDM)*, Nov 2017, pp. 127–136.

[8] L. Gauvin, A. Panisson, and C. Cattuto, "Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach," *PLOS ONE*, vol. 9, no. 1, pp. 1–13, 01 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0086028

[9] R. Hamon, P. Borgnat, P. Flandrin, and C. Robardet, "Extraction of temporal network structures from graph-based signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 215–226, June 2016.

[10] ——, "Nonnegative matrix factorization to find features in temporal networks," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1065–1069.

[11] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.

[12] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[13] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[15] A. N. Khambhati, K. A. Davis, B. S. Oommen, S. H. Chen, T. H. Lucas, B. Litt, and D. S. Bassett, "Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy," *PLoS computational biology*, vol. 11, no. 12, p. e1004608, 2015.

[16] A. Hegde, D. Erdogmus, and J. C. Principe, "Spatio-temporal clustering of epileptic ecog," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 4199–4202.

[17] J. Gonzalez-Castillo, C. W. Hoy, D. A. Handwerker, M. E. Robinson, L. C. Buchanan, Z. S. Saad, and P. A. Bandettini, "Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns," *Proceedings of the National Academy of Sciences*, vol. 112, no. 28, pp. 8762–8767, 2015.

[18] F. Liu, Y. Wang, M. Li, W. Wang, R. Li, Z. Zhang, G. Lu, and H. Chen, "Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic–clonic seizure," *Human brain mapping*, vol. 38, no. 2, pp. 957–973, 2017.

[19] L. Geerligs, N. M. Maurits, R. J. Renken, and M. M. Lorist, "Reduced specificity of functional connectivity in the aging brain during task performance," *Human brain mapping*, vol. 35, no. 1, pp. 319–330, 2014.

[20] X. Shi, W. Wang, and C. Zhang, "An Empirical Comparison of Latest Data Clustering Algorithms with State-of-the-Art," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 410–415, 2017.

[21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, 1996, pp. 226–231.

[22] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *science*, vol. 315, no. 5814, pp. 972–976, 2007.

[23] C. Bauckhage, "K-means clustering is matrix factorization," *arXiv preprint arXiv:1512.07548*, 2015.

[24] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 29.

[25] K. Nur'aini, I. Najahaty, L. Hidayati, H. Murfi, and S. Nurrohmah, "Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter," in *Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference on*. IEEE, 2015, pp. 123–128.

[26] A. Ben-Hur and I. Guyon, "Detecting stable clusters using principal component analysis," in *Functional genomics*. Springer, 2003, pp. 159–182.

[27] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A practical approach to microarray data analysis*. Springer, 2003, pp. 91–109.

[28] A. H. Hossny, T. Moschuo, G. Osborne, L. Mitchell, and N. Lothian, "Enhancing keyword correlation for event detection

in social networks using SVD and k-means: Twitter case study," *Social Network Analysis and Mining*, vol. 8, no. 1, p. 49, 2018.

[29] F. Pisana, T. Henzler, S. Schönberg, E. Klotz, B. Schmidt, and M. Kachelrieß, "High quality high spatial resolution functional classification in low dose dynamic CT perfusion using singular value decomposition (SVD) and k-means clustering," in *Medical Imaging 2017: Physics of Medical Imaging*, vol. 10132. International Society for Optics and Photonics, 2017, p. 101320M.

[30] A. A. Shabalin and A. B. Nobel, "Reconstruction of a low-rank matrix in the presence of Gaussian noise," *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, 2013.

[31] D. Feldman, M. Schmidt, and C. Sohler, "Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering," in *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2013, pp. 1434–1453.

[32] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay, "Clustering large graphs via the singular value decomposition," *Machine learning*, vol. 56, no. 1-3, pp. 9–33, 2004.

[33] K. Allab, L. Labiod, and M. Nadif, "A Semi-NMF-PCA unified framework for data clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 2–16, 2017.

[34] H. Zou and L. Xue, "A selective overview of sparse principal component analysis," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1311–1320, 2018.

[35] N. Leonardi, "Dynamic brain networks explored by structure-revealing methods," EPFL, Tech. Rep., 2014.

[36] L. R. Chai, A. N. Khambhati, R. Ciric, T. M. Moore, R. C. Gur, R. E. Gur, T. D. Satterthwaite, and D. S. Bassett, "Evolution of brain network dynamics in neurodevelopment," *Network Neuroscience*, vol. 1, no. 1, pp. 14–30, 2017.

[37] A. Cichocki, "Tensor decompositions: a new concept in brain data analysis?" *arXiv preprint arXiv:1305.0395*, 2013.

[38] M. J. Tobia, K. Hayashi, G. Ballard, I. H. Gotlib, and C. E. Waugh, "Dynamic functional connectivity and individual differences in emotions during social stress," *Human brain mapping*, vol. 38, no. 12, pp. 6185–6205, 2017.

[39] A. G. Mahyari, D. M. Zoltowski, E. M. Bernat, and S. Aviyente, "A tensor decomposition-based approach for detecting dynamic network states from eeg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 225–237, 2017.

[40] A. Ozdemir, E. M. Bernat, and S. Aviyente, "Recursive tensor subspace tracking for dynamic brain network analysis," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 4, pp. 669–682, 2017.

[41] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.

[42] M. Mørup, L. K. Hansen, and S. M. Arnfred, "Algorithms for sparse nonnegative tucker decompositions," *Neural computation*, vol. 20, no. 8, pp. 2112–2131, 2008.

[43] A. Zhang and D. Xia, "Tensor SVD: Statistical and Computational Limits," *IEEE Transactions on Information Theory*, 2018.

[44] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.

[45] D. Yang, Z. Ma, and A. Buja, "A sparse singular value decomposition method for high-dimensional data," *Journal of Computational and Graphical Statistics*, vol. 23, no. 4, pp. 923–942, 2014.

[46] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski, "SPAMS: A SPArse Modeling Software, v2. 3," *URL http://spams-devel. gforge. inria. fr/downloads. html*, 2014.

[47] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[48] C. C. Aggarwal and C. K. Reddy, *Data clustering: algorithms and applications*. CRC press, 2013.

[49] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[50] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.

[51] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 432–441, 2009.

[52] Y.-D. Kim and S. Choi, "Nonnegative tucker decomposition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[53] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[54] Y. Xu, "On the convergence of higher-order orthogonal iteration," *Linear and Multilinear Algebra*, vol. 66, no. 11, pp. 2247–2265, 2018.

[55] C. A. Andersson and R. Bro, "The N-way toolbox for MATLAB," *Chemometrics and intelligent laboratory systems*, vol. 52, no. 1, pp. 1–4, 2000.

[56] B. W. Bader and T. G. Kolda, "Matlab tensor toolbox version 2.5," *Available online, January*, vol. 7, 2012.

[57] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[58] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.

[59] F. Wendling, F. Bartolomei, J.-J. Bellanger, J. Bourien, and P. Chauvel, "Epileptic fast intracerebral eeg activity: evidence for spatial decorrelation at seizure onset," *Brain*, vol. 126, no. 6, pp. 1449–1459, 2003.

[60] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine learning refined: foundations, algorithms, and applications*. Cambridge University Press, 2016.