

Extraction of time-varying spatio-temporal networks using parameter-tuned constrained IVA

Suchita Bhinge*, Rami Mowakeaa, Vince D. Calhoun, and Tülay Adalı

Abstract—Dynamic functional connectivity (dFC) analysis is an effective way to capture the networks that are functionally associated and continuously changing over the scanning period. However, these methods mostly analyze the dynamic associations across the activation patterns of the spatial networks while assuming that the spatial networks are stationary. Hence, a model that allows for the variability in both domains and reduces the assumptions imposed on the data provides an effective way for extracting spatio-temporal networks. Independent vector analysis is a joint blind source separation technique that allows for estimation of spatial and temporal features while successfully preserving variability. However, its performance is affected for higher number of datasets. Hence, we develop an effective two-stage method to extract time-varying spatial and temporal features using IVA, mitigating the problems with higher number of datasets while preserving the variability across subjects and time. The first stage is used to extract reference signals using group independent component analysis (GICA) that are used in a parameter-tuned constrained IVA (pt-cIVA) framework to estimate time-varying representations of these signals by preserving the variability through tuning the constraint parameter. This approach effectively captures variability across time from a large-scale resting-state fMRI data acquired from healthy controls and patients with schizophrenia and identifies more functionally relevant connections that are significantly different among healthy controls and patients with schizophrenia, compared with the widely used GICA method alone.

Index Terms—Blind source separation, connectivity analysis, dimensionality reduction, fMRI analysis

I. INTRODUCTION

DYNAMIC functional connectivity (dFC) analysis has emerged due to evidence that the human brain exhibits changes in functional patterns over the scanning period [1]. A number of studies have shown the presence of multiple structured patterns corresponding to different functional connectivity in task-related and resting-state functional magnetic resonance imaging (fMRI) data, see *e.g.*, [2], [3], [4], [5], [6]. Analyzing these connectivity patterns in resting-state data has enabled the identification of distinct biomarkers in a variety of disorders such as schizophrenia [7], bipolar disorder [8], autism [9], [10], post-traumatic stress [11], generalized anxiety

disorder [12], attention deficit hyperactivity disorder [13] and mild cognitive impairment [14]. Studies have also shown changes in functional connectivity patterns in different stages of development [15] and due to hallucinations [16].

Most dFC analysis techniques examine time-varying associations among the activation patterns of spatial networks while assuming that the spatial evolution of the networks is stationary. However, studies have shown that changes in functional connectivity patterns imply changes in the spatial networks [17], [18], [19]. Region of interest (ROI)-based analyses on resting-state networks (RSNs) have shown better classification of subjects when variability in both spatial and temporal domains is considered compared with variability assumed in either spatial or temporal domain [17], [18]. Dynamic mode decomposition (DMD), a spatio-temporal modal decomposition technique, has demonstrated changes in the temporal activation of RSNs [19]. Although these techniques provide interesting results, the use of pre-defined RSNs causes the estimated functional connectivity to be sensitive to network selection whereas DMD requires significant dimension reduction that may restrict the method to estimation of few spatial components. Hence a more flexible model that simultaneously captures both time-varying patterns and spatial networks of the whole brain is desirable. Joint blind source separation techniques such as group independent component analysis (GICA) attempt to find a common spatial subspace whereas techniques such as joint ICA assumes a common temporal subspace. Independent vector analysis (IVA) relaxes these assumptions and estimates demixing matrices in order to decompose the data into dataset-specific time courses and spatial maps, providing an attractive approach for capturing spatially varying networks. IVA has been successfully applied to fMRI data to capture variability in spatial networks of patients with schizophrenia and healthy controls [20]. The method proposed in [20] divides each subject's data into overlapping windows and performs IVA on this setup treating each window as a dataset. While this approach successfully captures the dynamics in the spatial networks, it was limited to a small number of subjects (20 in the given case) due to curse of dimensionality to which IVA is susceptible. The flexibility of IVA comes at a cost that for a fixed number of samples, its performance degrades with the increase in number of datasets and number of sources since it requires estimation of high-dimensional probability density functions in addition to the increase in number of estimation parameters for the demixing matrices.

Analysis of fMRI data on a large number of subjects is common in order to obtain reliable results that can be used

Manuscript received October 23, 2018; revised January 6, 2019. This work is supported in parts by the National Institute of Biomedical Imaging and Bioengineering under Grant R01 EB 020407, National Science Foundation under Grant 1631838 and National Science Foundation-Computing and Communication Foundations under Grant 1618551. Asterisk indicates corresponding author (e-mail: suchita1@umbc.edu).

S. Bhinge*, R. Mowakeaa and T. Adalı are with the Department of Electrical and Computer Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250 USA.

V. D. Calhoun is with The Mind Research Network, Albuquerque, NM 87106 USA, and also with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131 USA.

to make robust inferences. We develop a two-stage procedure while addressing two important points: (1) use of a flexible model, like IVA, that captures variability in both spatial and temporal domain, and (2) address the performance degradation with high dimensionality in IVA, while preserving the variability in both domains for application to large scale fMRI data. One way to reduce the effect of high dimensionality in IVA is through the use of reference signals to limit the size of the solution space. Constrained IVA (cIVA) is a semi-blind source separation technique that incorporates information regarding reference signals [21]. However, it requires a user-defined constraint parameter that controls the influence of the reference signals on the source estimates. A higher value might constrain the source estimate more than necessary, thus affecting the model's ability to capture variability. Hence, we propose a new technique, parameter-tuned cIVA (pt-cIVA), to adaptively tune the constraint parameter to effectively capture the variability of these reference signals across time points. The two-stage procedure that includes extraction of reference signals through a data-driven approach and their use in pt-cIVA, enables us to capture time-varying features in the temporal and spatial domain while preserving variability across time windows and reduce the undesirable effects of high dimensionality by enabling analysis of each subject at a time.

The remainder of this paper is organized as follows: Section II introduces IVA and relevant IVA algorithms, followed by an overview of cIVA. It talks about an algorithm that jointly accounts higher order statistics (HOS) and second order statistics (SOS), namely, IVA-L-SOS, that provides a more suitable model for fMRI data and introduces the adaptive parameter tuning technique for cIVA. The proposed method to obtain spatio-temporal dynamics is described in Section III, Section IV and Section V shows the results on simulated and resting-state fMRI data.

II. METHODS

IVA is an extension of ICA to multiple datasets and estimates source components that are as statistically independent as possible within each dataset while accounting for dependence across datasets. Given M datasets, each comprised of L components, $\mathbf{x}^{[m]} \in \mathbb{R}^L$, $m = 1, \dots, M$, we have, $\mathbf{x}^{[m]} = \mathbf{A}^{[m]} \mathbf{s}^{[m]}$, $m = 1, \dots, M$, where $\mathbf{A}^{[m]} \in \mathbb{R}^{L \times L}$ is the mixing matrix. Given a set of observations, the IVA model can be written as $\mathbf{X}^{[m]} = \mathbf{A}^{[m]} \mathbf{S}^{[m]}$, $\mathbf{X}^{[m]} \in \mathbb{R}^{L \times V}$, the rows of $\mathbf{S}^{[m]} \in \mathbb{R}^{L \times V}$ are latent sources and V is the number of samples/voxel. IVA estimates M demixing matrices, $\mathbf{W}^{[m]}$, to compute the source estimates, $\hat{\mathbf{s}}^{[m]} = \mathbf{W}^{[m]} \mathbf{x}^{[m]}$, by minimizing the cost function given as [22], [23],

$$\mathcal{J}_{\text{IVA}} = \sum_{l=1}^L \left[\sum_{m=1}^M \mathcal{H}(\hat{s}_l^{[m]}) - \mathcal{I}(\hat{\mathbf{s}}_l) \right] - \sum_{m=1}^M \log |\det \mathbf{W}^{[m]}|, \quad (1)$$

where $\mathcal{H}(\hat{s}_l^{[m]})$ denotes the entropy of the l th source estimate for the m th dataset, and $\mathcal{I}(\mathbf{s}_l)$ denotes the mutual information of the l th source component vector (SCV), $\hat{\mathbf{s}}_l^T = [\hat{s}_l^{[1]}, \dots, \hat{s}_l^{[M]}]$. The optimization of the cost function jointly weighs the independence within the dataset through the entropy term along with the log determinant term and

dependence across the datasets through the mutual information term. SCV takes into account the dependence across the datasets and the l th SCV is formed by concatenating the l th component from all the M datasets as shown in Figure 1. For multisubject fMRI analysis, the M datasets corresponds to M subjects, forming an SCV of a similar brain region from M subjects. In this paper, the M datasets correspond to M time windows obtained using a sliding-window on a single subject's data, yielding an SCV that gives the time-evolution of a brain region and the corresponding columns of the mixing matrix are the time courses.

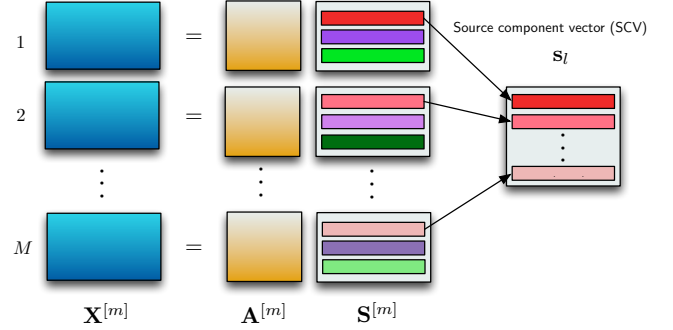


Fig. 1. Given a set of observations, the IVA model is given as $\mathbf{X}^{[m]} = \mathbf{A}^{[m]} \mathbf{S}^{[m]}$, $m = 1, \dots, M$, where $\mathbf{A}^{[m]}$ is the mixing matrix and the rows in $\mathbf{S}^{[m]}$, are the latent sources that are dependent across datasets. These dependent sources are concatenated together to form an SCV.

A. Choice of IVA algorithm

The assumption of a different model for the latent source distribution has led to the development of different IVA algorithms. IVA-Gaussian (IVA-G) assumes that the underlying SCVs are multivariate Gaussian [23], and thus only takes SOS into account and estimates the covariance matrix for each SCV, $\Sigma_1 \in \mathbb{R}^{M \times M}$. IVA-Laplacian (IVA-L) assumes the sources are multivariate Laplacian distributed [24] and takes only HOS into account. It assumes there is no second-order correlation within each SCV, i.e., the covariance matrix is an identity matrix for all SCVs. Although the assumption of no second-order correlation favors some applications, in many others, such as fMRI, it degrades the estimation performance since fMRI sources exhibit a significant level of correlation across windows [25], [26]. IVA-GL, another implementation of IVA that performs IVA-L initialized to the result of IVA-G, is a popular method for fMRI analysis since it has shown more robust performance than using IVA-G or IVA-L alone [23] and since it benefits from the advantages of both algorithms, although sequentially [26]. Since the SCVs for fMRI applications correspond to brain regions that have multivariate heavy-tailed distributions, like the multivariate Laplacian distribution, with a significant level of correlation across subjects/time windows, an algorithm that simultaneously exploit the benefits of these algorithms is preferable. In this paper, we use an IVA algorithm, IVA-L-SOS, that assumes the sources are multivariate Laplacian distributed, like IVA-L, but also takes second-order correlation of the SCVs into account, like IVA-G, for full statistical characterization of a Laplacian multivariate random vector.

1) *IVA-L-SOS algorithm*: The multivariate generalized Gaussian distribution (MGGD) covers a wide range of unimodal distributions by controlling the shape parameter, β , such as super-Gaussian ($\beta < 1$), normal ($\beta = 1$) and sub-Gaussian ($\beta > 1$), and assumes second-order correlation within an SCV [27]. The MGGD is given by,

$$p(\mathbf{s}; \Sigma, \beta) = \frac{M\Gamma(M/2) |\Sigma|^{-1/2}}{\pi^{[M/2]}\Gamma(1 + \frac{M}{2\beta})2^{1+\frac{M}{2\beta}}} e^{-\frac{1}{2}[\mathbf{s}^T \Sigma^{-1} \mathbf{s}]^\beta}, \quad (2)$$

where Σ is a positive definite scatter matrix and $\Gamma(\cdot)$ is the Gamma function. By setting the shape parameter β to 0.5, the MGGD distribution is equivalent to a multivariate Laplacian distribution that accounts for second-order correlation through Σ and is expressed as,

$$p(\mathbf{s}; \Sigma) = \frac{\Gamma(M/2) |\Sigma|^{-1/2}}{2^{M+1} \pi^{[M/2]} \Gamma(M)} \exp \left\{ -\frac{1}{2} \sqrt{\mathbf{s}^T \Sigma^{-1} \mathbf{s}} \right\}, \quad (3)$$

where Σ estimated at each iteration. Since fMRI sources are in general expected to have a super-Gaussian distribution, like Laplacian [26], and are dependent across subjects/windows, the IVA-L-SOS model is a good match for fMRI data.

Derivative of (1) with respect to $\mathbf{w}_l^{[m]}$ is given by,

$$\frac{\partial \mathcal{J}_{\text{IVA}}}{\partial \mathbf{w}_l^{[m]}} = E \left\{ \phi_l^{[m]} \mathbf{x}^{[m]} \right\} - \frac{\mathbf{u}_l^{[m]}}{\left(\mathbf{u}_l^{[m]} \right)^T \mathbf{w}_l^{[m]}}, \quad (4)$$

where $\mathbf{u}_l^{[m]}$ is a vector such that $\tilde{\mathbf{W}}^{[m]} \mathbf{u}_l^{[m]} = 0$, and $\tilde{\mathbf{W}}^{[m]}$ is formed by removing the l th row of $\mathbf{W}^{[m]}$. The last term in (1) can be written as $|\det(\mathbf{W}^{[m]})| = \left| \left(\mathbf{u}_l^{[m]} \right)^T \mathbf{w}_l^{[m]} \right| \sqrt{|\det(\tilde{\mathbf{W}}^{[m]} (\tilde{\mathbf{W}}^{[m]}))|}$ [28], [23], where $\sqrt{|\det(\tilde{\mathbf{W}}^{[m]} (\tilde{\mathbf{W}}^{[m]}))|}$ is independent of $\mathbf{w}_l^{[m]}$. The score function, $\phi_l^{[m]}$, for the IVA-GGD algorithm is given by,

$$\phi(\mathbf{s}) = \left\{ \frac{\Gamma\left(\frac{M+2}{2\beta}\right)}{M\Gamma\left(\frac{M}{2\beta}\right)} \right\}^\beta \beta \Sigma^{-1} \mathbf{s} (\mathbf{s}^T \Sigma^{-1} \mathbf{s})^{\beta-1}, \quad (5)$$

where the Gamma functions, $\Gamma\left(\frac{M+2}{2\beta}\right)$ and $\Gamma\left(\frac{M}{2\beta}\right)$ grow at a rate faster than the exponential function towards infinity as M increases, leading the score function to be undefined. Since $\beta = 0.5$ provides a better match for fMRI sources [26], by direct substitution of $\beta = 0.5$, which corresponds to multivariate Laplacian distribution, in (5), we obtain

$$\phi(\mathbf{s}) = (M+1)^{0.5} \frac{\Sigma^{-1} \mathbf{s}}{\sqrt{\mathbf{s}^T \Sigma^{-1} \mathbf{s}}},$$

which also reduces the effect of high dimensionality and enables a stable version for large M . However, there are other factors that affect the performance of IVA when the number of datasets is high. We discuss these factors in the next section.

2) *IVA: Negative effect of high dimensionality*: Although IVA provides a desirable framework for capturing time-varying spatial maps and time courses, and IVA-L-SOS mitigates the negative effects of large M , IVA requires the estimation of high-dimensional probability density functions (of dimension

M) for the SCVs and with increasing number of datasets, M , and number of sources, L , its performance degrades. It estimates M demixing matrices of dimension $L \times L$ yielding a total of ML^2 parameters. Algorithms such as IVA-G and IVA-L-SOS exploit second-order correlation and require the estimation of the scatter matrix, $\Sigma_l \in \mathbb{R}^{M \times M}$, $l = 1, \dots, L$, yielding $L \times M(M-1)/2$ parameters to be estimated. Thus, the number of parameters to be estimated increases linearly for $\mathbf{W}^{[m]}$ and quadratically for Σ_l with respect to the increase in M and L . Hence, for a fixed number of samples the performance of IVA degrades when a large number of sources and datasets is considered. To demonstrate this effect on IVA, we generate M datasets with $L = 3$ sources generated from a multivariate Laplacian distribution with $V = 10^4$ samples. The observations, $\mathbf{x}^{[m]}(v)$, are obtained using $\mathbf{x}^{[m]}(v) = \mathbf{A}^{[m]} \mathbf{s}^{[m]}(v)$, where $\mathbf{A}^{[m]} \in \mathbb{R}^{L \times L}$ is generated randomly from a uniform distribution. We obtain 10 estimates of the demixing matrices, $\mathbf{W}^{[m]}$, using four IVA algorithms: IVA-L, IVA-G, IVA-GL and IVA-L-SOS. The performance is measured in terms of joint inter-symbol interference (jISI) [23]. The jISI metric measures the ability of the algorithm to separate the sources ($0 \leq \text{jISI} \leq 1$), where 0 indicates better separation of underlying SCVs, i.e., $\mathbf{W}^{[m]} \mathbf{A}^{[m]} = \mathbf{I}$, $\forall m \in \{1, \dots, M\}$. The average of the jISI metric measured over 50 runs for each algorithm is shown in Fig. 2. IVA-GL and IVA-L-

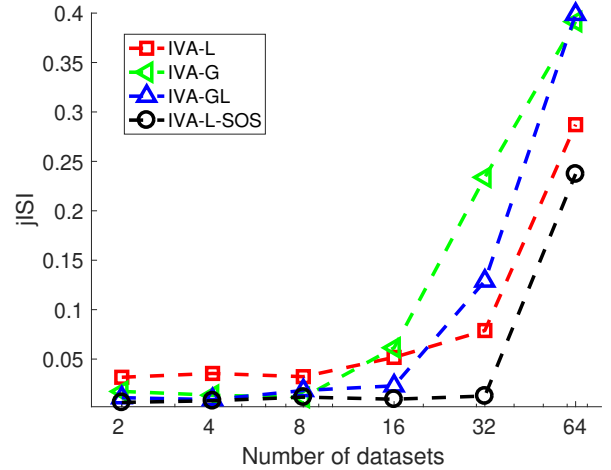


Fig. 2. Performance of three IVA algorithms, namely, IVA-L, IVA-G, IVA-GL and IVA-L-SOS, with respect to number of datasets in terms of jISI. Performance of IVA degrades after a certain number of datasets for a fixed number of samples and hence cannot be used for large number of datasets with fixed number of samples.

SOS both have similar performance for a lower number of datasets since both exploit signal properties that match the underlying source distribution: HOS and SOS. However, for IVA-G, IVA-GL and IVA-L the performance degrades at a faster rate for a large number of datasets as compared with IVA-L-SOS. The increase in jISI value for IVA-G algorithm with increase in number of datasets indicates that even for algorithms that have a convex cost function, as for IVA-G [23], the performance is affected due to increase in number of datasets. This indicates that the performance of IVA starts degrading as the number of samples available to estimate $Q = ML^2 + L \times M(M-1)/2$ parameters is approximately

less than $5Q$, hence degrading the estimation of the underlying sources. The method proposed in [20] performs IVA-GL on the windowed datasets from all 20 subjects, *i.e.*, a total of 140 datasets, and thus cannot be directly used for large number of subjects. Hence, we develop a two-stage procedure to reduce the effect of high dimensionality of IVA by performing pt-cIVA on windowed datasets of each subject. Before discussing the pt-cIVA technique, we first introduce the regular cIVA model in the next section.

B. Constrained IVA with fixed constraint parameter

Data-driven dFC analysis techniques, such as IVA, minimize the assumptions imposed on the data whereas model-driven dFC analysis techniques, such as ROI-based methods, make strong assumptions about the data making them robust to noise and other artifacts. However, the use of pre-defined RSNs limits the exploration of possible individual dynamic characteristics. Semi-blind source separation techniques, such as cIVA [21], [29], take advantage of the benefits of both techniques to efficiently capture individual-specific variability of the features extracted from the group of subjects. The cIVA method incorporates prior information about the sources into the IVA model [21] and limits size of the solution space, also addressing the high dimensionality issue. The cost function for cIVA is given as,

$$\mathcal{J} = \mathcal{J}_{\text{IVA}} - \sum_{l=1}^L \frac{1}{2\gamma_l} \sum_{m=1}^M \left\{ \left[\max\{0, \mu_l^{[m]} + \gamma_l g(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l)\} \right]^2 - (\mu_l^{[m]})^2 \right\}, \quad (6)$$

where $\mu_l^{[m]}$ is the regularization parameter, $(1/2\gamma_l)$ is the penalty parameter and $g(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l)$ is the inequality constraint function given as,

$$g(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l) = \rho_l - \epsilon(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l) \leq 0, \quad (7)$$

where $\hat{\mathbf{s}}_l^{[m]} = (\mathbf{w}_l^{[m]})^T \mathbf{x}^{[m]}$ is the estimated component, \mathbf{d}_l denotes the reference vector for the l th SCV, s_l , $\epsilon(\cdot, \cdot)$ is a function that defines the measure of similarity between the estimated SCV and reference signal, and ρ_l is the constraint parameter. One can also constrain the columns of the mixing matrix and the interested reader can refer to [21] for the procedure. The definition of the constraint function as in (7) allows for the use of different dissimilarity functions such as the inner product, mean square error, mutual information and correlation. In this paper, we use the absolute value of Pearson's correlation coefficient given by,

$$\epsilon(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l) = \left| \text{corr}(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l) \right|.$$

The absolute value of Pearson's correlation coefficient as a similarity measure restricts $\epsilon(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_l)$ to be between 0 and 1, steering $\rho_l \leq 1$. Thus, a higher value of ρ_l enforces the estimated source to be exactly similar to the reference signal, not allowing the reference component to vary across datasets, whereas a lower value results in the estimated component to deviate from the reference signal making it to be prone to

noise and other artifacts. Hence, the selection of ρ_l plays a crucial role in the performance of the cIVA algorithm. In this paper we propose an adaptive technique to select the constraint parameter, namely, pt-cIVA, which we introduce in the next section.

C. Parameter-tuned constrained IVA (pt-cIVA)

The use of reference signals provides an effective way to address the high dimensionality issue, however, use of a fixed value for the constraint parameter does not allow the model to efficiently capture the variability. We introduce pt-cIVA method such that it controls the amount of correspondence between the estimated source and the reference signal. In this case, the reference signals are the group components estimated using GICA and exhibit variability across time windows [17], [18], [19]. Using a fixed constraint parameter controls the amount of correspondence between the reference signal and estimated source, and constrains the variability of the reference signals across time windows. Hence, we adaptively tune the constraint parameter during the optimization of cIVA. We define N as the number of constraints, ρ_n the constraint parameter corresponding to the n th constraint and d_n , $n = 1, \dots, N$, as the reference signal used to constrain a single source among the L sources. The adaptive ρ -tuning method selects a value for ρ_n from a set of possible values for ρ_n , denoted as \mathcal{P} . We randomly initialize the demixing matrices, $\mathbf{W}^{[m]}$, set $\mu_n^{[m]} = 0$ and γ_n to a positive scalar value. At each iteration, we obtain an estimate of the sources, $s_l^{[m]}$, $m = 1, \dots, M$, $l = 1, \dots, L$ and select the first SCV to be the constrained component followed by an estimation of $\mu_n^{[m]}$ and $\hat{\rho}_n$ as given in line 8 and 9 of Algorithm 1, respectively. The update given in line 9 in Algorithm 1 identifies the highest value of ρ_n from set \mathcal{P} that satisfies the condition in (7) for all M datasets. Hence the distance of the estimated correlation, $\epsilon(\cdot, \cdot)$ is computed from all possible values of ρ_n from set \mathcal{P} , across all datasets, and the value of ρ_n with least distance is selected. The new value of the constraint parameter, $\hat{\rho}_n$, is then used to compute the gradient, $\partial \mathcal{J} / \partial \mathbf{w}_l^{[m]}$, and update the demixing matrix as in line 9 followed by obtaining a new estimate of the sources. The process is repeated until the convergence criterion, following the one proposed in [21], is met. Algorithm 1 describes the pt-cIVA technique. The parameter tuning technique improves the estimation of the constraint source at every iteration providing a better solution as compared with using a fixed ρ_n at every iteration.

III. IMPLEMENTATION

In this section, we present the methodology to capture time-varying spatial and temporal components using pt-cIVA. This method extracts steady-state representation of functionally relevant components from all subjects using GICA followed by performing pt-cIVA on each subject to obtain the time-varying representations of these components as shown in Figure 3.

A. Extraction of reference signals

GICA is one of the widely used, data-driven techniques used to extract components that are common across multiple subjects [30], [25]. Given datasets from K subjects, GICA first performs subject-level principal component analysis (PCA) in

Algorithm 1 pt-cIVA

Define set \mathcal{P} as possible values for ρ_n

2: **for** $n = 1, \dots, N$ **do**

 Randomly initialize demixing matrices, $[\mathbf{W}^{[1]}, \dots, \mathbf{W}^{[M]}]$ and set $\mu_n^{[m]} = 0, \gamma_n$ to be a positive scalar value

4: Compute $\hat{\mathbf{s}}^{[m]} = \mathbf{W}^{[m]} \mathbf{x}^{[m]}, m = 1, \dots, M$

for $l = 1, \dots, L$ **do**

6: **for** $m = 1, \dots, M$ **do**

if $l == 1$ **then**

8: $\mu_n^{[m]} = \max \left\{ 0, \gamma_n g \left(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_n \right) + \mu_n^{[m]} \right\}$

$\hat{\rho}_n = \arg \min_{\rho_n \in \mathcal{P}} \left[\min_m \left\{ \left| \rho_n - \left| \text{corr}(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_n) \right| \right| \right\}_{m=1}^M \right]$

10: $\partial \mathcal{J} / \partial \mathbf{w}_l^{[m]} = \partial \mathcal{J}_{\text{IVA}} / \partial \mathbf{w}_l^{[m]} - \frac{1}{\gamma_n} \left\{ \left[\max \left\{ 0, \gamma_n \left(\hat{\rho}_n - \epsilon(\hat{\mathbf{s}}_l^{[m]}, \mathbf{d}_n) \right) + \mu_n^{[m]} \right\} \right]^2 - \left(\mu_n^{[m]} \right)^2 \right\}$

else

12: $\partial \mathcal{J}_{\text{IVA}} / \partial \mathbf{w}_l^{[m]}$ using (4)

$\mathbf{w}_l^{[m]} = \mathbf{w}_l^{[m]} + \partial \mathcal{J}_{\text{IVA}} / \partial \mathbf{w}_l^{[m]}$

14: Repeat 3 to 13 until convergence

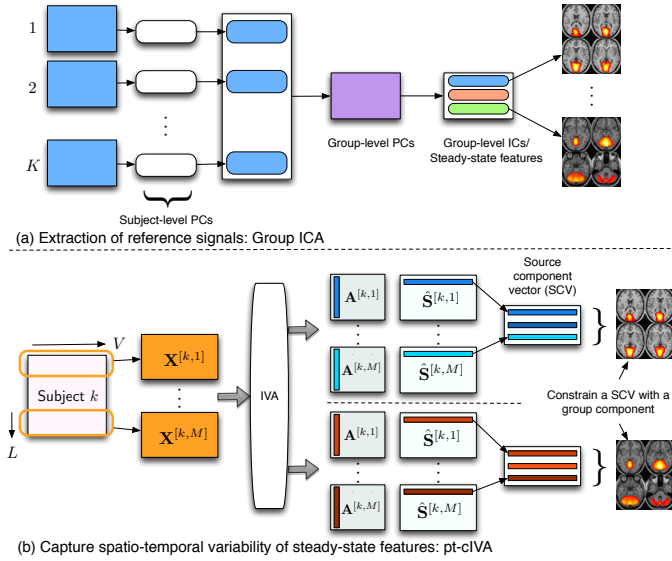


Fig. 3. The two-stage method for obtaining time-varying spatial networks and corresponding time courses. (a) Reference signals are obtained using group ICA from all subjects. (b) Each subject data is divided into windows and pt-cIVA is applied on the windowed datasets with features extracted from GICA used as reference.

order to obtain a lower-dimensional signal subspace. PCA estimates uncorrelated features in the order of highest variance, hence the signal subspace corresponds to features that capture most of the variability across time points and can be referred to as the steady-state representation of the time-varying spatial networks. GICA then vertically stacks the subject-level components from all subjects and performs a second group-level PCA on this matrix to obtain group-level principal components, which represent the components that account for most variability across subjects, *i.e.*, a common signal subspace. Since PCA estimates uncorrelated components that separates components using only second-order statistics, ICA is performed on the group-level principal components in order to obtain statistically independent spatial features. Among the estimated independent components (ICs), N functionally

relevant group components, denoted as $\mathbf{d}_n, n = 1, \dots, N$, are selected for further analysis. The N features are used as reference signals to obtain the variability of these components across windows by incorporating them into a sliding window pt-cIVA framework.

B. Parameter-tuned cIVA

In the second stage, we divide each subject's data into M windows of length L with an 50% overlap yielding a total of MK windows. Considering all the MK windows in the analysis results in IVA to model MK -dimensional SCVs resulting in MKL^2 and $LMK(MK - 1)/2$ parameters that need to be estimated from the fixed V samples. However, as discussed, the performance of IVA degrades with a large number of datasets and sources for a fixed number of samples. Thus, we perform a subject-level analysis to mitigate the high dimensionality issue by modeling a M -dimensional SCV instead of a MK -dimensional SCV by performing a subject-level IVA, where the windowed data from each subject defines a dataset. Using this setup, IVA also takes advantage of source dependence across windows since the spatial maps are expected to change smoothly across windows, thus aligning the components across windows. The N reference signals, $\mathbf{d}_n, n = 1, \dots, N$, obtained from GICA are used as constraints in pt-cIVA to constrain the first SCVs for each subject. The pt-cIVA technique we introduced in Section II-C enables since each window to have a different level of correlation with the constraint and setting a fixed value for the constraint parameter can deteriorate the estimation of the SCVs as shown in the simulation examples in Section IV.

IV. SIMULATION RESULTS

We generate $M = 5$ datasets such that $\mathbf{x}^{[m]} = \mathbf{A}^{[m]} \mathbf{s}^{[m]}, m = 1, \dots, M$, where the mixing matrix for each dataset, $\mathbf{A}^{[m]} \in \mathbb{R}^{L \times L}$, is randomly generated with elements drawn from a normal distribution with zero mean and unit variance. The $L = 10$ SCVs are formed from M -dimensional SCVs of $V = 10^4$ samples. Each SCV is generated from a

multivariate Laplacian distribution where the scatter matrix, Σ , has a AR-type correlation structure given as,

$$\Sigma_l = \begin{bmatrix} 1 & \psi & \psi^2 & \dots \\ \psi & 1 & \psi & \dots \\ \psi^2 & \psi & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (8)$$

Among the 10 SCVs, first five are generated with medium to high second-order correlation, $\psi \sim \mathcal{U}(0.5, 0.9)$, and the remaining five with lower second-order correlation, $\psi \sim \mathcal{U}(0.2, 0.5)$. A value of ψ between $[0.8, 0.9]$ models the components that have low variability across datasets, while a value of ψ between $[0.5, 0.8]$ models the components that have high variability. A value of ψ below 0.5 models the artifactual components. A reference signal is generated such that it has ρ_{true} correlation with the average component of the constrained SCV. We consider three scenarios to test the performance of our method to cover the range of possibilities: Scenario A: $\rho_{true} = 0.6$; Scenario B: $\rho_{true} = 0.3$; and Scenario C: $\rho_{true} = 0$. For each scenario, pt-cIVA is applied with the set \mathcal{P} defined as $0.001, \dots, 0.9$, $\gamma_n = 3$ and cIVA with fixed constraint parameter for 50 runs using the IVA-L-SOS algorithm. We tested the performance of the pt-cIVA approach using different values of γ_l between 1 and 1000, and observed no change in the performance. In this work, we set $\gamma_l = 3$ following [21]. For cIVA with a fixed constraint parameter, we vary ρ from 0.001 to 0.9, where $\rho = 0.9$ corresponds to stronger influence of the constraint and $\rho = 0.001$ corresponds to weaker influence.

We measure the performance of the methods in terms of jISI, and dissimilarity between the constrained estimated source and ground truth. The average of the jISI metric computed over 50 runs for each method is shown in Figure 4. For each scenario in Figure 4, jISI obtained using regular cIVA increases when the constraint parameter is fixed to a value above the true parameter value indicating poor separation of the sources. On the other hand, pt-cIVA demonstrates lower jISI for all three scenarios indicating good separation performance. The constraint parameter selected at each IVA iteration for the three scenarios for all 50 runs is shown in Figure 5. We can see that the parameter converges to the true value (indicated by ‘*’) for all the scenarios. The true value is computed by plugging in the true constrained sources, $s_l^{[m]}$, into the equation in line 9 of Algorithm 1. Hence, it is lower than ρ_{true} for scenarios A and B. In order to verify if the proposed method accurately estimates the constrained source across time windows, we measure the dissimilarity factor, α , between the constrained estimated source, $\hat{s}_l^{[m]}$, and corresponding ground truth, $s_l^{[m]}$, computed as, $\alpha = 1 - 1/M \sum_{m=1}^M |\text{corr}(\hat{s}_l^{[m]}, s_l^{[m]})|$. A higher value of this metric indicates poor estimation of the sources. Figure 6 shows the dissimilarity factor obtained using regular cIVA and pt-cIVA for the three scenarios. The estimation of the constrained component degrades using regular cIVA when a higher constraint parameter is used whereas the proposed method has low dissimilarity factor for scenarios A and B. Noting the lower jISI and dissimilarity factor metric for lower

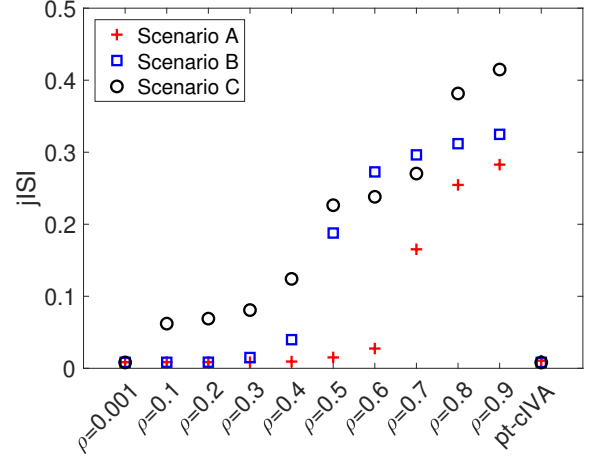


Fig. 4. Performance of cIVA with fixed constraint parameter varied from $\rho = 0.001, \dots, 0.9$ and pt-cIVA with $\mathcal{P} \in \{0.001, \dots, 0.9\}$ in terms of jISI for the three scenarios. The performance of cIVA with fixed constraint parameter degrades if the parameter is fixed to value higher than the true constraint parameter whereas pt-cIVA has low jISI for all three scenarios.

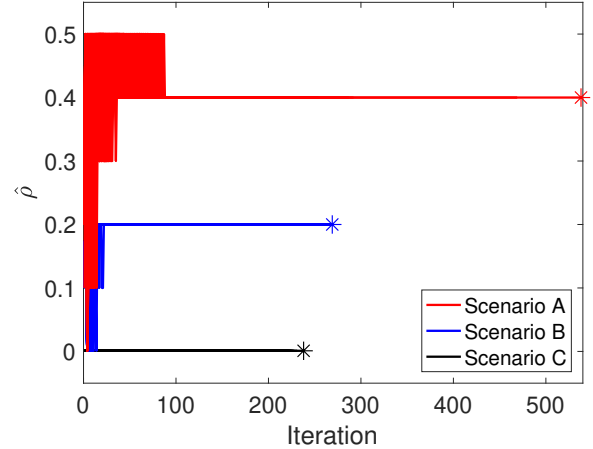


Fig. 5. Constraint parameter selected at every IVA iteration for all 50 runs for scenarios A, B and C. The marker ‘*’ indicates the corresponding true value of ρ . The estimated constraint parameter using pt-cIVA converges to the true value for all scenarios.

values of ρ in regular cIVA, it might be initially thought that cIVA might be preferable rather than pt-cIVA. However, since in real world applications, one does not know the true value of the constraint parameter and whether the constraint is present or not, setting a lower value for ρ might adversely affect the performance of the estimation. For scenario C, i.e., when the constraint is not present, pt-cIVA demonstrates better performance than regular cIVA for lower values of ρ . At $\rho = 0.001$, which is equivalent to performing unconstrained IVA, the jISI value is similar to that of pt-cIVA, however the dissimilarity factor is high for all scenarios, indicating a weaker influence of the constraints on the source. For scenario C the estimated constraint parameter, $\hat{\rho}_n$, is 0.001, imposing a weaker constraint on the IVA decomposition. This is equivalent to performing regular IVA that holds permutation ambiguity. Thus the dissimilarity factor between the estimated source and constraint source is high even though the jISI value is low.

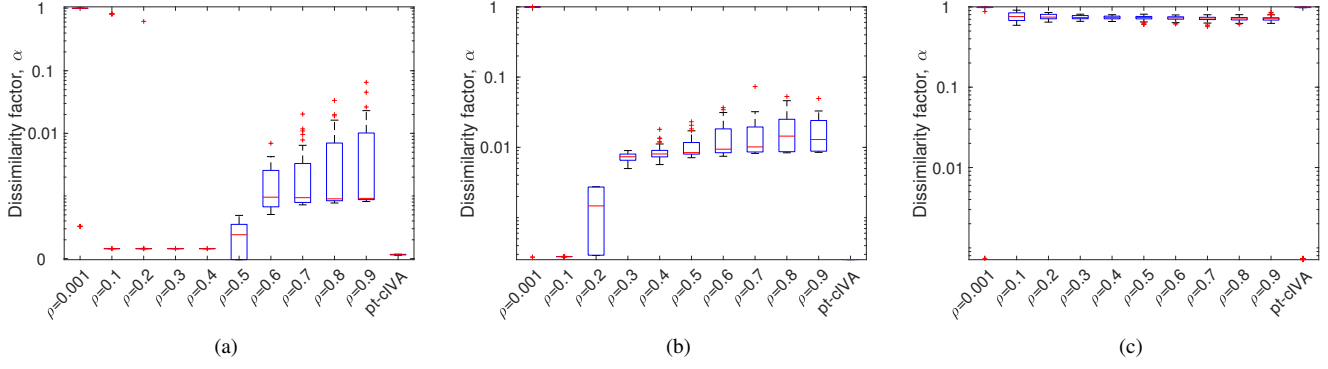


Fig. 6. Performance of cIVA with fixed constraint parameter varied from $\rho = 0.001, \dots, 0.9$ and pt-cIVA with $\mathcal{P} \in \{0.001, \dots, 0.9\}$ in terms of dissimilarity factor for (a) Scenario A, (b) Scenario B and (c) Scenario C. For each box, the horizontal red line indicates the median, the top and bottom edges indicate the 75th and 25th percentiles, respectively, the whiskers show the extreme points not considered as outliers and the '+' symbol indicate outliers. The dissimilarity factor of the constrained component is low for pt-cIVA whereas it increases using regular cIVA when a higher constraint parameter is used.

V. IDENTIFICATION OF RESTING-STATE DYNAMICS

We use a large-scale resting-state fMRI data obtained from the Center for Biomedical Research Excellence (COBRE), which is available on the collaborative informatics and neuroimaging suite data exchange repository (<http://coins.mrn.org/dx>) [31], to capture the variability using the proposed pipeline. This resting-state fMRI data includes $K = 179$ subjects: 91 healthy controls (HCs) (average age: 38 ± 12) and 88 patients with schizophrenia (SZs) (average age: 37 ± 14). For this study, the participants were asked to keep their eyes open during the entire scanning period. The resting fMRI data were obtained using a 3-Tesla TIM Trio Siemens scanner with TE = 29 ms, TR = 2 s, flip angle = 75° , slice thickness = 3.5 mm, voxel size = $3.75 \times 3.75 \times 4.55$ mm³ and slice gap = 1.05 mm. Image scans were obtained over five minutes with a sampling period of 2 seconds yielding 150 timepoints per subject. We removed the first 6 timepoints to address T1-effect and each subject's image data was pre-processed including motion correction, slice-time correction, spatial normalization and slightly re-sampled to $3 \times 3 \times 3$ mm³ yielding $53 \times 63 \times 46$ voxels. We perform masking on each image volume to remove the non-brain voxels and flatten the result to form an observation vector of $V = 58604$ voxels, giving $T = 144$ time evolving observations for each subject. Each subject's data is normalized to zero mean per time point and whitened.

In order to extract the reference signals using a data-driven approach, we perform GICA using the Group ICA for fMRI (GIFT) toolbox (<http://mialab.mrn.org/software/gift>) on the resting-state fMRI data. The number of group components are estimated using a modified minimum description length criterion that accounts for sample dependence [32], for each subject's data and the mean (30) plus one standard deviation (5) across subjects is used as the final number of estimated group components. ICA using the entropy rate bound minimization algorithm [33], [34] is used to estimate 35 components/networks. Of these, $N = 17$ functionally relevant networks are selected based on visual inspection. These networks are categorized into 8 domains: auditory, sensorimotor (SM), frontal, fronto-parietal (FP), parietal (PAR), default mode network (DMN), visual (VIS) and cerebellum. The

DMN domain also consists of voxels corresponding to anterior DMN (ADMN) and insular (INS) regions. The frontal, parietal and fronto-parietal networks comprise the cognitive control domain. The FRO domain consists of two networks: FRO1 and FRO2 corresponding to their peak activation in the frontal cortex situated anterior to the premotor cortex and dorsolateral prefrontal cortex, respectively. The PAR domain consists of three networks: PAR1, PAR2 and PAR3, corresponding to their peak activation in the primary somatosensory cortex, supramarginal gyrus and somatosensory association cortex, respectively. The VIS domain consists of two networks: VIS1 and VIS2, corresponding to their peak activation in the lateral and medial visual cortex, respectively. The components in each domain and the corresponding number of components are shown in Figure. 7.

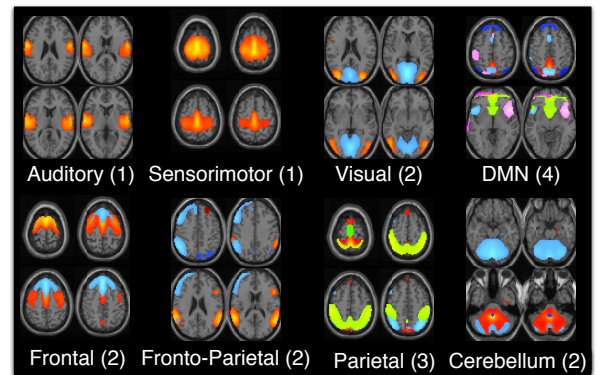


Fig. 7. 17 features are selected from GICA as constraints for pt-cIVA. The features are categorized into 8 domains: auditory (AUD), sensorimotor (SM), frontal (FRO), fronto-parietal (FP), parietal (PAR), default mode network (DMN), visual (VIS) and cerebellum (CB).

Each of these $N = 17$ components is used as a reference signal in the pt-cIVA model in order to capture their variation in both the spatial and temporal domain. For the pt-cIVA model, we divide each subject's data into $M = 17$ windows of length $L = 16$ with a 50% overlap, resulting in a total of $MK = 3043$ windows. By performing pt-cIVA on each subject's data, we reduce the dimensionality of the SCV from 3043 to 17. The first SCV is constrained to be correlated with

one of the 17 group components. The pt-cIVA method using IVA-L-SOS algorithm, with the set \mathcal{P} defined as 0.001, \dots , 0.9 and $\gamma_n = 3$, is applied on the windowed datasets of each subject to estimate 10 solutions. Since IVA is an iterative algorithm, optimization of IVA results in different solutions depending on the initialization. Hence, in order to select the most representative run, we perform the method in [35] to select the most consistent run across multiple runs with different initializations. This method computes the distance between solutions obtained for each pair of runs and selects the runs that has the least average distance. Along with addressing the issue of high dimensionality by limiting the size of the solution space, the use of references in IVA also results in components that are ordered across multiple subject-level IVA decompositions, thus yielding SCVs that are aligned across subjects. In order to verify that the estimated constrained SCVs are ordered across subjects, we visually inspected the estimated components and observed that these components were similar to the reference signal. We also inspected the final constraint parameter for all reference signals and for all subjects and the range of these values was between 0.4 to 0.9, indicating that the components are ordered as per the reference signal. The estimated source corresponding to n th constraint for the k th subject at m th window from the consistent run is denoted as $\mathbf{y}_n^{[m,k]}$. The corresponding time courses at each window are further processed to correct for quadratic, linear and cubic trends, and low-pass filtered with a cutoff of 0.15Hz [6]. We obtain M graphs for each subject, $R^{[m,k]}, k = 1, \dots, K, m = 1, \dots, M$, using N nodes and $N(N-1)$ edges, denoted as $r_{n_1 n_2}^{[m,k]}$. The N nodes represent spatial maps or time courses and an edge defines the Pearson's correlation coefficient between the n_1 th and n_2 th nodes, $n_1, n_2 = 1, \dots, N$. Thus, we obtain M temporal dFC (tdFC) and spatial dFC (sdFC) graphs of dimension $N \times N$ from time courses and spatial nodes respectively for each subject.

QUANTIFICATION OF DYNAMICS

A number of studies have focused on identifying biomarkers that show differences in the HC and SZ groups [36], [37], [38]. We were interested in determining if any of the estimated spatio-temporal component features would be sensitive to mental illness. One feature that is of interest is variability of functional connectivity and spatial maps [39], [20]. In this paper, we define two metrics: component similarity and functional connectivity fluctuation, to identify the spatial components and functional connections that are variable. To evaluate differences between the HC and SZ groups, we perform a permutation test on each metric, that is, a non-parametric statistical test which controls the false alarm rate under the null hypothesis [40], [41]. The idea of a permutation test is to determine whether the difference between the two groups is large enough to reject the null hypothesis that two groups have identical distributions. The test first obtains the observed difference between the two groups using the true labels of the subjects. The labels for the subjects from the two groups are randomly pooled and a difference statistic using the new labels is obtained for every permutation of

the labels. A distribution of the calculated differences is the exact distribution of possible differences under the null hypothesis. If the observed difference is within 95% of the exact distribution, then we do not reject the null hypothesis. This test hence assumes that there are no differences between the two groups and tests if this hypothesis is true or not. We use the t -statistic obtained from a two-sample t -test to measure the difference between the two groups and identify whether a particular group has higher intensity using the sign of the t -statistic.

A. Functional connectivity fluctuation

The functional connectivity fluctuation, $\sigma_{n_1 n_2}^{[k]}$, for each subject using tdFC and sdFC graphs is computed as follows,

$$\sigma_{n_1 n_2}^{[k]} = \sqrt{\frac{1}{M-1} \sum_{m=1}^M \left(r_{n_1 n_2}^{[m,k]} - \bar{c}_{n_1 n_2}^{[k]} \right)^2}, \quad (9)$$

where $\bar{c}_{n_1 n_2}^{[k]} = \frac{1}{M} \sum_{m=1}^M r_{n_1 n_2}^{[m,k]}$ is the mean of the connectivity metric, $r_{n_1 n_2}^{[m,k]}$, computed across M windows for nodes n_1 and n_2 , and $r_{n_1 n_2}^{[m,k]}$ denotes the Pearson's correlation coefficient between the nodes n_1 and n_2 . Each node represents a spatial map/time course obtained using pt-cIVA. We also compute this metric on the tdFC graphs obtained from GICA, which estimates time courses while assuming the spatial networks are stationary. The estimated reference signals are back-reconstructed to estimate subject specific time courses, and a sliding window of length $L = 16$ is applied with a 50% overlap yielding $M = 17$ windows. For each subject, tdFC graphs are obtained as mentioned above. The mean and standard deviation of functional connectivity fluctuation metric across the HC and SZ group are obtained for tdFC graphs from GICA and proposed approach, and sdFC graphs, and are provided as supplementary material ¹.

The permutation test results on the functional connectivity fluctuation metric identified a number of distinct and relevant connections. These also showed lower p -values using the proposed method as compared with GICA. Figure. 8 shows the connections identified as significantly different using tdFC: GICA (Figure. 8(a)), tdFC: pt-cIVA (Figure. 8(b)) and sdFC: pt-cIVA (Figure. 8(c)). The combined result using tdFC and sdFC graphs computed from our method suggests lower variability within the cognitive control network and within the default mode network for the SZ group. Studies have reported decreased hemodynamic response in the insula region in the SZ group causing low variability in this region [42]. Higher variability is observed across components in different clusters, namely, the visual and cognitive control cluster, visual and DMN cluster, visual and frontal component, and fronto-parietal and sensorimotor component for the SZ group. This variability across brain regions may be due to dysfunction in the working memory, attention and visual learning [43] and the tendency of patients with schizophrenia to engage more brain regions than healthy controls [44]. However, GICA results identifies higher temporal variability in HC group between

¹Supplementary materials are available in the supporting documents /multimedia tab.

DMN and the frontal component of the cognitive network and higher variability in the SZ group between visual and auditory component, and between auditory and sensorimotor component. These results suggest that the use of tdFC graphs alone does not fully characterize the dynamic functional connectivity and assuming variability in both spatial and temporal domains results in identification of more distinct biomarkers.

B. Component similarity

In order to quantify the variability of the n th feature for each subject k , we compute the absolute value of the Pearson's correlation coefficient between the n th component at window m , $\mathbf{y}_n^{[m,k]}$, and the n th component at window $m+1$, $\mathbf{y}_n^{[m+1,k]}$. Component similarity is then obtained by computing the mean across all adjacent windows. A higher value of this metric suggests that the spatial network is less variable. Figure. 8(e) shows the results for the components that demonstrated significant differences ($p < 0.05$, corrected) using a permutation test between healthy controls and schizophrenia patients. The VIS, SM, FP and PAR1 components exhibited less variability within the HC group whereas the PAR2 component exhibited less variability in schizophrenia. These components were also identified as less variable among healthy individuals in a previous dynamic study [20]. Deficits in visual perception, attention and motor regions have been previously shown in schizophrenia, which may lead to variability in these components. Figure. 8(f) shows an example of the changes in the visual component of one subject for whom the stationarity is estimated as the highest within the HC and SZ groups. The activated voxels corresponding to the visual component also shows disrupted activation patterns across time for the SZ subject. This is consistent with previous work showing disruptions in the perceptual functions in SZ subjects including abnormalities of smooth pursuit in this group of subjects [45].

VI. DISCUSSION

In recent years, extracting time-varying spatial and temporal features has become of interest in neuroimaging studies. IVA provides a simple linear formulation with minimum assumptions and allows for estimation of spatio-temporal features. However, due to the effect of high dimensionality in IVA, it has been applied to a small number of subjects. We develop a technique that reduces the effect of high dimensionality and extracts time-varying features in both spatial and temporal domain for a large number of subjects. This method extracts reference signals using GICA followed by pt-cIVA to extract their variability across time windows through tuning the constraint parameter. The tuning method effectively captures the variability of the components across time windows using simulated data as compared with using regular cIVA. It also identifies functional connections that differentiate healthy controls and patients with schizophrenia from a large-scale fMRI data.

The two-stage procedure allows for flexibility in the use of different methods for extraction of reference signals. Methods such as dictionary learning [46] and sparse ICA [47] can be used to exploit sparsity of the components. Matrix decomposition techniques such as multiset canonical correlation analysis (MCCA) [48], population value decomposition [49],

shared dictionary learning [50], joint and individual variation explained (JIVE) [51] and common orthogonal basis extraction (COBE) [52] can be used to extract common and individual features from subjects and used as reference signals for the second stage.

The success of the proposed method leads to a number of future directions. Identification of spatial connectivity states and studying potential gains from analyzing them is of interest. Studying the ability of the spatial features to classify subjects can provide a quantification to study the effectiveness of spatial features. Graph theoretical metrics such as dynamic connectivity strength, clustering coefficient and centrality measures can be used to summarize the dynamic spatial networks [53]. Along with capturing variability of the features across time windows, the pt-cIVA technique can be used for applications where the effect of prior information is unknown. In seed-based analysis, selection of 'seeds' is crucial to the analysis of fMRI data, *i.e.*, incorrect selection of 'seeds' may lead to incorrect detection of connectivity. In analysis of EEG data, prior information regarding target frequencies is imposed on all subjects without any knowledge about the presence of these frequencies for each subject. The use of pt-cIVA model would automatically weigh the influence of incorrect reference selection.

REFERENCES

- [1] R. M. Hutchison, T. Womelsdorf, J. S. Gati, S. Everling, and R. S. Menon, "Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques," *Human Brain Mapping*, vol. 34, no. 9, pp. 2154–2177, 2013.
- [2] Ü. Sakoğlu, G. D. Pearson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun, "A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, no. 5-6, pp. 351–366, 2010.
- [3] A. Kucyi and K. D. Davis, "Dynamic functional connectivity of the default mode network tracks daydreaming," *NeuroImage*, vol. 100, pp. 471–480, 2014.
- [4] L. Wang, C. Yu, H. Chen, W. Qin, Y. He, F. Fan, Y. Zhang, M. Wang, K. Li, Y. Zang *et al.*, "Dynamic functional reorganization of the motor execution network after stroke," *Brain*, vol. 133, no. 4, pp. 1224–1238, 2010.
- [5] V. D. Calhoun, R. Miller, G. Pearson, and T. Adali, "The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery," *Neuron*, vol. 84, no. 2, pp. 262–274, 2014.
- [6] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.
- [7] E. Damaraju, E. A. Allen, A. Belger, J. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearson, S. Potkin, A. Preda *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.
- [8] B. Rashid, E. Damaraju, G. D. Pearson, and V. D. Calhoun, "Dynamic connectivity states estimated from resting fMRI identify differences among Schizophrenia, bipolar disorder, and healthy control subjects," *Frontiers in Human Neuroscience*, vol. 8, p. 897, 2014.
- [9] M. Assaf, K. Jagannathan, V. D. Calhoun, L. Miller, M. C. Stevens, R. Sahl, J. G. O'Boyle, R. T. Schultz, and G. D. Pearson, "Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients," *NeuroImage*, vol. 53, no. 1, pp. 247–256, 2010.
- [10] Z. Fu, Y. Tu, X. Di, Y. Du, J. Sui, B. B. Biswal, Z. Zhang, N. de Lacy, and V. D. Calhoun, "Transient increased thalamic-sensory connectivity and decreased whole-brain dynamism in autism," *NeuroImage*, 2018.
- [11] R. L. Bluhm, P. C. Williamson, E. A. Osuch, P. A. Frewen, T. K. Stevens, K. Boksman, R. W. Neufeld, J. Théberge, and R. A. Lanius, "Alterations in default network connectivity in post-traumatic stress disorder related to early-life trauma," *Journal of Psychiatry & Neuroscience: JPN*, vol. 34, no. 3, p. 187, 2009.

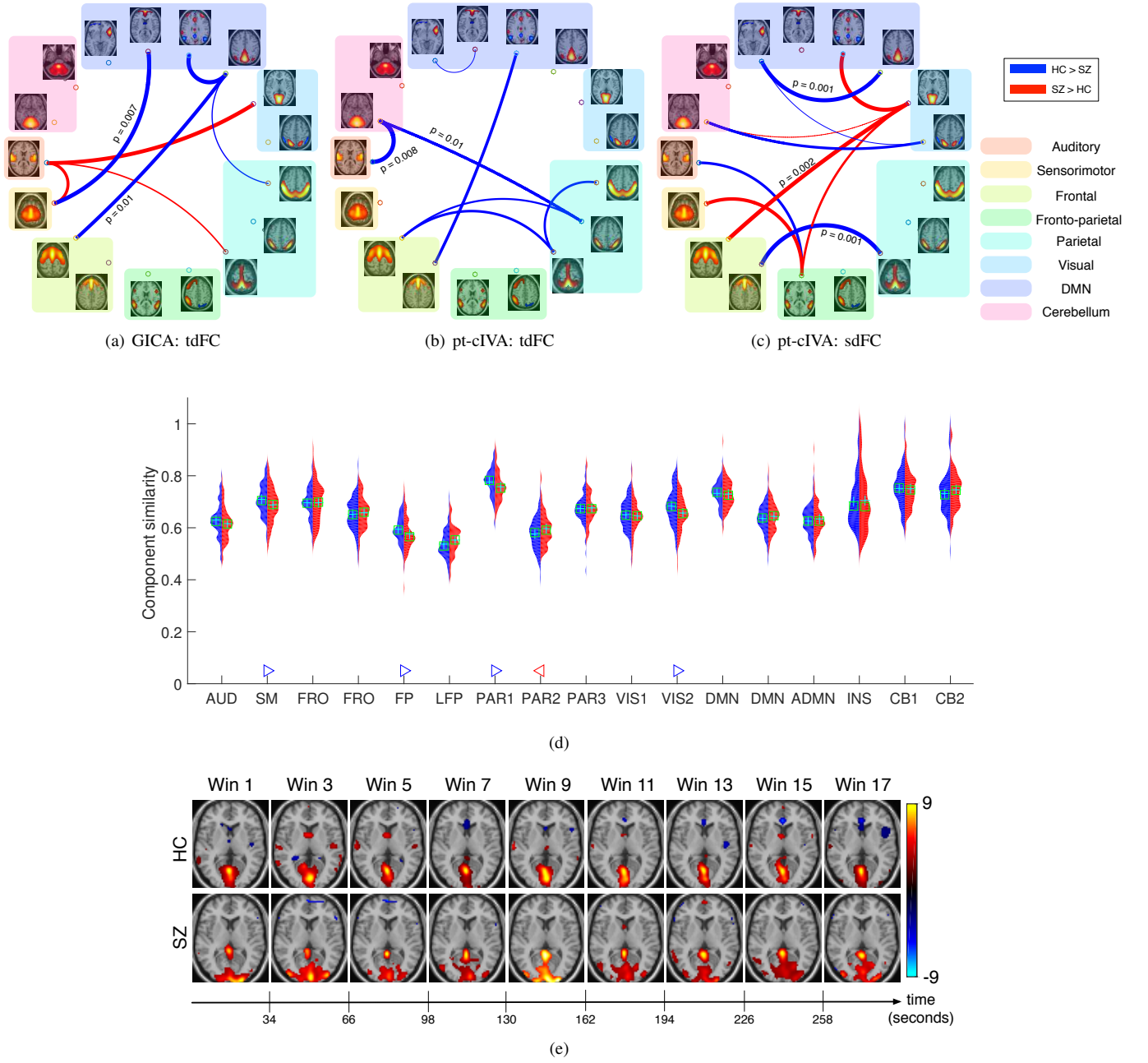


Fig. 8. (a-c) Associations that demonstrate significant difference ($p < 0.05$, corrected) between HC and SZ group. Blue connections indicate higher measures in controls whereas red indicates higher measure in patients. Thickness of the connection indicates a more significant difference (lower p -value) between HCs and SZs. More group differentiating and relevant connections with significantly lower p -values are obtained using the proposed method as compared with GICA. (d) Component similarity of all components. Red indicates the distribution of this metric for the SZ group and blue indicates the distribution of this metric across HC group. Components that demonstrate significant difference ($p < 0.05$, corrected) between HC and SZ group are indicated by a triangle. A blue ' $>$ ' denotes the corresponding component is less variable in the HC group whereas a red ' $<$ ' denotes the corresponding component is less variable in the SZ group. The results indicate that SM, FP, PAR1, and VIS components are less variable in HC whereas PAR2 is less variable in SZ group. (e) Changes in the visual component (VIS1) for one subject corresponding to lowest variability within the HC group and SZ group. The visual component shows disrupted activation pattern for the SZ subject as compared with the HC subject.

- [12] A. Etkin, K. E. Prater, A. F. Schatzberg, V. Menon, and M. D. Greicius, "Disrupted amygdalar subregion functional connectivity and evidence of a compensatory network in generalized anxiety disorder," *Archives of General Psychiatry*, vol. 66, no. 12, pp. 1361–1372, 2009.
- [13] N. de Lacy and V. D. Calhoun, "Dynamic connectivity and the effects of maturation in youth with attention deficit hyperactivity disorder," *Network Neuroscience*, no. Just Accepted, pp. 1–41, 2018.
- [14] X. Chen, H. Zhang, Y. Gao, C.-Y. Wee, G. Li, D. Shen, and A. D. N. Initiative, "High-order resting-state functional connectivity network for MCI classification," *Human Brain Mapping*, vol. 37, no. 9, pp. 3282–3296, 2016.
- [15] L. Q. Uddin, K. S. Supekar, S. Ryali, and V. Menon, "Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development," *Journal of Neuroscience*, vol. 31, no. 50, pp. 18 578–18 589, 2011.
- [16] S. Lefebvre, M. Demeulemeester, A. Leroy, C. Delmaire, R. Lopes, D. Pins, P. Thomas, and R. Jardri, "Network dynamics during the different stages of hallucinations in schizophrenia," *Human Brain Mapping*, vol. 37, no. 7, pp. 2571–2586, 2016.
- [17] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Medical Image Analysis*, vol. 47, pp. 81–94, 2018.
- [18] A. Kottaram, L. Johnston, E. Ganella, C. Pantelis, R. Kotagiri, and A. Zalesky, "Spatio-temporal dynamics of resting-state brain networks improve single-subject prediction of schizophrenia diagnosis," *Human Brain Mapping*, 2018.
- [19] J. M. Kunert-Graf, K. M. Eschenburg, D. J. Galas, J. N. Kutz, S. D. Rane, and B. W. Brunton, "Extracting reproducible time-resolved resting state networks using dynamic mode decomposition," *bioRxiv*, p. 343061, 2018.
- [20] S. Ma, V. D. Calhoun, R. Phlypo, and T. Adali, "Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis," *NeuroImage*, vol. 90, pp. 196–206, 2014.
- [21] S. Bhinge, Q. Long, Y. Levin-Schwartz, Z. Boukouvelas, V. D. Calhoun, and T. Adali, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 2666–2670.
- [22] T. Adali, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.
- [23] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.
- [24] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2006, pp. 165–172.
- [25] V. D. Calhoun, T. Adali, G. D. Pearson, and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, pp. 140–151, 2001.
- [26] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 60–73, 2012.
- [27] E. Gómez, M. Gomez-Vilegas, and J. Marin, "A multivariate generalization of the power exponential family of distributions," *Communications in Statistics-Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.
- [28] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1803–1814, 2007.
- [29] A. H. Khan, M. Taseska, and E. A. Habets, "A geometrically constrained independent vector analysis algorithm for online source extraction," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 396–403.
- [30] V. Calhoun, T. Adali, G. Pearson, and J. Pekar, "Group ICA of functional MRI data: Separability, stationarity, and inference," in *Proc. Int. Conf. on ICA and BSS San Diego, CA*, p. vol. 155, 2001.
- [31] A. Scott, W. Courtney, D. Wood, R. De la Garza, S. Lane, R. Wang, M. King, J. Roberts, J. A. Turner, and V. D. Calhoun, "COINS: An innovative informatics and neuroimaging tool suite built for large heterogeneous datasets," *Frontiers in Neuroinformatics*, vol. 5, p. 33, 2011.
- [32] Y.-O. Li, T. Adali, and V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human Brain Mapping*, vol. 28, no. 11, pp. 1251–1266, 2007.
- [33] X.-L. Li and T. Adali, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 1934–1937.
- [34] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li *et al.*, "Blind source separation by entropy rate minimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4245–4255, 2014.
- [35] Q. Long, C. Jia, Z. Boukouvelas, B. Gabrielson, D. Emge, and T. Adali, "Consistent run selection for independent component analysis: Application to fMRI analysis," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2581–2585.
- [36] Y. Du, G. D. Pearson, D. Lin, J. Sui, J. Chen, M. Salman, C. A. Tamminga, E. I. Ivleva, J. A. Sweeney, M. S. Keshavan *et al.*, "Identifying dynamic functional connectivity biomarkers using GIG-ICA: Application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder," *Human Brain Mapping*, vol. 38, no. 5, pp. 2683–2708, 2017.
- [37] Y. Du, Z. Fu, and V. D. Calhoun, "Classification and prediction of brain disorders using functional connectivity: Promising but challenging," *Frontiers in Neuroscience*, vol. 12, 2018.
- [38] M. S. Salman, Y. Du, D. Lin, Z. Fu, E. Damaraju, J. Sui, J. Chen, Q. Yu, A. R. Mayer, S. Posse *et al.*, "Group ica for identifying biomarkers in schizophrenia: 'adaptive' networks via spatially constrained ica show more sensitivity to group differences than spatio-temporal regression," *bioRxiv*, p. 429837, 2018.
- [39] A. Zalesky, A. Fornito, L. Cocchi, L. L. Gollo, and M. Breakspear, "Time-resolved resting-state brain networks," *Proceedings of the National Academy of Sciences*, p. 201400181, 2014.
- [40] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," *NeuroImage*, vol. 53, no. 4, pp. 1197–1207, 2010.
- [41] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.
- [42] K. P. Wylie and J. R. Tregellas, "The role of the insula in schizophrenia," *Schizophrenia research*, vol. 123, no. 2-3, pp. 93–104, 2010.
- [43] R. W. Heinrichs and K. K. Zakzanis, "Neurocognitive deficit in schizophrenia: a quantitative review of the evidence," *Neuropsychology*, vol. 12, no. 3, p. 426, 1998.
- [44] S. Ma, T. Eichele, N. M. Correa, V. D. Calhoun, and T. Adali, "Hierarchical and graphical analysis of fMRI network connectivity in healthy and schizophrenic groups," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 1031–1034.
- [45] L. Abel, S. Levin, and P. Holzman, "Abnormalities of smooth pursuit and saccadic control in schizophrenia and affective disorders," *Vision Research*, vol. 32, no. 6, pp. 1009–1014, 1992.
- [46] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 562–573.
- [47] Z. Boukouvelas, Y. Levin-Schwartz, and T. Adali, "Enhancing ICA performance by exploiting sparsity: Application to FMRI analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2532–2536.
- [48] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.
- [49] C. M. Crainiceanu, B. S. Caffo, S. Luo, V. M. Zipunnikov, and N. M. Punjabi, "Population value decomposition, a framework for the analysis of image populations," *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 775–790, 2011.
- [50] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5160–5175, 2017.
- [51] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (JIVE) for integrated analysis of multiple data types," *The Annals of Applied Statistics*, vol. 7, no. 1, p. 523, 2013.
- [52] G. Zhou, A. Cichocki, Y. Zhang, and D. P. Mandic, "Group component analysis for multiblock data: Common and individual feature extraction," *IEEE Transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2426–2439, 2016.
- [53] A. E. Sizemore and D. S. Bassett, "Dynamic graph metrics: Tutorial, toolbox, and tale," *NeuroImage*, 2017.