The Role of Diversity in Data-driven Analysis of Multi-subject fMRI Data: Comparison of Approaches Based on Independence and Sparsity Using Global Performance Metrics

Qunfang Long[a,*], Suchita Bhinge[a], Yuri Levin-Schwartz[b], Zois Boukouvalas[c], Vince D. Calhoun[d,e], and Tülay Adalı[a]

[a]*Department of CSEE, University of Maryland Baltimore County, Baltimore, MD 21250, USA*

[b]*Department of EMPH, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA*

[c]*Department of ENME, University of Maryland College Park, College Park, MD 20742, USA*

[d]*The Mind Research Network, Albuquerque, NM 87106, USA*

[e]*Department of ECE, University of New Mexico, Albuquerque, NM 87131, USA*

*Corresponding author:

Qunfang Long
E-mail: qunfang1@umbc.edu
Tel: (+1)240-535-7959
This work was performed at University of Maryland Baltimore County

How to cite this article:

Long Q, Bhinge S, Levin-Schwartz Y, Boukouvalas Z, Calhoun VD, Adalı T. The role of diversity in data-driven analysis of multi-subject fMRI data: Comparison of approaches based on independence and sparsity using global performance metrics. Hum Brain Mapp. 2018;1–16. https://doi.org/10.1002/hbm.24389

Abstract

Data-driven methods have been widely used in functional magnetic resonance imaging (fMRI) data analysis. They extract latent factors, generally, through the use of a simple generative model. Independent component analysis (ICA) and dictionary learning (DL) are two popular data-driven methods that are based on two different forms of diversity—statistical properties of the data—statistical independence for ICA and sparsity for DL. Despite their popularity, the comparative advantage of emphasizing one property over another in the decomposition of fMRI data is not well understood. Such a comparison is made harder due to the differences in the modeling assumptions between ICA and DL, as well as within different ICA algorithms where each algorithm exploits a different form of diversity. In this paper, we propose the use of objective global measures, such as time course frequency power ratio, network connection summary and graph theoretical metrics, to gain insight into the role that different types of diversity have on the analysis of fMRI data. Four ICA algorithms that account for different types of diversity and one DL algorithm are studied. We apply these algorithms to real fMRI data collected from patients with schizophrenia and healthy controls. Our results suggest that no one particular method has the best performance using all metrics, implying that the optimal method will change depending on the goal of the analysis. However, we note that in none of the scenarios we test the highly popular Infomax provides the best performance, demonstrating the cost of exploiting limited form of diversity.

***Keywords**:* Data-driven analysis, ICA, dictionary learning, fMRI analysis, diversity, independence, sparsity, performance evaluation, global metric

1. Introduction

In blind source separation problems, such as functional magnetic resonance imaging (fMRI) data analysis, it is beneficial to summarize the observed data through a latent factor model. However, little is known about the processes underlying the generation of the factors. This motivates the development of data-driven methods, which extract latent factors through the use of a simple generative model. Data-driven methods have proven useful for the analysis of fMRI data since the underlying assumptions are minimized [Li et al., 2009a; Li et al., 2009b; Van Den Heuvel and Pol, 2010]. These assumptions are, in part, based on different forms of diversity—types of statistical properties—of either the observations or the underlying sources, such as independence and sparsity, thus resulting in the development of different data-driven methods.

One such technique that has proven quite popular is independent component analysis (ICA), the goal of which is to maximize a measure of (statistical) independence among the underlying sources [Calhoun et al., 2009; Li et al., 2009a; Van Den Heuvel and Pol, 2010]. This popularity has, in part, led to the development of different ICA algorithms that can be derived from the maximum likelihood principle, each designed to achieve the independent component decomposition through exploiting different forms of diversity of the signals, such as higher-order statistics, noncircularity, and sample dependence [Adalı et al., 2014; Calhoun et al., 2001; Du et al., 2016; Hyvärinen et al., 2001]. For example, the Infomax algorithm [Bell and Sejnowski, 1995], which was the first algorithm used for fMRI analysis [Mckeown et al., 1998], exploits higher-order statistics through the use of a fixed tangent hyperbolic nonlinearity. In contrast to Infomax, entropy bound minimization (EBM) [Li and Adalı, 2010b] and entropy rate bound minimization (ERBM) [Li and Adalı, 2010a] are two more recently introduced ICA algorithms and have been shown to provide desirable performance on both simulated and real fMRI data, see *e.g.*, [Adalı et al., 2014; Du et al., 2011; Li and Adalı, 2010a; Long et al., 2017], through the use of a dynamic nonlinearity, and in the case of ERBM, sample dependence of each source as well. The use of a dynamic nonlinearity enables EBM and ERBM to match a wide variety of distributions.

Besides independence, sparsity—another form of diversity that has been exploited in many fields—has proven to be a useful property in the decomposition of fMRI signals [Abolghasemi et al., 2015; Lee et al., 2011]. Dictionary learning (DL) makes use of sparsity as the starting point for the decomposition and has been shown to be useful for the analysis of fMRI data as well [Abraham et al., 2013; Lee et al., 2011; Varoquaux et al., 2011]. Most ICA algorithms can be cast under the maximum likelihood framework and their performance is a function of the assumed source distribution, or the nonlinearity that is used which implies a certain type of distribution. On the other hand, DL aims at balancing the decomposition accuracy and source sparsity through a regularization parameter and in the estimation of functional networks of interest, there is no implicit or explicit assumption of a source distribution.

While initially sparsity has been presented as a competitor for ICA in fMRI analysis [Daubechies et al., 2009], the claims in this article regarding the roles of sparsity and independence are later clarified [Calhoun et al., 2013], and both sparsity and independence have now been recognized as meaningful starting points for use in fMRI analysis. Reference [Boukouvalas et al., 2017a, 2017b] introduces a unified mathematical framework that enables dynamic exploitation of both independence and sparsity, which we refer to as the SparseICA framework. Source sparsity is incorporated through the use of the ICA cost function, penalized by an ℓ^1 regularization term. EBM is utilized to demonstrate its application and the corresponding algorithm is referred to as SparseICA-EBM [Boukouvalas et al., 2017a, 2017b]. It inherits the advantages of EBM with enhanced performance due to the exploitation of sparsity when the sources are

sparse. Though the desirable performance of the algorithm has been shown using simulated fMRI data, it has not been tested on real fMRI data.

Although independence and sparsity have demonstrated their utility for the analysis of fMRI data, there is no study that explores the role of each type of diversity in terms of global metrics for real fMRI datasets. Few studies that have investigated the performance of Infomax, EBM and ERBM [Correa et al., 2007; Du et al., 2011] used a limited number of subjects and based the evaluation on subjective metrics, such as visual inspection of a few well-matched components [Adalı et al., 2014; Du et al., 2011]. Additionally, there has been no comparison of the performance of these methods with the SparseICA framework and DL. Such an exploration raises the issue of how to determine a metric for comparing the component estimation performance of different data-driven algorithms on real fMRI data without a ground truth. Algorithmic comparison is difficult as decompositions can be quite different depending on the modeling assumptions of a particular algorithm, thus matching of all the estimated components one-to-one is usually not possible. In addition, each method is based on a different cost function. This motivates the identification of objective global metrics that are independent of cost function for algorithmic performance.

The purpose of this work is to provide insight into the role of different types of diversity on a decomposition. For this task, we choose Infomax, EBM, ERBM, DL [Mairal et al., 2010] and SparseICA-EBM for the comparison as they are good representations of algorithms that use different types of diversity. Infomax is one of the most widely used data-driven algorithm for the task. EBM, on the other hand, provides a more flexible estimation of source distribution thus enabling better fulfillment of independence among the sources while ERBM extends EBM to account for sample dependence. DL puts the emphasis on sparsity, and SparseICA-EBM balances the roles of the two types of diversity, independence and sparsity. Additionally, the relative performance of these algorithms is assessed on a large real fMRI dataset consisting of 179 subjects. Due to the increasing number of large fMRI datasets that include hundreds or even thousands of subjects, understanding the performance of these techniques in this scenario increases our confidence in the generalizability of the results. Note that each algorithm has its own target cost function thus using the cost for performance assessment is not possible. Even if the algorithms had the same cost function, the lowest value of the cost function does not necessarily lead to better separation results depending on the overall goal. In order to be fair to each algorithm, we propose to use objective global measures, *e.g.*, time course frequency power ratio [Allen et al., 2011; Robinson et al., 2009], network connection summary [Du et al., 2017; Ma et al., 2012] and graph-theoretical metrics [Bullmore and Sporns, 2009; Telesford et al., 2013]. Time course frequency power ratio indicates whether a component is describing the blood-oxygenation-level-dependent (BOLD) response in fMRI data or not. Network connection summary provides a general idea on how well each algorithm can reconstruct the complex connections in brain. Graph-theoretical metrics are an efficient tool for studying the heterogeneity between different groups of subjects, such as patients with schizophrenia (SZs) and healthy controls (HCs) [Alexander-Bloch et al., 2010; Bassett et al., 2012; Lynall et al., 2010; Vértes et al., 2012]. These metrics perform a global comparison of the algorithms based on all the components or the whole brain functional network connectivity.

Through this comparison, we find that the use of global metrics for a performance comparison can provide a general guide to the practitioners about the selection of the appropriate algorithm for a specific situation. For instance, DL produces components comprising signal that is more likely to be derived from the BOLD response, EBM yields better clustering within functional networks, and ERBM is better in capturing group differences and yields higher variance in SZs than HCs when using graph-theoretical

metrics. We discuss such tradeoffs and aim at providing guidance to the practitioner.

The rest of the paper is organized as follows. In Section 2, we introduce the data that is used, describe the group ICA framework as well as the four ICA algorithms, DL, and describe the three global measures in detail. In Section 3, we present the experimental results. We discuss the results in Section 4 and conclude with Section 5.

2. Materials and methods

2.1. Data acquisition

The data used in this study is a resting state fMRI data from the Center of Biomedical Research Excellence (COBRE), which is available on the collaborative informatics and neuroimaging suite data exchange repository (http://coins.mrn.org/dx) Aine et al., 2017; Çetin et al., 2014; Scott et al., 2011]. The data includes 88 SZs (average age: 37 ± 14) and 91 HCs (average age: 38 ± 12). All images were collected on a single 3-Tesla Siemens Trio scanner with a 12-channel radio frequency coil using the following parameters: TE = 29 ms, TR = 2 s, flip angle = 75°, slice thickness = 3.5 mm, slice gap = 1.05 mm, voxel size $3.75 \times 3.75 \times 4.55$ mm^3. Participants were instructed to keep their eyes open during the scan and stare passively at a central fixation cross. Each resting state scan consists of 150 volumes. To eliminate the T1-related signal fluctuations (T1 effect) [Shin et al., 2013], the first 6 volumes are removed in this study, thus 144 volumes remain for each subject. The fMRI data is realigned with INRIalign algorithm [Freire et al., 2002], slice-timing correction is applied using the middle slice as the reference frame in the functional data pipeline and spatially normalized to the standard Montreal Neurological Institute space [Friston et al., 1994] and resampled to $3 \times 3 \times 3$mm^3, resulting in $53 \times 63 \times 46$ voxels. Afterwards, the fMRI data is smoothed using a Gaussian kernel with a full-width at half-maximum of 5mm.

2.2. Group ICA

The group ICA framework enables analysis of fMRI data from multiple subjects using ICA [Calhoun et al., 2001; Calhoun and Adalı, 2012]. Let the observed fMRI data from the kth subject be denoted by $\widetilde{\mathbf{X}}^{[k]} \in \mathbb{R}^{T\times V}$, $1 \le k \le K$, where T denotes the number of time points and V denotes the number of voxels. To reduce the contamination from noise, principal component analysis (PCA), using an order suggested by the entropy rate based order selection technique described in [Fu et al., 2014], is employed to reduce the dimension of the data for each subject, *i.e.*, to estimate the order of signal subspace. The order estimation method proposed in [Fu et al., 2014] takes sample dependence into consideration without downsampling, which leads to improved estimation of the signal subspace. For each subject, the dimension of $\widetilde{\mathbf{X}}^{[k]}$ is reduced from T to T', computed as $\mathbf{X}^{[k]} = \left(\mathbf{F}^{[k]}\right)^{\dagger}\widetilde{\mathbf{X}}^{[k]}$, where † represents the pseudoinverse and $\left(\mathbf{F}^{[k]}\right)^{\dagger} \in \mathbb{R}^{T'\times T}$ is the subject level reduction matrix, whose columns are the eigenvectors of $\widetilde{\mathbf{X}}^{[k]}\left(\widetilde{\mathbf{X}}^{[k]}\right)^{\mathrm{T}}$ and the reduced data is $\mathbf{X}^{[k]} \in \mathbb{R}^{T'\times V}$. The reduced data matrix, $\mathbf{X}^{[k]}$, consists of the first T' principal components of $\widetilde{\mathbf{X}}^{[k]}$ that represent the informative signals from the kth subject. It is assumed that the subjects share a common component subspace [Calhoun and Adalı, 2012]. In order to estimate the order of the common subspace across subjects, the datasets are temporally concatenated to form a single data matrix $\widetilde{\mathbf{Y}} \in \mathbb{R}^{KT'\times V}$, which is then reduced to $\mathbf{Y} \in \mathbb{R}^{N\times V}$ by a group level PCA, $\mathbf{Y} = \mathbf{G}^{\dagger}\widetilde{\mathbf{Y}}$, with $\mathbf{G}^{\dagger} \in \mathbb{R}^{N\times KT'}$ as the group level reduction matrix and N as the order for the common observation subspace. Group components $\mathbf{S} \in \mathbb{R}^{N\times V}$ are then estimated by performing ICA on the common group subspace $\mathbf{Y}$,

$$\mathbf{Y} = \mathbf{AS}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the mixing matrix. ICA seeks to find a group demixing matrix, $\mathbf{W}$, such that the estimated sources are obtained as $\hat{\mathbf{S}} = \mathbf{WY}$. The use of a single ICA on the

common subspace of all datasets helps to preserve the order of the components across subjects. Following the completion of ICA, back-reconstruction is performed on $\hat{\mathbf{S}}$ to generate the corresponding subject-specific source estimates $\hat{\mathbf{S}}^{[k]} \in \mathbb{R}^{N\times V}$. In order to obtain the back-reconstructed signals, $\hat{\mathbf{S}}^{[k]}$, the group level reduction matrix $\mathbf{G}^{\dagger}$ is blocked by columns, $\mathbf{G}^{\dagger} = [(\mathbf{G}^{[1]})^{\dagger}, (\mathbf{G}^{[2]})^{\dagger}, \cdots, (\mathbf{G}^{[K]})^{\dagger}]$ with $(\mathbf{G}^{[k]})^{\dagger} \in \mathbb{R}^{N\times T'}$ and, $\hat{\mathbf{S}}^{[k]}$ is reconstructed using $\hat{\mathbf{S}}^{[k]} = \mathbf{W}(\mathbf{G}^{[k]})^{\dagger}(\mathbf{F}^{[k]})^{\dagger}\tilde{\mathbf{X}}^{[k]}$. The corresponding subject-specific time courses are obtained using $\hat{\mathbf{A}}^{[k]} = \mathbf{F}^{[k]}\mathbf{G}^{[k]}\mathbf{W}^{-1}$.

2.3. Independence-based ICA: Algorithm choice

The differences in separation performance for separate ICA algorithms, such as Infomax, EBM and ERBM, are related to differences in their assumed latent source models. In order to estimate the demixing matrix W, Infomax and EBM equivalently aim at minimizing the mutual information between the source estimates $\hat{\mathbf{s}}_1, \cdots, \hat{\mathbf{s}}_N$, given by

$$I(\mathbf{W}) = \sum_{n=1}^{N} H(\hat{\mathbf{s}}_n) - H(\hat{\mathbf{S}})$$
$$= \sum_{n=1}^{N} H(\hat{\mathbf{s}}_n) - \log|\det \mathbf{W}| - H(\mathbf{Y}) \qquad (2)$$

where $H(\cdot) = -\mathrm{E}[\log p(\cdot)]$ refers to (differential) entropy of a random variable, with p refering to the probability density function (PDF) of corresponding variable. The last term $H(\mathbf{Y})$ is a constant as it is independent of $\mathbf{W}$. Expression in (2) can be written using maximum likelihood (ML) formulation for the given observations. While ICA is achieved through the estimation of the demixing matrix $\mathbf{W}$, for the estimator to achieve the desirable large sample properties of ML principle, the PDF of the sources needs to be modeled or estimated as well [Adalı et al., 2014]. Among different versions of Infomax algorithms [Bell and Sejnowski, 1995, Amari et al., 1996, Lee et al., 1999], the most widely used in fMRI data analysis is the original version proposed in [Bell and Sejnowski, 1995]. In this version, the algorithm takes only higher-order statistics into consideration, and achieves this by using a fixed sigmoidal nonlinearity to model the source PDF [Bell and Sejnowski, 1995], which implies a super-Gaussian PDF as well as a soft sparsity for the sources. This means that Infomax is a good match for very focal regions of activation, however it might significantly bias latent sources relating to broad regions, such as the default mode network (DMN) [Du et al., 2016]. In contrast to Infomax, EBM does not assume one specific distribution for the latent sources but instead attempts to upper bound their entropy through the use of several measuring functions [Li and Adalı, 2010b]. Each of these functions provides bounds on the entropy, with the tightest bound being closest to the true entropy. The use of these measuring functions makes it possible to match a wide variety of distributions, including those that are sub-Gaussian, super-Gaussian, unimodal, bimodal, symmetric, as well as skewed [Li and Adalı, 2010b], thus potentially leading to more accurate estimation of the latent sources.

Instead of bounding the entropy of latent sources, ERBM attempts to bound their entropy rate using measuring functions. The cost function, thus, is given by

$$I_r(\mathbf{W}) = \sum_{n=1}^{N} H_r(\hat{\mathbf{s}}_n) - \log|\det\mathbf{W}| - H_r(\mathbf{Y}) \qquad (3)$$

where $H_r(\hat{\mathbf{s}}_n) = H[(\hat{s}_n(1), \cdots, \hat{s}_n(V))]/V$ refers to the entropy rate of $\hat{\mathbf{s}}_n$, thus the goal is now minimizing the mutual information *rate* among the source estimates. Hence this cost function accounts for two types of diversity—sample dependence and nonstationarity—of the signals. ERBM models the sample dependence of fMRI data and filters the samples using a time-invariant invertible linear filter. Consequently, sample dependence is taken into consideration by ERBM. Since EBM and ERBM relax the assumptions placed on the fMRI sources by assuming flexible source distributions, they are expected to provide improved performance over Infomax. Additionally, ERBM is expected to have superior performance over EBM as well

as Infomax, since it takes advantage of multiple underlying properties of the fMRI components, namely, higher-order statistics and voxel-wise dependence. All three algorithms can be found in the group ICA of fMRI toolbox (GIFT) (http://mialab.mrn.org/software/gift/index.html).

2.4. Sparsity-based DL

DL is one of the data-driven methods that makes use of the sparsity of latent sources as the starting point for the decomposition and aims at balancing the decomposition accuracy and source sparsity through a regularization parameter. Recently, it has been successfully applied to fMRI data analysis [Abolghasemi et al., 2015; Lee et al., 2011]. DL expresses the observations as sparse combinations of the atoms (columns) in a dictionary $\mathbf{D} \in \mathbb{R}^{KT\prime \times N}$, seeking to estimate the latent spatial components $\hat{\mathbf{S}} \in \mathbb{R}^{N \times V}$ that are conveyed in the sparse loadings. The cost of DL procedure is given by

$$\ell(\mathbf{D}) = \min_{\mathbf{D},\hat{\mathbf{S}}} \| \tilde{\mathbf{Y}} - \mathbf{D}\hat{\mathbf{S}} \|_F^2 + \lambda \| \hat{\mathbf{S}} \|_1 \qquad (4)$$

where $\hat{\mathbf{S}}$ are the spatial maps, $\| \hat{\mathbf{S}} \|_1 = \sum_{i=1}^{N} \sum_{j=1}^{V} |\hat{s}_{ij}|$, and λ is the regularization parameter that controls the level of sparsity of the sources. The DL algorithm used in this work achieves the decomposition using an online optimization algorithm, making it suitable for large datasets [Mairal et al., 2010].

2.5. Balancing independence and sparsity

In the case of fMRI data analysis, both independence and sparsity are natural properties. A unified mathematical framework is proposed to account for both by incorporating a sparsity term into the cost function of ICA [Boukouvalas et al., 2017a, 2017b]. The cost function of SparseICA-EBM, which is utilized to demonstrate the application of this framework, is constructed using two terms, *i.e.*, an independence term and a sparsity term. The independence refers to (2) and the sparsity is written as the sum of the regularization function of all estimates. The final cost function is given by

$$\mathcal{J}(\mathbf{W}) = \sum_{n=1}^{N} H(\hat{\mathbf{s}}_n) - \log|\det(\mathbf{W})| + \sum_{n=1}^{N} \lambda_n f(\hat{\mathbf{s}}_n) \qquad (5)$$

$$f(\hat{\mathbf{s}}_n) = \lim_{\varepsilon_n \to 0} \sum_{v=1}^{V} \sqrt{\hat{\mathbf{s}}_{nv}^2 + \varepsilon_n} \qquad (6)$$

where $f(\hat{\mathbf{s}}_n)$ is the regularization term with respect to the nth source estimate $\hat{\mathbf{s}}_n$, λ_n is the sparsity parameter and ε_n is the smoothing parameter. The third term in (2) is removed here since it is a constant with respect to $\mathbf{W}$. By tuning the sparsity and smoothing parameters, SparseICA-EBM enables different decompositions that yield different estimates with different levels of sparsity.

2.6. Parameter selection

One important parameter in the application of an ICA algorithm is the model order, *i.e.*, the order of signal subspace, *N*. However, for fMRI data, classical order estimation techniques based on information theoretic criteria may overestimate the order due to the inherent sample dependence of fMRI data [Li et al., 2007, 2011]. A common way to overcome this issue is by using downsampling to obtain effectively independent and identically distributed samples [Li et al., 2007, 2011]. However, methods based on downsampling suffer from a loss of information associated with it. More recently, two entropy rate (ER)-based order estimation techniques are proposed that account for sample dependence without the use of downsampling: ER using a finite memory length model (ER-FM) and ER using an autoregressive model (ER-AR) [Fu et al., 2014]. Since the sample correlation structure in ER-FM is a better match to that in fMRI data due to the finite span of correlation in the point spread function, ER-FM is used in this paper to estimate the order of signal subspace.

As introduced above, SparseICA-EBM seeks to achieve the decomposition by balancing the roles of independence and sparsity using a sparsity parameter λ_n, which enables the sparse solution for the nth source, $\hat{\mathbf{s}}_n$. The balance

between independence and sparsity can be adjusted by tuning λ_n . Smaller λ_n in (5) emphasizes statistical independence while larger λ_n places greater weight on the sparsity of the sources. Another parameter for SparseICA-EBM is the smoothing parameter ε_n. Since in (5), the regularization term, $f(\hat{\mathbf{s}}_n) = \| \hat{\mathbf{s}}_n \|$, is non-differentiable, it is replaced by the sum of multi-quadratic functions [Lee et al., 2006], as given by (6). A higher smoothing parameter ε_n will produce smoother sources [Boukouvalas et al., 2017b]. However, the effect of this parameter with application to the analysis of real fMRI data has not been explored. For this reason, in this work, different values of λ_n and ε_n are considered and the details of the parameter selection will be presented in Section 3.1.3.

Similarly, an appropriate value should be set for the regularization parameter λ for DL. The performance of DL is highly dependent on the selection of λ to achieve a desirable tradeoff between accuracy and level of sparsity. In this work, we investigate the influence of different values of λ on the performance for our relatively large real fMRI data. Another parameter to be determined is the number of atoms in dictionary, and in this work, the number of atoms is set to be *N*, the order of signal subspace estimated by the data-driven order selection method ER-FM [Fu et al., 2014].

2.7. Global metrics for performance evaluation

We propose the use of global measures to compare the performances of data-driven algorithms on real data. Following are the three global measures that are used.

2.7.1. Frequency analysis

The first global measure is the ratio of time course power spectra in low-frequency band ($<$ 0.1Hz) to the high-frequency band ($>$ 0.15 Hz) for each component. Since the activation in the components is due to the BOLD response, which corresponds to low frequencies, higher power ratios imply that the components are more closely associated with true neural function. Conversely, the lower the ratio, the more likely the component is to be describing the cardiac or respiratory noise as opposed to true BOLD activation [Allen et al., 2011; Cordes et al., 2001].

2.7.2. Network connection summary

Another global measure of the performance is the network connection summary. For each algorithm, M components out of N that refer to brain functional networks are selected based on their time course power ratio and visual inspection. The two-stage procedure helps us to ensure that components with meaningful functional regions are included in the post analyses. If some of them are removed or some noisy components and components with heavy ventricle effects are included in the post analyses, the results may change. First, all the components are separated into two categories, one containing those with time course power ratio higher than a threshold t_{PR}, and the other set containing the remaining ones. Through visual inspection the components with large edge effects and ventricles are removed from the first category, and those whose power ratios are slightly lower than t_{PR} but with activation in meaningful functional network regions are put back to the first category, which is finally used for network connection analysis. Using time course power ratio to separate components into two categories in the first step significantly reduces the difficulty of visual inspection in the second step.

The components extracted from data-driven methods generally have some dependence due to their functional relevance. Consequently, after identifying the functional networks that are conveyed in the components, we estimate brain connectivity by calculating the full-order statistical dependence, mutual information (MI), among them. The functional network connectivity is constructed using normalized MI, $I_{\mathrm{norm}}(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$, as the measure. The normalized MI is given by

$$e(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2) = I_{\mathrm{norm}}(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2) = \frac{2I(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)}{I(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_1) + I(\hat{\mathbf{s}}_2, \hat{\mathbf{s}}_2)} \quad (7)$$

where $I(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$ is the MI between two estimated components $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$. The performance of three Shannon entropy based MI estimation methods, the k-nearest neighborhood [Goria et al., 2005], the analytical formula corresponding in the chosen exponential family [Nielsen and Nock,

2011] and the Parzen window based method [Peng et al., 2005] are investigated. The former two methods are from information theoretical estimators (ITE) toolbox (https://bitbucket.org/szzoli/ite/). Our exploration shows that for observations with super-Gaussian distribution, the method proposed in [Peng et al., 2005] performs the best, and is used in our work. After the normalization, the connectivity between two similar sources becomes close to 1 and that of two dissimilar sources close to 0.

The complete brain network connections are summarized based on the M identified functional networks with the normalized MI as connectivity measure. Connectivity patterns within the meaningful functional modules, such as the default mode cluster, are expected to be observed [Allen et al., 2014; Yu et al., 2015]. Compact connections within modules improve the results of post analyses, such as clustering, on these estimated networks. To measure the modularity of the functional networks, the ratio of the average intra-module connectivity to the average inter-module connectivity is defined as:

$$R = \frac{\frac{1}{N_{\text{intra}}}\sum_{i}^{Q}\sum_{u}^{N_i} e_u}{\frac{1}{N_{\text{inter}}}\sum_{i,j}^{Q}\sum_{v}^{N_{i,j}} e_v} \qquad (8)$$

where Q is the number of modules, N_i is the number of connections within the ith module, $N_{i,j}$ is the number of connections between the ith and jth modules, e_u and e_v refer to the intra- and inter-module connectivity respectively, and N_{intra} and N_{inter} refer to the total number of intra- and inter-module connectivity, again respectively. The larger the ratio, the more compact the modules are.

2.7.3. Graph-theoretical metrics

We also use graph-theoretical metrics as global measures of performance comparison. For each algorithm and subject, the M selected components are used to construct the graph. With the M components of interest as nodes and the normalized MI between them as the edges, a fully connected graph, G, is constructed. Beginning with G, an edge threshold, e_t, is used to retain only the highest P percent of the edges, thus generating a new graph G'. We define the percentage of the edges that remain after thresholding as the link density, which increases as the threshold decreases. The weighted graphs are binarized, with the edges below the threshold e_t having a value of 0 and those above having a value of 1. In order to avoid very sparse graphs with small link densities and those with too high link densities, we limit the link density to range from 20% to 70%.

Graph metrics highlight different topological characteristics of graphs, see *e.g.*, [Boccaletti et al., 2006; Bullmore and Sporns, 2009; Bullmore and Bassett, 2011]. The characteristic path length (PL), global efficiency and centrality are measures that can quantify the ability of a node to facilitate functional integration in graphs [Rubinov and Sporns, 2010]. These are global measures as all the nodes are taken into consideration in their calculation, and provide measures of how information is transferred in the functional network. The clustering coefficient (CC) can capture the segregation of networks by measuring the transfer of information in the immediate neighborhood of each node. We use global graph CC that is averaged across all the nodes. The small-worldness (SW) of the network is also calculated to measure the degree of small-world organization in the overall functional network. Table. I presents the metrics that are used in this work. The formulas of these metrics are described in detail in [Bonacich, 1987, 2007; Rubinov and Sporns, 2010]. All the implementations are performed using Matlab code from the Brain Connectivity Toolbox (https://sites.google.com/site/bctnet/).

3. Results

3.1. Implementation

3.1.1 Model order

The order estimation method, ER-FM, is applied to estimate the signal subspace order for the COBRE data. The mean and standard deviation of the order across subjects are 72.86 ± 10.40. We use an order equal to the mean plus one standard deviation, which is rounded up to 85 to retain a significant level of the signal across multiple subjects while

introducing minimal noise. The use of this high model order is also well motivated in the literature for achieving a more useful functional segmentation of the brain, see *e.g.*, [Allen et al., 2011; Kiviniemi et al., 2009].

Finally, in group ICA, the order of subject-level PCA is set to $T' = 100$, a little higher than 85, seeking to retain as much variability as possible [Erhardt et al., 2011]. In DL, we first try to determine a decomposition without any dimension reduction but this results in very noisy estimates regardless of the parameter values. We thus perform subject-level PCA using the same order, T', with group ICA, then perform DL on the concatenated data, $\widetilde{\mathbf{Y}} \in \mathbb{R}^{KT' \times V}$, and set the number of dictionary atoms to be 85, which is estimated by ER-FM in a data-driven manner.

3.1.2 Back-reconstruction

Direct group ICA (GICA) back-reconstruction is used to obtain subject-specific time courses and spatial components for the ICA algorithms [Calhoun et al., 2001]. Group ICA estimates the group components after performing two levels of PCA on the original data as introduced in Section 2.2. GICA provides a clear flow of the back-reconstruction of the subject-specific time courses and spatial components [Erhardt et al., 2011] as all the dimension reduction matrices in PCA are known. Indirect back-reconstruction approaches, *i.e.*, spatio-temporal regression (STR), or dual regression as used in [Beckmann et al., 2009; Calhoun et al., 2004] for fMRI analysis can also recover the subject-specific time courses and spatial components. In this paper, we perform STR, which is the most suitable back-reconstruction approach for DL to obtain the time courses and spatial maps. Because only the dictionary D is given by the online DL and regression is suggested for the recovery of the spatial maps given a dictionary D [Mairal et al., 2010]. To make the DL estimates comparable with those from ICA, GICA1, a GICA back-reconstruction method that partitions the group-level PCA reducing matrix $\mathbf{G}^{\dagger}$ as described in Section 2.2, is used to perform back-reconstruction for ICA, since it has been shown to have similar performance to dual regression [Erhardt et al., 2011].

Random initialization is used for each algorithm. For DL, the dictionary **D** is randomly initialized from the input data. The most stable run is selected using minimum spanning tree method [Du et al. 2016] for EBM from 10 runs and for ERBM and SparseICA-EBM from 25 runs. We use fewer number of runs for EBM because compared with ERBM and SparseICA-EBM, EBM takes only higher-order statistics as type of diversity into consideration, resulting in less variability across solutions. Hence using smaller number of runs is sufficient for EBM to provide a robust solution compared with ERBM and SparseICA-EBM. For Infomax, multiple runs are performed and the result shows that the spatial maps and time course power ratio statistics are very similar across different runs. Multiple runs are not performed for DL due to the consistency of its decompositions. Using back-reconstruction, the 85 components are estimated for individual subjects associated with their time courses. All the analyses are performed on the mean components that are generated by averaging the back-reconstructed components across all subjects.

3.1.3 SparseICA-EBM

As introduced above, there are two key parameters for SparseICA-EBM, the regularization parameter λ_n and the smoothing parameter ε_n. In this work, the regularization and smoothing parameters are kept the same for all estimates, thus the index n is dropped, resulting in λ_s and ε, respectively. As noted in [Boukouvalas et al., 2017b], higher values of ε yield consistent results across different levels of noise and sparsity and lead to decompositions that are better than those of lower values. While for λ_s, the performance varies based on the sparsity levels of the sources. Hence, in this work, we first try to find a suitable value of λ_s for our dataset by fixing $\varepsilon = 10$, a relatively high value. Further, with the selected λ_s, different values of ε are used. The final values of each parameter are selected by investigating the cost function, spatial maps and time course power ratio.

In order to select λ_s, a wide range of values from 10^{-8} to 10^{4} are used to study the contribution of independence and sparsity in the sparse ICA cost function. As shown in Fig. 1(a) and (c), when $\lambda_s = 10^{-5}$, the independence term is larger which means that independence contributes more and when $\lambda_s = 10^{-4}$, sparsity contributes more. For $10^{-4.5}$, the contribution of independence and sparsity is balanced as seen in Fig. 1(b). The spatial components are visually inspected for different values of λ_s. We notice that the decomposition fails to estimate all possible latent sources when λ_s is too large (*e.g.*, $\geq 10^{2}$). The estimated sources are very focal, which means that the activated area of this estimate is compact with little noise and similar components are estimated multiple times. This comes at the cost of not estimating sources whose activated area is expected to be larger or dispersive, such as the DMN and fronto-parietal, which are of interest in resting-state fMRI data [Allen et al., 2011; Greicius et al., 2003]. When λ_s is small (*e.g.*, $\leq 10^{-6}$), the decomposition becomes very similar to that of EBM, *i.e.*, sparsity scarcely contributes in these cases.

The frequency power ratio of time courses for different values of λ_s is shown in Fig. 1(d). Only those values of λ_s that between 10^{-5} and 0 are investigated. The black curve is the smoothed density histogram of power ratios of all subjects. The box plot displays the median, the 25th and 75th percentiles of the time course power ratio with whiskers extending to the 99.3% confidence interval and some outliers in red asterisks beyond whisker. The mean and standard deviation are shown in magenta. Horizontal magenta and blue lines refer to the global average, 4.20, and median power ratio, 3.50, across components from all decompositions, respectively. From Fig. 1(d), we can see that when $\lambda_s = 10^{-4.5}$, the independence and sparsity is balanced, the decomposition does not give the best statistics for power ratio summary. The mean, 25th and 75th percentiles are lower for $\lambda_s = 10^{-4.5}$ as compared with $\lambda_s = 10^{-5}$ and $\lambda_s = 10^{-4}$. Even though when $\lambda_s = 10^{-5}$ (with mean as 4.55 and median as 3.58), the decomposition gives slightly better power ratio summary than when $\lambda_s = 10^{-4}$ (with mean as 4.42 and median as 3.57). The components when $\lambda_s = 10^{-4}$ look more focal compared with those when $\lambda_s = 10^{-5}$, *e.g.*, in Fig. 1(e), the motor components when $\lambda_s = 10^{-4}$ are cleaner than those when $\lambda_s = 10^{-5}$. Therefore, we decide to use $\lambda_s = 10^{-4}$, in which case that sparsity contributes slightly more than independence.

To explore the influence of ε, we fix $\lambda_s = 10^{-4}$ and change the value of ε from 10^{-2} to 10^{4}. By visual inspection, we notice that with small ε (*e.g.*, ≤ 1), certain components are estimated repeatedly. Most of the repeated components are the ventricle components whose activated regions are focal and localized. Usually, the time course power ratio of ventricle component is low. But when they are estimated repeatedly, some of the repeated ones have power ratios as high as 10. Though this leads to slightly better power ratio, they are not reliable. For larger ε (*e.g.*, ≥ 10), the spatial maps are not estimated repeatedly. The power ratio statistics are higher when $\varepsilon = 10^{2}$ (with mean as 4.65 and median as 3.90) than when $\varepsilon = 10$ (with mean as 4.42 and median as 3.57), $\varepsilon = 10^{3}$ (with mean as 4.57 and median as 3.47) and $\varepsilon = 10^{4}$ (with mean as 4.37 and median as 3.69). Therefore, we use $\lambda_s = 10^{-4}$ and $\varepsilon = 10^{2}$ for SparseICA-EBM.

3.1.4. DL

There is only one regularization parameter λ for DL. It controls the balance between decomposition accuracy and sparsity. Different values, from 10^{-2} to 10^{2}, are explored to study its effect. We set the number of atoms as 85. For each λ, we check both the spatial components and the time course frequency power ratio. Components estimated from smaller λ are much noisier than those from larger λ, which can be seen from the examples in Fig. 2. The activated area of the spatial maps from $\lambda = 10^{2}$ is a little smaller than that from $\lambda = 10$. Additionally, components from $\lambda = 10^{2}$ are not well-aligned with those from the other values of λ, and some of them seem to split, such as the fronto-parietal component as shown in Fig. 2. Power ratio distributions obtained for each λ are shown in Fig. 3(a). Notice that when $\lambda = 10$, both the mean, 4.83, and median, 4.12, are much higher

than the global mean, 3.76, and median, 3.33. When $\lambda = 10$, the decomposition appears to provide the best performance, and hence we use this value for DL in our further experiments.

3.2. Power ratio comparison

Spatial maps of a set of mean components from each algorithm are shown in Fig. 4. These are selected among the components that can be easily matched. For those functional networks that only contain one single relevant area, all five algorithms produce a good decomposition. However, for the DMN component, EBM and ERBM estimate more relevant activated areas since they use a more flexible density estimation model. Infomax only uses a fixed unimodal super-Gaussian density model which can hurt the quality of its decomposition if the data does not correspond to this model. Consequently, the components estimated by Infomax all have very focal activated areas and have similar PDFs. SparseICA-EBM and DL emphasize the sparsity of the sources, which makes the estimated components have fewer activated voxels outside the regions of interest. However, the amplitude levels in the parietal regions marked by the cyan circles in Fig. 4 for the DMN component from Infomax, SparseICA-EBM and DL are significantly lower that that from EBM and ERBM.

The time course power ratio comparison of the 85 mean components among the five algorithms is shown in a violin plot in Fig. 3(b). Horizontal magenta and blue lines refer to the global average, 4.52, and median power ratio, 3.53, across components from all five algorithms, respectively. The results indicate that Infomax yields the lowest mean, 4.26, and median, 3.13, of the power ratio and DL yields both the highest mean, 4.83, and median, 4.12. This implies that emphasizing sparsity results in components that are more likely to be related to BOLD signal on average.

3.3. Network connection summary

In order to find M functionally relevant components to construct the network connection, a threshold of t_{PR} is selected by rounding off the global median value (3.53) across all components from all five algorithms to the closest integer less than this value, which is selected as $t_{\mathrm{PR}} = 3$ experimentally. Finally, $M = 50$ components corresponding to known networks are selected for each algorithm. The network connection summary of the 50 mean components is created for each algorithm using the normalized MI as a measure. We group these components into six domains, motor, cognitive control (COG), default mode (DM), auditory (AUD), visual (VIS) and cerebellum (CB), according to their anatomical and presumed functional properties as in [Allen et al., 2011]. Fig. 5 shows the composited spatial maps for each cluster, the functional network connectivity matrix and the network connection summary for EBM. The functional network connectivity matrix exhibits some patterns of the brain network connection, for example, the modular organization within motor, DM and so on. These patterns are consistent with the observations in prior literature [Allen et al., 2014; Yu et al., 2015]. The connection summary visually illustrates these patterns. The intra- and inter-module connectivity ratio R is 2.4, 3.2, 1.9, 3.1 and 1.7 for Infomax, EBM, ERBM, SparseICA-EBM and DL, respectively. This reveals that EBM and SparseICA-EBM yield better clustering within each functional network cluster rather than across.

3.4. Graph-theoretical analysis

Graph-theoretical analysis is performed on the selected 50 components for each subject. The graph, G, is formed, where the retained components are nodes and pairwise normalized MI between the spatial components forms the edges. For each binarized graph, G', both the nodal and global metrics are calculated. Permutation testing [Bullmore et al., 1999; Zalesky et al., 2010] is performed on nodal metrics to obtain the corrected p-value to detect the significance level ($p < 0.05$) of difference between SZ and HC groups. For each nodal graph metric, 10,000 random permutations are generated independently. We first perform two-sample t-test on the graph metric based on the

SZ and HC groups, and record the test statistic, t_0. To implement the permutation test, all the subjects are randomly divided into two groups. A two-sample t-test is performed between the two groups in each division and all the test statistics are stored. Finally, the p-value is calculated by counting the number of permutations for which the test statistic is greater than t_0 and normalizing by 10,000. In order to prevent nodes from being declared significant by chance, only those nodes that show significant differences in at least three successive graphs are declared to be truly significant.

Our results reveal that for the globally calculated nodal graph metrics, *i.e.*, degree, characteristic PL, global efficiency and three versions of centrality, more components from ERBM show consistent significance in graphs for the majority of link density values, which gives us a greater ability to explore the differences between the patient and control groups. Fig. 6 shows the plot for the number of significant components at each link density for PL and betweenness centrality, and the plots for the other nodal graph metrics are similar with these two (not shown here due to the space limitation).

Fig. 7 compares the standard deviation (STD) of CC and SW between SZs and HCs. Usually, higher variability is expected for SZs than for HCs in fMRI analysis [Demirci et al., 2009]. The normalized STD difference is calculated by first subtracting the STD of HCs from that of SZs then normalizing it by the STD of HCs. The result reveals that ERBM yields the highest contrast between SZs and HCs using both CC and SW.

4. Discussion

Data-driven methods are now very widely used in the analysis of fMRI data, especially for resting state studies, and there are a number of toolboxes that incorporate multiple algorithms for the task. In this paper, we consider the two key classes of algorithms used for the task, those that are based on independence and sparsity and compare their performances by fully taking the properties of the task, fMRI analysis into account. We also include a promising new method SparseICA in the comparison that balances the contribution of the two objectives and demonstrate its first application to real fMRI data. An important goal of this paper has been providing feedback to practitioners who are mainly focused on using the methods without the need to perform full performance evaluations. We address issues in two categories in that respect in this section, and also discuss some of the limitations of the work.

4.1. Parameter selection

Parameter choice for optimized performance is always an important issue, and for practitioners, this might not always be an easy task, especially for approaches that aim to include the right amount of sparsity through regularization parameters into the decomposition such as DL and SparseICA. We investigate the influence of parameters in SparseICA-EBM and DL, and rather than using cross-validation that focuses on prediction performance [Kohavi, 1995; Wernick et al., 2010], we consider the ultimate goal in fMRI analysis and use suitable metrics for the task. For different values of λ_s and ε in SparseICA-EBM, we use cost, time course power ratio and visual inspection to determine the suitable values. Our experimental results from real fMRI data show that when λ_s is too large, components are estimated repeatedly by SparseICA-EBM. Some important components, such as DMN and fronto-parietal, are missed. Large ε appears to produce good performance, which is consistent with the results in [Boukouvalas et al., 2017b]. For different values of λ in DL, time course power ratio and visual inspection are performed to determine its suitable value. The experimental results suggest that $\lambda = 10$ yields the best estimation. Even though different values of sparsity parameters are selected, *i.e.*, $\lambda_s = 10^{-4}$ for SparseICA-EBM and $\lambda = 10$ for DL, the sparsity level imposed by them are similar. The Gini index is computed to measure the sparsity level of a source. It is a normalized quantity with 1 corresponding to very sparse sources while 0 to dense ones [Hurley et al., 2009]. The results show that the distribution of Gini indices of all

85 components estimated from DL is close to that from SparseICA-EBM. The Kullback-Leibler divergence from the distribution of DL to that of SparseICA-EBM is estimated as 0.0393.

For ICA, the main parameter of choice is the signal subspace order that determines the total number of components to estimate. In this work, the model order is set as 85. Recently, many fMRI studies that use data-driven techniques such as ICA favor a higher model order, typically within the range [70, 100] since this yields a more detailed decomposition of the resting-state networks which are also more repeatable [Allen et al., 2011; Abou-Elseoud et al., 2010; Damaraju et al., 2014; Kiviniemi et al., 2009]. Small orders such as those in the range [20, 40] are not favored in newer studies since they most often include overlapping or combined estimates of resting-state networks. However, model orders that are greater than 100 decrease the stability of ICA which is as expected in other data-driven techniques as well [Abou-Elseoud et al., 2010]. There are a number of methods for performing order selection in a data-driven manner, and ER-FM provides a desirable match to fMRI data characteristics as the span of sample correlation is modeled as finite in this model, and is what we have used in this analysis. Thus for parameter-free algorithms like Infomax and EBM, the fMRI analysis can be totally data driven, which might also explain the popularity ICA has enjoyed from the point of view of practitioners. For DL, one advantage that has been noted is that dimensionality reduction stage can be bypassed [Abraham et al., 2013; Varoquaux et al., 2011], however our study using global metrics demonstrate that estimates tend to be noisy without dimensionality reduction regardless of the parameter values including dictionary size. Hence the observation of our experiment suggests that DL is not able to separate signal components from noise components effectively without dimensionality reduction when applied to the COBRE data used in this paper. Thus it is advisable to incorporate dimension reduction into DL as well either directly using PCA or by imposing low rank structure. In this work, subject-level PCA with the same order as in ICA is performed before applying DL.

4.2. Component estimation

From the comparison results using time course power ratio, network connection summary and graph-theoretical analysis, we can see that different algorithms stand out in different cases. This can be explained by the intrinsic variance of these decompositions which depend on dissimilar modeling assumptions.

DL provides better identification of the components that describe the BOLD response. As observed in Fig. 3(b), the power ratio mean and median of the components from DL are higher than those from the other algorithms, which illustrates that on average, DL has the capability to estimate components that are more likely to describe the BOLD response. Fig. 4 shows that the activated area in components that are estimated from Infomax, SparseICA-EBM and DL are very focal and localized. However, components, such as DMN and fronto-parietal, that have multiple relevant regions are not completely estimated by these three algorithms. It is known that DMN contains multiple areas, the ventral anterior cingulate cortex (vACC), posterior cingulate cortex (PCC), left and right inferior parietal cortex (IPC). Usually, PCC is estimated together with the left and right IPC, and vACC is estimated in another individual component called anterior DMN. The DMN components from Infomax, SparseICA-EBM and DL are only activated in PCC. Similar observations are noted for fronto-parietal network where ERBM yields a fronto-parietal component with both the frontal and parietal region activated. Consequently, EBM and ERBM provide better estimation of functionally relevant components.

Fig. 5 shows the network connection summary of EBM and it illustrates that components from EBM can help to reconstruct a good functional network connectivity of brain. The reconstructed connection from EBM shows dense connectivity within functional networks and also demonstrates reasonable inter-connectivity among different functional networks. The close-packed connectivity within

functional networks may provide valuable clues for the development of those automatic clustering algorithms for the estimated components.

Though DL provides better identification of the components that describe the BOLD response, it yields the lowest intra- and inter-module connectivity ratio *R*. To account for higher-order dependence, we use mutual information between pairwise spatial maps to measure connectivity. Time course power ratio is a metric for time course comparison, while *R* is used to summarize the spatial dependence. Therefore, there is no expectation of higher *R* for DL. This is also an evidence that it is to study higher-order dependence rather than only concentrating on second order dependence, *i.e.*, correlation.

In the event that one is analyzing data from two groups and seeking to find group differences, ERBM appears to be more preferable as ERBM can produce components with more relevant regions activated thus leading to more effective discrimination in a graph-theoretical analysis. From the graph-theoretical analysis results shown in Fig. 6, ERBM estimates more components that show significant group difference when using globally calculated nodal graph metrics, such as degree, PL, global efficiency and centrality. Usually, dysconnectivity is observed for individuals with schizophrenia [Bullmore et al., 1997; Stephan et al., 2009]. Therefore, higher variance is expected in connectivity analysis for patients compared to HCs. In Fig. 7, ERBM demonstrates improved ability to capture the contrast of the variability of SZs and HCs using the global graph metrics, CC and SW. We also note that there is no scenario where widely-used Infomax algorithm provides the best performance, demonstrating the cost of exploiting fewer forms of diversity. This suggests that different algorithms can be chosen for different purposes.

Finally, there are some limitations in our work. (1) We do not provide a generalized parameter selection for all algorithms. For example, the filter length used in ERBM is a parameter that might be further investigated. The actual filter length that can make the samples of fMRI data as independent as possible is not fixed throughout the data so the value we use, which is determined empirically, might be suboptimal. Similarly, the regularization and smoothing parameters, λ_s and ε, in SparseICA-EBM can be tuned differently for individual components to offer adaptive and robust estimation. Larger and smaller values of the number of atoms in D for DL can be investigated to study the influence of over- and under-determined scenarios. (2) The best run selection strategy that is used for EBM, ERBM and SparseICA-EBM has not been thoroughly investigated. There are a number of issues in best run selection, such as whether to select the most accurate or the most consistent run, which metrics to use for evaluation and whether they are reliable. However, taking all of these issues in a computationally reasonable manner is a challenge. Furthermore, our work is based on the selected run, but the average performance across multiple runs is also informative.

Some future directions of this work are as follows. (1) A reliability test of these metrics can be done for further justification. Even though these metrics have shown their capability to evaluate the algorithmic performance, it is not clear that they are stable for different datasets come from the same group of subjects. As test-retest strategy is now very popular in reliability testing [Du et al., 2017; Shirer et al., 2015; Termenon et al., 2016]. Similar experiments can be performed to further evaluate the faithfulness of these global metrics in fMRI data analysis. (2) The relationship between the cost and global measures can be explored by tuning the parameters of the algorithm. The objective global measures are independent of cost functions. However, changing the parameters of the algorithm will lead to different source estimates thus varying the results obtained for the global measures. The effect of varying these parameters on the global measures is an interesting topic for further analysis.

5. Conclusion

Data-driven methods that make use of independence and sparsity have proven useful in many applications in particular for the analysis of fMRI data. This motivates a performance comparison of those algorithms that incorporate

independence and/or sparsity on real fMRI data. Consequently, appropriate metrics for performance evaluation are needed since the decompositions can be significantly different and it is usually impossible to match all the estimated components across methods. In this work, we propose to use three global metrics to assess the performance, namely, time course spectra power ratio, network connection summary and graph-theoretical metrics. With the use of time course spectra power ratio, DL yields components that are most likely to describe the BOLD response. By using network connection summary, EBM illustrates better ability to reconstruct the brain connectivity with normalized mutual information as the measure. When graph-theoretical analysis is applied, ERBM demonstrates better capability to capture the group differences between SZs and HCs, especially the higher variance in SZs as expected. This successful application motivates the use of additional global metrics to assess the performance for unbiased algorithmic comparison on real fMRI data analysis, or even on other real data analysis such as electroencephalograph data. Additionally, we should note that there is no scenario where the widely used Infomax has shown best performance, which demonstrates that incorporating multiple types of diversity is more desirable.

Acknowledgment

This work was supported by grants NSF-CCF 1618551, NSF 1631838, NIH R01EB 020407, and NIH P20GM103472. The authors thank the research staff at COBRE who collected, preprocessed and shared the data. The authors appreciate the valuable feedback provided by the members of Machine Learning for Signal Processing Laboratory in University of Maryland, Baltimore County.

References

Abolghasemi, V., Ferdowsi, S., and Sanei, S. (2015). Fast and incoherent dictionary learning algorithms with application to fMRI. *Signal, Image and Video Processing*, 9(1), 147–158.

Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., and Kiviniemi, V. (2010). The effect of model order selection in group PICA. *Human Brain Mapping*, 31(8), 1207–1216.

Abraham, A., Dohmatob, E., Thirion, B., Samaras, D., and Varoquaux, G. (2013, September). Extracting brain regions from rest fMRI with total-variation constrained dictionary learning. *In*: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 607–615.

Adalı, T., Anderson, M., and Fu, G.-S. (2014). Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging. *IEEE Signal Processing Magazine*, 31(3), 18–33.

Aine, C., Bockholt, H., Bustillo, J., Cañive, J., Caprihan, A., Gasparovic, C., Hanlon, F., Houck, J., Jung, R., Lauriello, J., et al. (2017). Multimodal neuroimaging in schizophrenia: Description and dissemination. *Neuroinformatics*, 15(4), 343–364.

Alexander-Bloch, A.F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., Lenroot, R., Giedd, J., and Bullmore, E.T. (2010). Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in Systems Neuroscience*, 4, 147.

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., and Calhoun, V.D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3), 663–676.

Allen, E.A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Frontiers in Systems Neuroscience*, 5, 2.

Amari, S., Cichocki, A., and Yang, H.H. (1996). A new learning algorithm for blind signal separation. in *Advances in Neural Information Processing Systems*, 8, 757–763

Bassett, D.S., Nelson, B.G., Mueller, B.A., Camchong, J., and Lim, K.O. (2012). Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3), 2196–2207.

Beckmann, C.F., Mackay, C.E., Filippini, N., and Smith, S.M. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *NeuroImage*, 47(Suppl 1), S148.

Bell, A., and Sejnowski, T. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129–1159.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4), 175– 308.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5), 1170–1182.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29 (4), 5.

Boukouvalas, Z., Levin-Schwartz, Y., and Adalı, T. (2017a, March). Enhancing ICA performance by exploiting sparsity: Application to fMRI analysis. *In: Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, 2532–2536.

Boukouvalas, Z., Levin-Schwartz, Y., Calhoun, V. D., and Adalı, T. (2017b). Sparsity and independence: Balancing two objectives in optimization for source separation with application to fMRI analysis. *Journal of the Franklin Institute*.

Bullmore, E., Frangou, S., and Murray, R. (1997). The dysplastic net hypothesis: an integration of developmental and

dysconnectivity theories of schizophrenia. *Schizophrenia Research*, 28(2), 143–156.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10 (3), 186–198.

Bullmore, E.T., and Bassett, D.S. (2011). Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, 7, 113–140.

Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M.J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(1), 32–42.

Calhoun, V.D., and Adalı, T. (2012). Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery. *IEEE Reviews in Biomedical Engineering*, 5, 60–73.

Calhoun, V.D., Adalı, T., Pearlson, G.D., and Pekar, J.J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140–151.

Calhoun, V.D., Liu, J., and Adalı, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45(1), S163–S172.

Calhoun, V.D., Pekar, J.J., and Pearlson, G.D. (2004). Alcohol intoxication effects on simulated driving: Exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology*, 29(11), 2097–2107.

Calhoun, V.D., Potluru, V.K., Rhlypo, R., Silva, R.F., Pearlmutter, B.A., Caprihan, A., Plis, S.M., and Adalı, T. (2013). Independent component analysis for brain fMRI does indeed select for maximal independence. *PLoS ONE*, 8(10), e73309.

Çetin, M.S., Christensen, F., Abbott, C.C., Stephen, J.M., Mayer, A.R., Cañive, J.M., Bustillo, J.R., Pearlson, G.D., and Calhoun, V.D. (2014). Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *NeuroImage*, 97, 117–126.

Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., and Meyerand, M.E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in resting-state data. *American Journal of Neuroradiology*, 22(7), 1326–1333.

Correa, N., Adalı, T., and Calhoun, V.D. (2007). Performance of blind source separation algorithms for fMRI analysis using a group ICA method. *Magnetic Resonance Imaging*, 25(5), 684–694.

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D. H., Mueller, B. A., Pearlson, G. D., Potkin, S. G., Preda, A., Turner, J. A., Vaidya, J. G., van Erp, T. G., and Calhoun, V. D. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, 5, 298-308.

Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C., D'Ardenne, K., Richter, W., Cohen, J.D., and Haxby, J. (2009). Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences*, 106(26), 10415–10422.

Demirci, O., Stevens, M.C., Andreasen, N.C., Michael, A., Liu, J., White, T., Pearlson, G.D., Clark, V.P., and Calhoun, V.D. (2009). Investigation of relationships between fMRI brain networks in the spectral domain using ICA and Granger causality reveals distinct differences between schizophrenia patients and healthy controls. *NeuroImage*, 46(2), 419–431.

Du, W., Levin-Schwartz, Y., Fu, G.-S., Ma, S., Calhoun, V.D., and Adalı, T. (2016). The role of diversity in complex ICA algorithms for fMRI analysis. *Journal of Neuroscience Methods*, 264, 129–135.

Du, W., Li, H., Li, X.-L., Calhoun, V. D., and Adalı, T. (2011, April). ICA of fMRI data: Performance of three ICA algorithms and the importance of taking correlation information into account. *In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1573–1576.

Du, Y., Lin, D., Yu, Q., Sui, J., Chen, J., Rachakonda, S., Adalı, T., and Calhoun, V.D. (2017). Comparison of IVA and GIG-ICA in brain functional network estimation using fMRI data. *Frontiers in Neuroscience*, 11.

Erhardt, E.B., Rachakonda, S., Bedrick, E.J., Allen, E.A., Adalı, T., and Calhoun, V.D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Human Brain Mapping*, 32(12), 2075–2095.

Freire, L., Roche, A., and Mangin, J.-F. (2002). What is the best similarity measure for motion correction in fMRI time series? *IEEE Transactions on Medical Imaging*, 21(5), 470–484.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., and Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.

Fu, G.-S., Anderson, M., and Adalı, T. (2014). Likelihood estimators for dependent samples and their application to order detection. *IEEE Transactions on Signal Processing*, 62(16), 4237–4244.

Goria, M.N., Leonenko, N.N., Mergel, V.V., and Novi Inverardi, P.L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Journal of Nonparametric Statistics*, 17(3), 277–297.

Greicius, M.D., Krasnow, B., Reiss, A.L., and Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1), 253–258.

Hurley, N., and Rickard S. (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10), 4723–4741.

Hyvärinen, A., Karhunen, J., and Oja, E. (2009). Independent Component Analysis. *John Wiley & Sons, Inc.*

Kiviniemi, V., Starck, T., Remes, J., Long, X., Nikkinen, J., Haapea, M., Veijola, J., Moilanen, I., Isohanni, M., Zang, Y.-F., and Tervonen, O. (2009). Functional segmentation of the brain cortex using high model order group PICA. *Human Brain Mapping*, 30(12), 3865–3886.

Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In: International Joint Conference on Artificial Intelligence (IJCAI)*, 14, 1137–1145.

Lee, K., Tak, S., and Ye, J. C. (2011). A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. *IEEE Transactions on Medical Imaging*, 30(5), 1076–1089.

Lee, S.-I., Lee, H., Abbeel, P., and Ng, A. Y. (2006). Efficient L_1 regularized logistic regression. *In: Proceedings of the National Conference on Artificial Intelligence*, 6, 401–408.

Lee, T.-W., Girolami, M., and Sejnowski, T.J. (1999). Independent component analysis using an extended Infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, 11(2), 417–441.

Li, K., Guo, L., Nie, J., Li, G., and Liu, T. (2009a). Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics*, 33(2), 131–139.

Li, X.-L., and Adalı, T. (2010a, March). Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization. *In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 1934–1937.

Li, X.-L., and Adalı, T. (2010b). Independent component analysis by entropy bound minimization. *IEEE Transactions on Signal*

Processing, 58(10), 5151–5164.

Li, X.-L., Ma, S., Calhoun, V.D., and Adalı, T. (2011 April). Order detection for fMRI analysis: Joint estimation of downsampling depth and order by information theoretic criteria. *In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1019–1022.

Li, Y.-O., Adalı, T., and Calhoun, V.D. (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Human Brain Mapping*, 28(11), 1251–1266.

Li, Y.-O., Adalı, T., Wang, W., and Calhoun, V. D. (2009b). Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing*, 57(10), 3918–3929.

Long, Q., Bhinge, S., Levin-Schwartz, Y., Calhoun, V.D., and Adalı, T. (2017, March). A graph theoretical approach for performance comparison of ICA for fMRI analysis. *In: Information Sciences and Systems (CISS), 2017 51st Annual Conference on*, 1–6.

Lynall, M.-E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., and Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, 30(28), 9477–9487.

Ma, S., Calhoun, V.D., Eichele, T., Du, W., and Adalı, T. (2012). Modulations of functional connectivity in the healthy and schizophrenia groups during task and rest. *NeuroImage*, 62(3), 1694–1704.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11 (Jan), 19–60.

Mckeown, M.J., Makeig, S., Brown, G.G., Jung, T.-P., Kindermann, S.S., Bell, A.J., and Sejnowski, T.J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6, 160–188.

Nielsen, F., and Nock, R. (2011). A closed-form expression for the Sharma–Mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical*, 45(3), 023003.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238.

Robinson, S., Basso, G., Soldati, N., Sailer, U., Jovicich, J., Bruzzone, L., Kryspin-Exner, I., Bauer, H., and Moser, E. (2009). A resting state network in the motor control circuit of the basal ganglia. *BMC Neuroscience*, 10(1), 1.

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52 (3), 1059–1069.

Scott, A., Courtney, W., Wood, D., De la Garza, R., Lane, S., Wang, R., King, M., Roberts, J., Turner, J.A., and Calhoun, V.D. (2011). COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Frontiers in Neuroinformatics*, 5, 33.

Shin, J., Ahn, S., and Hu, X. (2013). Correction for the T1 effect incorporating flip angle estimated by Kalman filter in cardiac-gated functional MRI. *Magnetic Resonance in Medicine*, 70(6), 1626–1633.

Shirer, W.R., Jiang, H., Price, C.M., Ng, B., and Greicius, M.D. (2015). Optimization of rs-fMRI preprocessing for enhanced signal-noise separation, test-retest reliability, and group discrimination. *NeuroImage*, 117, 67–79.

Stephan, K.E., Friston, K.J., and Frith, C.D. (2009). Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin*, 35(3), 509–527.

Telesford, Q.K., Burdette, J.H., and Laurienti, P.J. (2013). An exploration of graph metric reproducibility in complex brain networks. *Frontiers in Neuroscience*, 7, 67.

Termenon, M., Jaillard, A., Delon-Martin, C., and Achard, S. (2016). Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *NeuroImage*, 142, 172–187.

Van Den Heuvel, M.P., and Pol, H.E.H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519–534.

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., and Thirion, B. (2011, July). Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. *In: Biennial International Conference on Information Processing in Medical Imaging*, 562–573.

Vértes, P.E., Alexander-Bloch, A.F., Gogtay, N., Giedd, J.N., Rapoport, J.L., and Bullmore, E.T. (2012). Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences*, 109(15), 5868–5873.

Wernick, M.N., Yang, Y., Brankov, J.G., Yourganov, G., and Strother, S.C. (2010). Machine learning in medical imaging. *IEEE Signal Processing Magazine*, 27(4), 25–38.

Yu, Q., Erhardt, E.B., Sui, J., Du, Y., He, H., Hjelm, D., Cetin, M.S., Rachakonda, S., Miller, R.L., Pearlson, G., and Calhoun, V.D. (2015). Assessing dynamic brain graphs of time-varying connectivity in fMRI data: Application to healthy controls and patients with schizophrenia. *NeuroImage*, 107, 345–355.

Zalesky, A., Fornito, A., and Bullmore, E.T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4), 1197–1207.

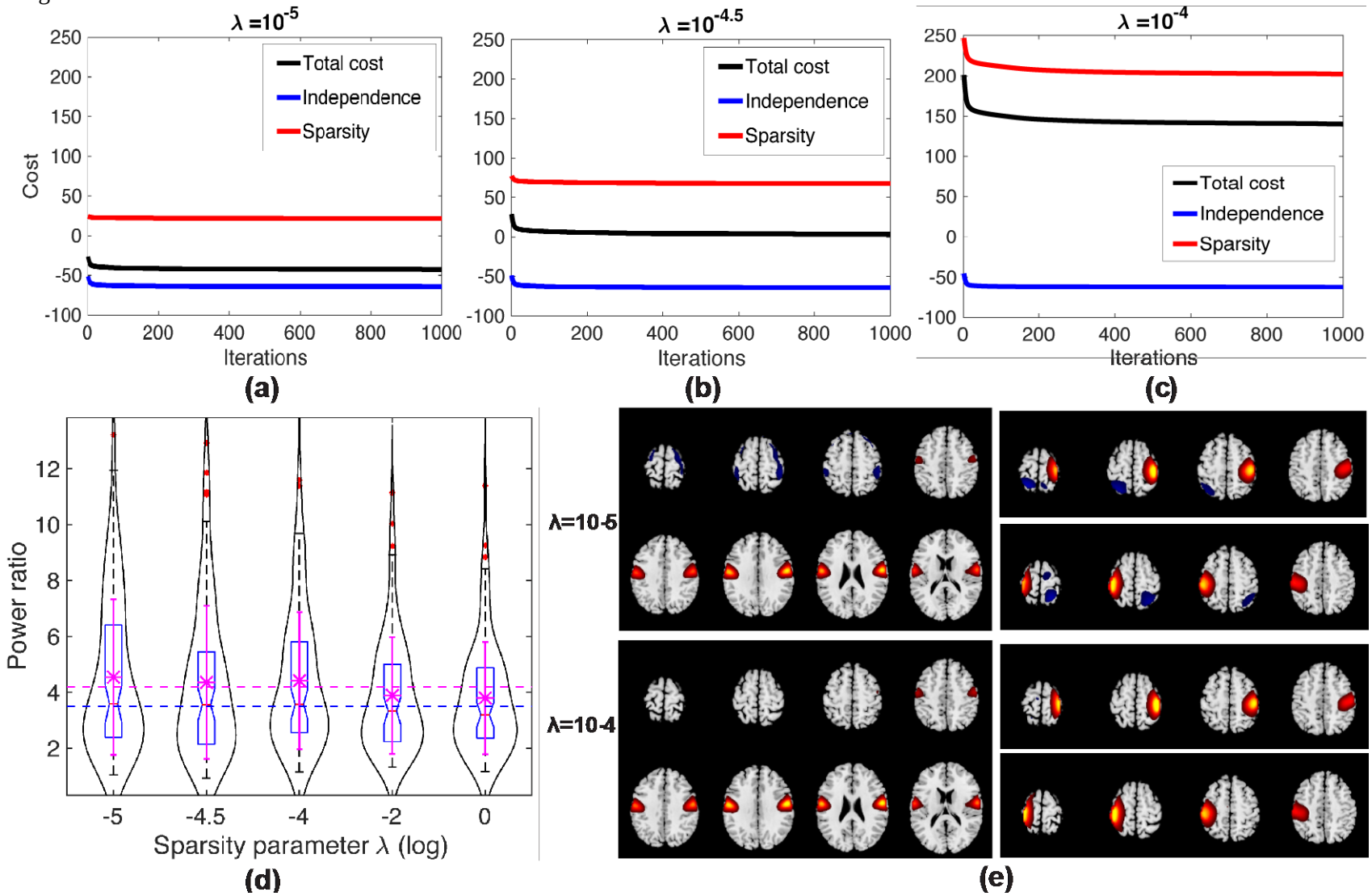

Figure. 1 The relative contribution of statistical independence and sparsity (a)--(c) in SparseICA-EBM, the violin plot of the time course power ratio (d) and the spatial maps of motor components estimated by SparseICA-EBM when $\lambda_s = 10^{-5}$and $\lambda_s = 10^{-4}$ (e). In (a)-(c), the x-axis denotes the number of SparseICA-EBM iterations and y-axis denotes the cost. In (e), Z-maps corresponding to the mean components averaged across all subjects for motor components are shown. The maps are thresholded using a threshold $Z_t = 2$.

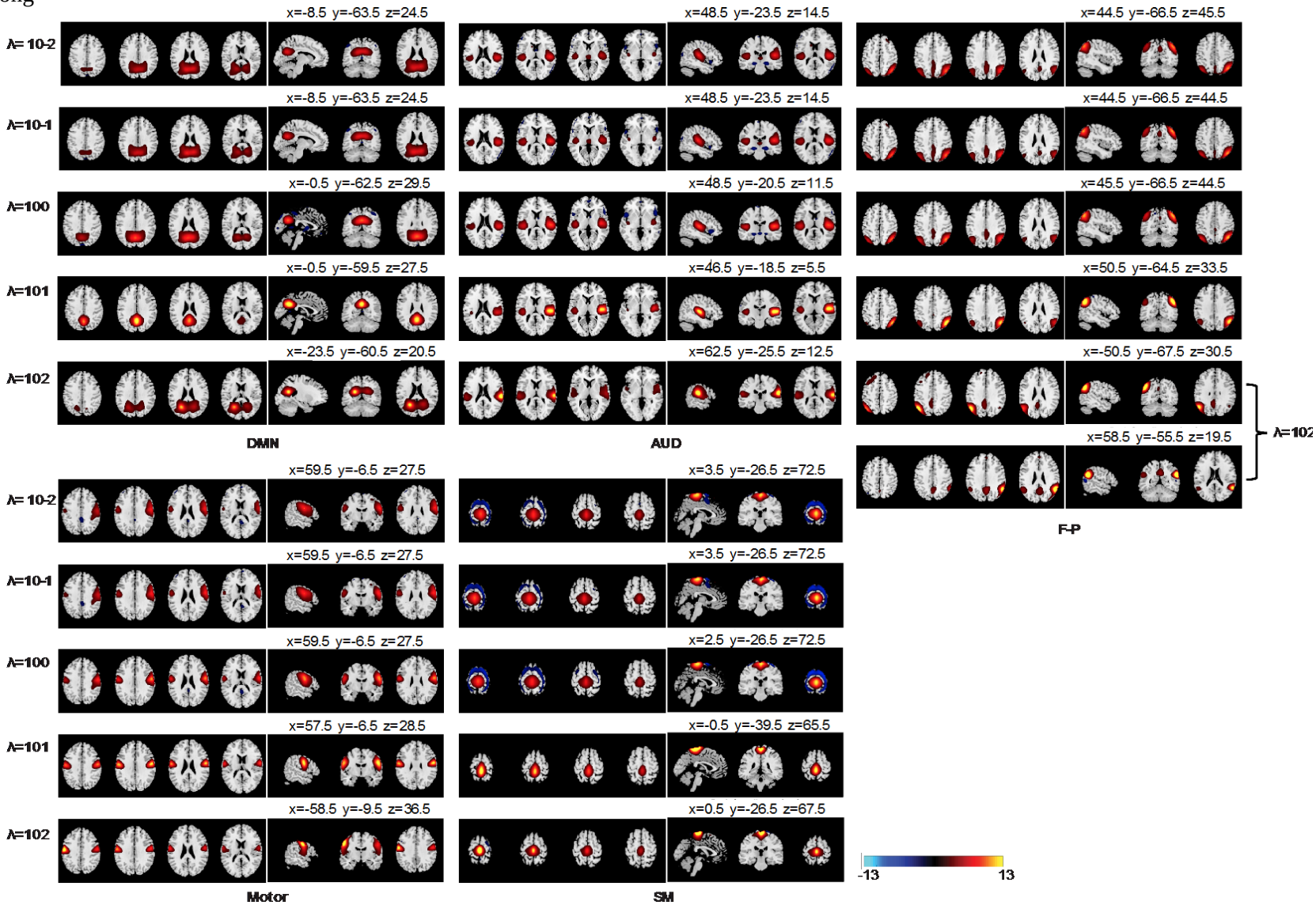

Figure. 2 Comparison of several components from DL using different values of λ. Z-maps corresponding to the mean components averaged across all subjects for DMN, auditory (AUD), motor, sensorimotor (SM) and fronto-parietal (F-P) component are shown. The maps are thresholded using a threshold $Z_t = 2$. The peak coordinates in mm are shown to the right. Note that for $\lambda = 10^2$, there are two fronto-parietal

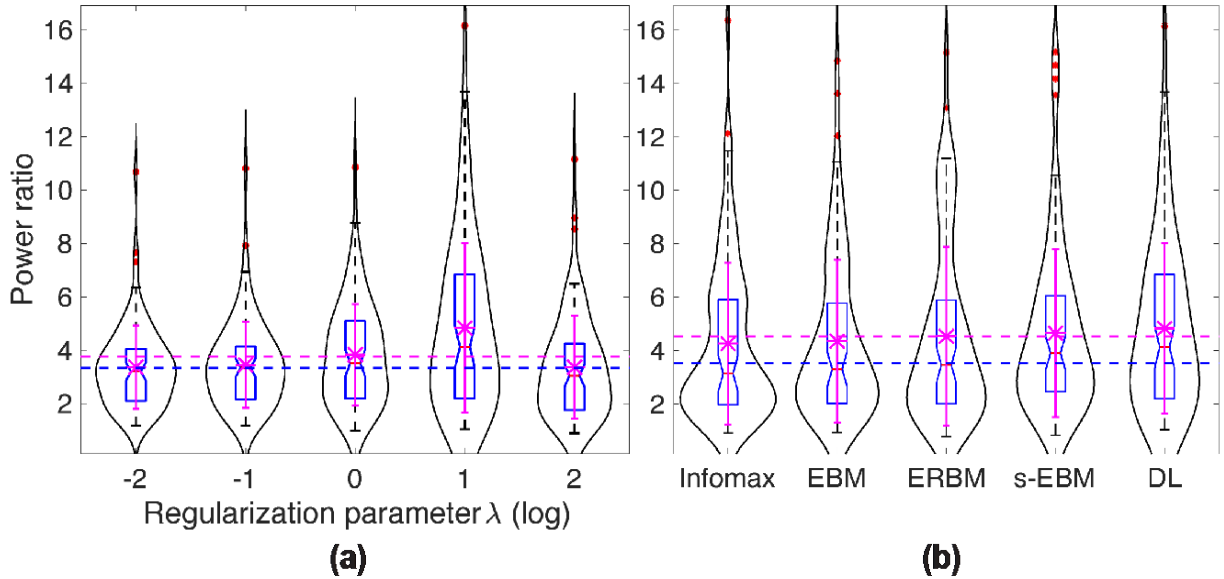

Figure. 3 The violin plot of the time course power ratio for all 85 mean components of (a) DL using different values of λ and (b) five algorithms. Note that s-EBM refers to SparseICA-EBM.

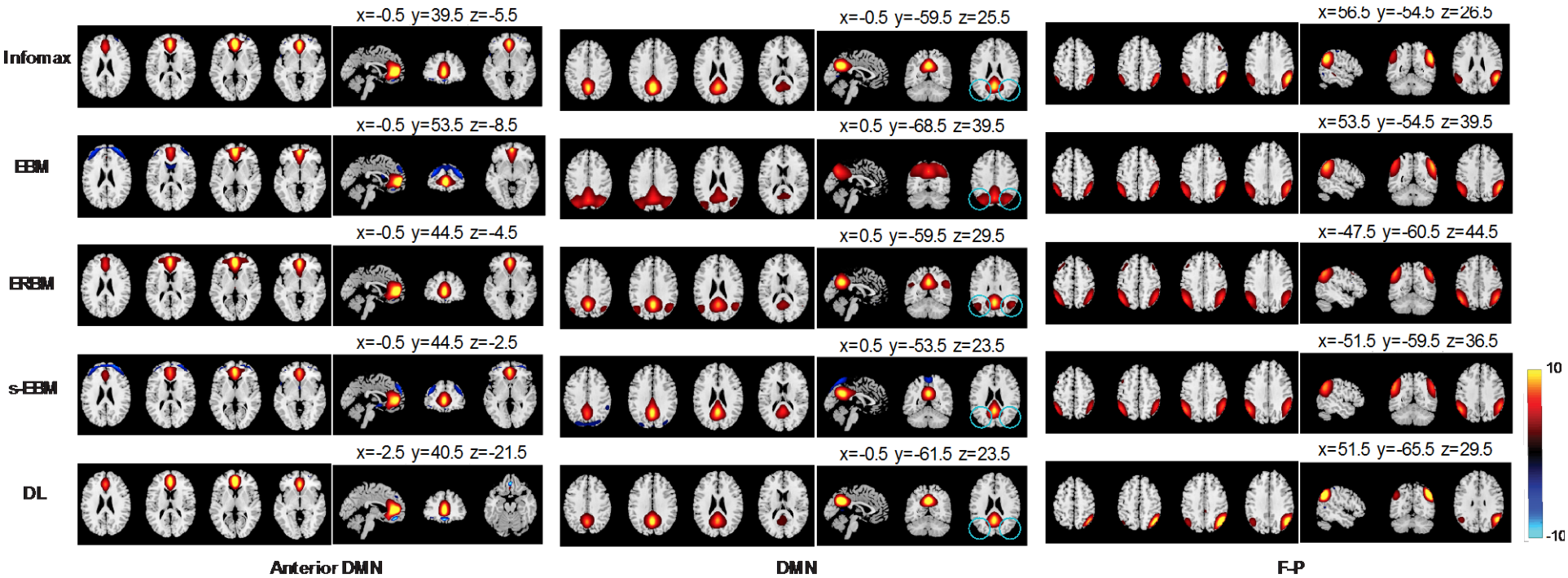

Figure. 4 Comparison of several components from five algorithms. Z-maps corresponding to the mean components averaged across all subjects for the anterior DMN, DMN and fronto-parietal (F-P) component are shown. The maps are thresholded using a threshold $Z_t = 2$. The peak coordinates in mm are shown to the right. Note that s-EBM refers to SparseICA-EBM.

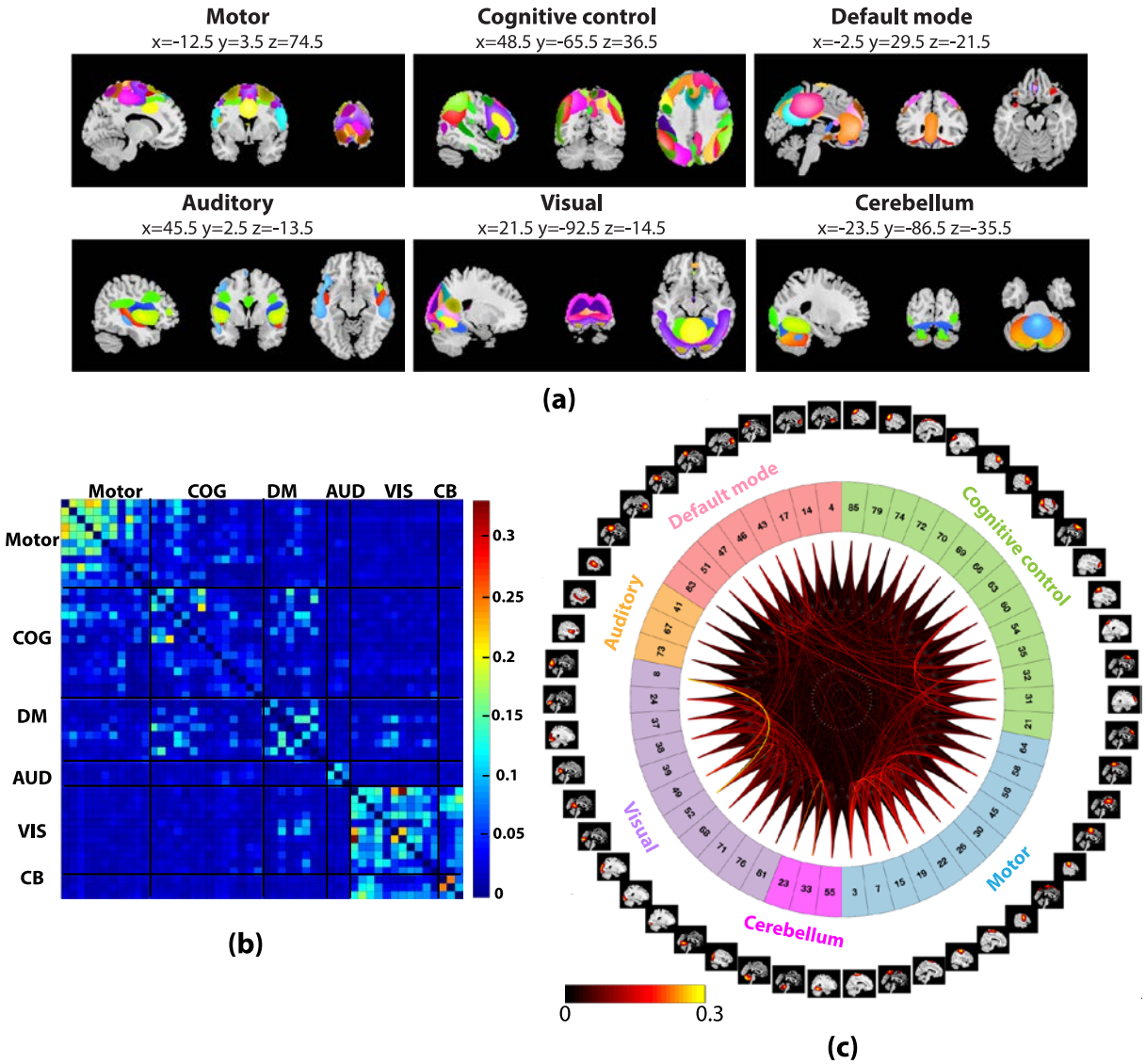

Figure. 5 Network connection summary of components from EBM. (a) Spatial maps of 50 functional networks. Functional networks are divided into groups based on their anatomical and functional properties. COG: Cognitive control; DM: default mode; AUD: auditory; VIS: visual; CB: cerebellum. (b) The functional connectivity matrix. Connectivity is measured using normalized MI. (c) Network connectivity visualization. The outermost circle demonstrates one slice of the spatial map of individual mean component. The colored circle indicates the index corresponding to each component and different colors refer to different functional network clusters. Curves in the center carry the exact connectivity among these components.

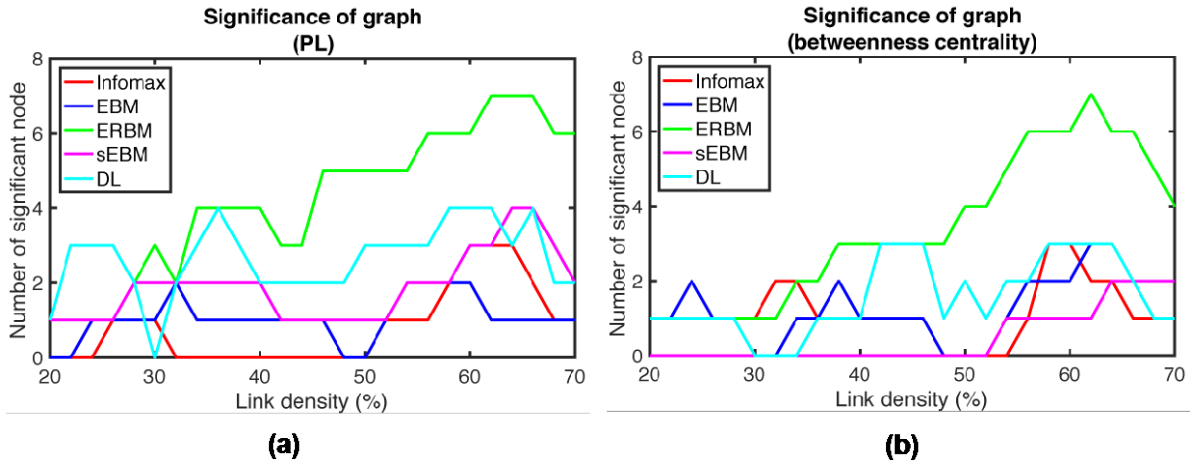

Figure. 6 Number of nodes showing significant group difference in each graph as a function of link density for (a) characteristic path length (PL) and (b) betweenness centrality. Note that s-EBM refers to SparseICA-EBM.

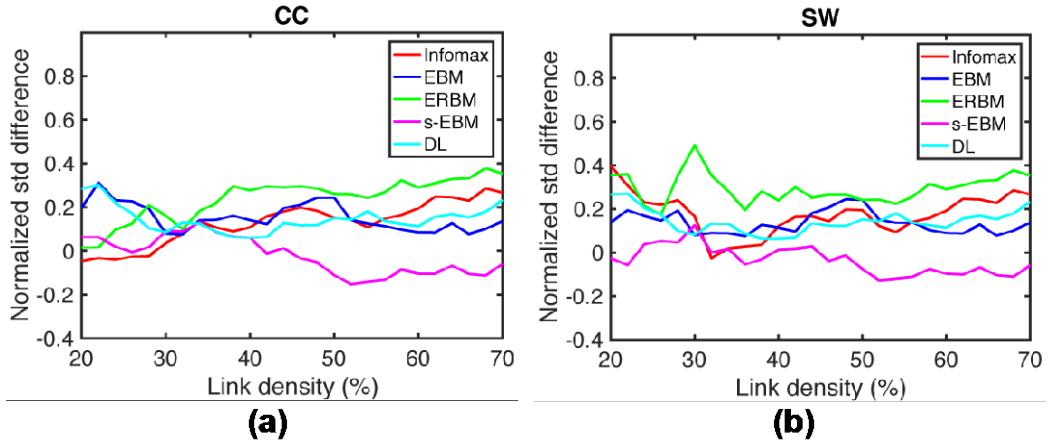

Figure. 7 Comparison of the normalized STD of (a) clustering coefficient (CC) and (b) small-worldness (SW) in SZs and HCs. Note that s-EBM refers to SparseICA-EBM.

Metric	Abbreviation	Description
Nodal metrics (globally calculated)		
Degree	—	Number of links directly connected to node *i*
Characteristic path length	PL	Average distance of node *i* to all the others
Global efficiency	—	Communication efficiency of node *i* with all the others
Centrality		
Betweenness centrality	—	Prominence of node *i* in information transfer in network
Closeness centrality	—	
Eigenvector centrality	—	
Global metrics		
Clustering coefficient	CC	Measure of functional segregation of the network
Small-worldness	SW	Quantifying the ability of combining functional integration and segregation

TABLE I: Description of graph-theoretical metrics in both node and graph level.