Published in final edited form as:
Med Image Anal. 2014 December ; 18(8): 1337–1348. doi:10.1016/j.media.2014.06.006.

Identifying group discriminative and age regressive sub-networks from DTI-based connectivity via a unified framework of non-negative matrix factorization and graph embedding

Yasser Ghanbari[a], **Alex R. Smith**[a], **Robert T. Schultz**[b], and **Ragini Verma**[a,*]

[a]Center for Biomedical Image Computing and Analytics, Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States

[b]Center for Autism Research, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, United States

Abstract

Diffusion tensor imaging (DTI) offers rich insights into the physical characteristics of white matter (WM) fiber tracts and their development in the brain, facilitating a network representation of brain's traffic pathways. Such a network representation of brain connectivity has provided a novel means of investigating brain changes arising from pathology, development or aging. The high dimensionality of these connectivity networks necessitates the development of methods that identify the connectivity building blocks or sub-network components that characterize the underlying variation in the population. In addition, the projection of the subject networks into the basis set provides a low dimensional representation of it, that teases apart different sources of variation in the sample, facilitating variation-specific statistical analysis. We propose a unified framework of non-negative matrix factorization and graph embedding for learning sub-network patterns of connectivity by their projective non-negative decomposition into a reconstructive basis set, as well as, additional basis sets representing variational sources in the population like age and pathology. The proposed framework is applied to a study of diffusion-based connectivity in subjects with autism that shows localized sparse sub-networks which mostly capture the changes related to pathology and developmental variations.

Keywords

Diffusion MRI; Connectivity analysis; Non-negative matrix factorization; Graph embedding; Autism and development

1. Introduction

Computational techniques applied to neuroimaging data have helped unveil the underlying structural or functional differences between groups of interest, e.g. patients and healthy controls. Study of brain connectivity has recently gained a lot of attention in investigating

*Corresponding author. Address: Section of Biomedical Image Analysis, 3600 Market St, Suite 380, Philadelphia, PA 19104, United States. Ragini.Verma@uphs.upenn.edu (R. Verma).

the origin of many brain disorders such as autism spectrum disorder (Jou et al., 2011; Vissers et al., 2012), schizophrenia (Price et al., 2007; Skudlarski et al., 2010), and Alzheimer's disease (Bozzali et al., 2011; Matthews et al., 2013) as well as in the study of development (Dennis et al., 2013; Ingalhalikar et al., 2013). Hence, advanced techniques of brain connectivity analysis are emerging as a powerful tool in studies of brain disorders as well as in the study of brain development. Such tools quantify the connectivity between two regions of interest (ROIs) in DTI, fMRI, EEG, or MEG recordings by using measures such as fiber tracking (Mori and van Zijl, 2002; Friman et al., 2006), independent component analysis (Calhoun et al., 2008), mutual information (Salvador et al., 2010), and synchronization likelihood (Ghanbari et al., 2013). Connectivity is then represented by a square matrix of size equal to the number of ROIs, where each element of the matrix represents the connectivity measure between corresponding ROIs (Rubinov and Sporns, 2010).

Diffusion MRI (dMRI) (Bihan et al., 1986, 2001, 2003; Basser and Pierpaoli, 1996; Basser et al., 2000; Alexander et al., 2007; Paul et al., 2007) offers a rich insight into the complex "highway network" of white matter (WM) fiber pathways in the brain by capturing the diffusion patterns of water molecules aligned with microscopic tissue architecture. Diffusion is a three dimensional process which is mathematically modeled by 3-D tensors, which is representative of how water diffuses in the underlying tissue. Thus, dMRI provides unique clues of neural organization structures in terms of fiber bundles (Bihan, 2003). This brings the opportunity to quantify the anatomical connectivity using fiber bundle density measures (Lanyon, 2012; Clayden, 2013). The connectivity measures are subsequently used to construct the brain structural network, known as the connectome, to analyze the structural brain connectivity as a large complex network (Sporns, 2011). Structural connectome is hence represented by non-negative quantities indicative of anatomical pathways between regions of the brain. Its relationship with anatomy is the constraining factor for it to be non-negative, where a zero means no pathway, while a positive number indicates presence of fiber pathway with its magnitude representing the connectivity strength.

A number of established analysis methods are available for studying the underlying brain structure via a network representation. Graph theory metrics (Bullmore, 2009; Rubinov and Sporns, 2010) have been recently introduced to analyze complex organization of brain networks by providing features such as small-worldness, modularity, and participation coefficient (Bassett et al., 2011; Ingalhalikar et al., 2013). A successful analysis methodology must possess a means of identifying relevant sub-networks, providing an interpretable representation of the brain network with non-negative quantities, while also facilitating the statistical analysis that describes how this representation is affected by disease. The traditional approaches, i.e. principal and independent components analysis (PCA and ICA) used for investigating brain functional networks (Calhoun et al., 2008) provide dimensionality reduction but may lack physiological interpretability, due to no non-negativity constraints, when applied to structural connectivity networks.

Recently, non-negative matrix factorization (NMF) and its alternatives have received extensive attention and proven effective in providing an interpretable set of bases characterizing multivariate data. After being first introduced by Lee and Seung (1999), NMF

has been successfully employed in many applications such as signal processing, pattern recognition, and data mining (Berry et al., 2007; Yan et al., 2007; Wang et al., 2009; Yang and Oja, 2010; Batmanghelich et al., 2012). Its part-based representation of data, as well non-negativity constraints on both the bases and coefficients, facilitates interpretability, and its small size of the basis set categorizes NMF among the dimensionality reduction techniques. Although these methods are useful for interpretation due to their positivity and sparsity, they do not necessarily provide discriminative bases, only bases which best reconstruct the original data.

We consider the connectivity matrix of a subject as linear combination of several fundamental connectivity matrices called connectivity component sub-networks. The approach taken here is the decomposition of such connectivity matrices into dominant components, while enforcing positivity on both the components and coefficients. Such decomposition maintains the interpretation of each component as a network connectivity matrix and the coefficients associated with these components as weights of those networks, while providing a low dimensional representation of the population amenable to statistical analysis. Such an interpretation is obtained by enforcing non-negativity constraints to both the component sub-networks that represent structural connectivity, as well as their coefficients that represent their contributing weights. Fig. 1 illustrates the decomposition of a connectivity matrix into several sub-network modules.

The analysis of connectomes of a population requires meaningful reduction of the dimensionality to obtain sub-networks that describe the population. The weight of each sub-network describes the contribution of that sub-network to a subject which can be used to learn about the subject's standing within its population group spectrum. These weights can also be used to perform statistical analysis and regression. Such a representation is able to capture different aspects of population and facilitates different types of analyses. Hence, three types of components are needed to extract: discriminative, regressive, and reconstructive components. Discriminative components are used to identify groups (such as patients-control) within the population and their pattern of connectivity difference. Regressive components characterize the connectivity pattern of a continuous score such as age or a clinical score. Reconstructive components are also needed to ensure that the original connectome input data would be reconstructed with minimal error in conjunction with discriminative and regressive components.

To compute our basis components towards discriminative, regressive, and reconstructive sub-networks, we take advantage of graph embedding. Graph embedding approaches have recently gained a lot of attention in data analysis (Craddock et al., 2012), dimensionality reduction (van den Maaten et al., 2009), and clustering (Belkin and Niyogi, 2002; Ng et al., 2002). These methods mainly use the intrinsic geometry of the points on the data manifold to embed a graph into an objective function to keep neighborhood structure in the lower dimensional space. Locally linear embedding (Roweis and Saul, 2000; Saul and Poweis, 2000) and Laplacian eigenmaps (Belkin and Niyogi, 2002; He and Niyogi, 2003) are among the most popular techniques applying graph embedding to find an appropriate low dimensional representation. Such techniques have proven to successfully provide low dimensional discriminative/regressive features in medical imaging applications such as in

MRI based classification of Alzheimer's disease (Liu et al., 2013) and in characterizing trajectories of brain development (Aljabar et al., 2010).

Hence, we design a new framework accounting for the intrinsic geometrical information of the data manifold that helps to categorize our non-negative basis set into discriminative (i.e. providing group differences), regressive (age related in this paper) and reconstructive (i.e. providing low reconstruction error) components. This is modeled by minimizing the Frobenius norm of the decomposition residual error when graph embedding is added. We form a unified objective function to be minimized by a gradient descent approach with appropriate step size while guaranteeing the positivity of bases and their coefficients.

While the method is generalizable to any type of non-negative network, we demonstrate the applicability of the framework to DTI-based structural connectivity for a population of subjects with autism spectrum disorder (ASD). Our method is able to extract components that describe the underlying sub-network patterns of pathology related variability (discriminative components), as well as sub-network patterns of age variation (developmental components). Differences seen are pertinent to deficiencies in thalamic network and sub-cortical inter-hemespheric connectivity in autism, while there are other differentiating sub-networks in autism that are linked with age as well.

2. Material

2.1. Participants

24 individuals with ASD and 59 age-matched typically developing controls were recruited with no known genetic conditions associated with ASD. Ages of ASD subjects ranged from 7.8 to 18.3 years with mean = 12.9 and SD = 3.0. The TDCs aged from 6.2 to 17.2 years with mean = 11.6 and SD = 3.2. The groups did not significantly differ in age. Participants with a community diagnosis of an ASD were recruited in part through autismMatch (https://autismmatch.org), and diagnoses were confirmed using gold standard diagnostic instruments (ADOS (Lord et al., 2000) and ADI-R (Lord et al., 1994)) and expert consensus clinical judgment by two independent psychologists following Collaborative Programs of Excellence in Autism (CPEA) diagnostic guidelines. Tests include performance based measures collected by a psychologist with each study participant, as well as parent rating forms designed to quantify behavioral features of ASD.

2.2. Image acquisition

DTI data was acquired on a Siemens 3T Verio™ scanner, using a 32 channel head coil and a single shot spin-echo, echo-planar sequence with the following parameters: TR/TE = 11,000/76 ms, *b*-value of 1000 s/mm^2, 30 gradient directions. Eighty 2 mm contiguous axial slices of 128 × 128 matrix (FOV 256 mm) yielded 2 mmisotropic data. Quality assurance (QA) of the images was performed manually and the ones with poor quality were removed leaving 83 images with acceptable quality.

2.3. DTI processing and tractography algorithm

Brain extraction procedure was performed on each diffusion volume of DTI data acquired for each subject, from which fractional anisotropy (FA) volume was computed. The high resolution T1 structural images were parcellated into 79 regions (68 cortical and 11 sub-cortical) using Desikan atlas (Desikan et al., 2006) in Freesurfer. The 79 region labels were then transferred to the diffusion space where the GM-WM boundaries were determined. Probabilistic fiber tracking (Behrens et al., 2003a,b) was performed from each of these regions with 5000 streamline fibers sampled per voxel, resulting in a 79 × 79 matrix of weighted connectivity values, where each element of the matrix represents the conditional probability of a pathway between regions, normalized by the active surface area of the seed ROI. Fig. 2 demonstrates this procedure.

3. Methods

We view the connectivity matrix obtained from the structural brain connectivity network of subjects in a population as linear combination of several fundamental connectivity matrices called connectivity component sub-networks. There are challenges in computing such sub-networks from a set of DTI structural connectomes: (a) these sub-networks have to be represented by connectivity matrices with positive elements because there is no concept of negative connections in DTI and (b) brain structure does not deactivate and hence the coefficients of those sub-networks have to be positive as well. To pursue this goal, we develop a method here that guarantees these constraints while producing sub-networks representative of pathology and development.

Due to the symmetry of connectivity matrices, a vector of all elements of the upper triangular part of any connectivity matrix is considered as a representative of that matrix, and is used as an observation vector $\boldsymbol{x}_i$ for the corresponding subject i. We note that this is not a constraint of the method, i.e. if the connectivity network is represented by an asymmetric matrix (or a directed graph) then all elements of the connectivity matrix are vectorized into $\boldsymbol{x}_i$, not just the upper triangular ones.

To compute the connectivity components, a matrix factorization model is used as follows.

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{\Phi} + \varepsilon \quad (1)$$

where represents the residual error of the decomposition operation, columns of $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \boldsymbol{R}^{m\times n}$, i.e. $\boldsymbol{x}_i$ $(1 \le i \le n)$, are the vectorized connectivity matrix, and columns of $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_r] - \in \boldsymbol{R}^{m\times r}$, i.e. $\boldsymbol{w}_j$ $(1 \le j \le r)$, are representative of the normalized basis connectivity components, i.e. the upper triangular elements of the matrix of the corresponding connectivity component. These components $\boldsymbol{w}_j$ are then mixed by the elements of each column of the loading matrix $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_n] \in \boldsymbol{R}^{r\times n}$ to approximate the corresponding column of $\boldsymbol{X}$ (Yang and Oja, 2010), i.e. $\boldsymbol{x}_i \approx \sum_{j=1}^{r} \Phi_{ji} w_j$; $1 \le i \le n$. Fig. 3 illustrates this procedure.

3.1. Projective non-negative basis learning

We assume that $\boldsymbol{\Phi}$ is the projection of $\boldsymbol{X}$ into $\boldsymbol{W}$, i.e. $\boldsymbol{\Phi}=\boldsymbol{W}^T\boldsymbol{X}$, the non-negativity constraint on the elements of $\boldsymbol{W}$ and $\boldsymbol{\Phi}$ makes our non-negative component analysis an optimization problem of minimizing the cost function $F_1(\boldsymbol{W}) = \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\|^2$ with respect to $\boldsymbol{W}$, where $\|.\|$ represents the matrix norm. Considering the Frobenius norm, the minimization problem can be denoted by

$$\min_{w\geq 0} F_1(\boldsymbol{W}) = \min_{w\geq 0} trace\{(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})\,(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})^T\}. \quad (2)$$

In order to demonstrate the effectiveness of non-negativity constraints in identifying the true underlying components, we apply the method of Eq. (2) to simulated connectivity matrices. We compare this with ICA which is the most widely-adopted method for similar purposes. The simulation is performed by using random non-negative numbers as the elements of three 10×10 simulated symmetric matrices (as components) plus a slight (non-negative) random background noise. Ten linear mixtures of the simulated connectivity matrices are composed by a random mixing matrix with non-negative elements. The upper triangular part (excluding the diagonal) of the 10 connectivity matrices are vectorized to form the 10 columns of $\boldsymbol{X}_{45\times 10}$. We apply the fast ICA algorithm (Hyvarinen and Oja, 2000) as well as the projective NMF of Eq. (2) (Ghanbari et al., 2012) to solve for $p = 3$ normalized components as columns of $\boldsymbol{W}_{45\times 3}$. The results are shown in Fig. 4. For visualization, the matrices and the elements of the resulting components are displayed by grayscale images. It can be seen that ICA is unable to resolve the components correctly while the solution to Eq. (2) can. This is due to the fact that ICA forces the components to be statistically independent whereas Eq. (2) forces the components to be non-negative and orthogonal which leads to localized components.

3.2. Locality preserving bases with graph embedding

In order to regularize our connectivity sub-network components towards possessing discriminative and regressive properties, we take advantage of the locality of the points in the high dimensional space of the vectorized connectivity. Firstly, we split the set of projective bases into two sets of $\boldsymbol{W} = [\hat{\boldsymbol{W}}, \boldsymbol{W}, \breve{\boldsymbol{W}}, \tilde{\boldsymbol{W}}]$ in which $\hat{\boldsymbol{W}} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q]$ are the *discriminative* basis components, $\boldsymbol{W} = [\boldsymbol{w}_{q+1}, \ldots, \boldsymbol{w}_{q+p}]$ are the *developmental* basis components, and $\breve{\boldsymbol{W}} = [\boldsymbol{w}_{q+p+1}, \ldots, \boldsymbol{w}_r]$ is the complimentary space containing the *reconstructive* basis components which minimizes the reconstruction error together with $\hat{\boldsymbol{W}}$ and $\boldsymbol{W}$. Thus, the coefficient matrix $\boldsymbol{\Phi}$ is also split into $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{W}}^T\boldsymbol{X}$, $\breve{\boldsymbol{W}} = \breve{\boldsymbol{W}}^T\boldsymbol{X}$, and $\tilde{\boldsymbol{\Phi}} = \tilde{\boldsymbol{W}}^T\boldsymbol{X}$. A proper modeling of such intent would provide at most q of those bases which are likeliest to provide population discrimination to belong to $\hat{\boldsymbol{W}}$, and p of those which are likeliest to account for the developmental variations in $\boldsymbol{W}$.

The locality preserving property that we impose in our method is based on the assumption that the two-group multivariate m-dimensional connectivity points lie on a manifold in the high dimensional space. In the first embedding, if the connectivity matrices are close in the m-dimensional spaces, then their proximity is preserved in the lower dimension using the set of discriminative components. The developmental components will embed subjects close in

age together, in order to preserve the age continuum. The third set of reconstructive components will preserve large distances to capture the global data variations so that the basis decomposition scheme will ensure reconstruction of the original connectivity data.

To clarify the idea behind our mathematical modeling, suppose that the *m*-dimensional points of two groups lie on a 3-D manifold in the *m*-dimensional space, as illustrated in Fig. 5(a), and are to be projected into an $r = 3$ dimensional subspace with $q = 1$ discriminative (say $\vec{\boldsymbol{y}}$), and $p = 1$ developmental (say $\vec{\boldsymbol{x}}$), and $r - p - q = 1$ reconstructive components (say $\vec{\boldsymbol{z}}$). To achieve this, we construct three separate graphs with the *m*-dimensional connectivity points as their vertices. The first graph is an intrinsic graph of ***k***-nearest neighbors (He and Niyogi, 2003; Yan et al., 2007; Wang et al., 2009) which connects point *i* to point *j* if point *j* is among the *k* nearest neighbors of point *i* (as illustrated in Fig. 5(b)). This graph is employed to obtain the discriminative component sub-networks $\hat{\boldsymbol{W}}$ by keeping the *m*-dimensional points as close to each other as possible in the subspace spanned by $\hat{\boldsymbol{W}}$. The second graph, used for the developmental components, connects point *i* to point *j* if subject *j*'s age is among the *k* nearest ages to subject *i*, irrespective of the two points' distance (illustrated in Fig. 5(c)). This graph is used to compute the age-regressive sub-networks $\breve{\boldsymbol{W}}$ by keeping the connectivity points of subjects with similar ages near each other in the subspace of $\breve{\boldsymbol{W}}$. Finally, the third graph is formed by connecting point *i* to point *j* if point j is among the *k* farthest points to point *i* (as illustrated in Fig. 5(d)). Making far points close to each other will avoid the inclusion of the corresponding sub-network wrongly as representative of a discriminative or developmental component. As seen in Fig. 5(a), the axis $\vec{\boldsymbol{x}}$ could be an incorrect potential candidate for the discriminative component since it keeps nearby points closer to each other compared to axis $\vec{\boldsymbol{y}}$ (the desired discriminative component). Hence, having it captured in the third set of components will avoid wrongly categorizing this component into the discriminative, instead of reconstructive set. It will also help capture other variations of the data that was not captured by the discriminative/ developmental bases to help a better reconstruction of the data (i.e. better representation of the data variations).

There are a variety of approaches that can characterize separability of multivariate datapoints. Most of such techniques can be unified in the framework of graph embedding (Yan et al., 2007). Let $G = \{\boldsymbol{X}, \boldsymbol{S}\}$ be an undirected weighted graph of *n* vertices, i.e. datapoints $\boldsymbol{x}_i$, with a symmetric similarity matrix $\boldsymbol{S} \in \Re^{n \times n}$ with non-negative elements, within the range of 0–1, corresponding to the edge weight of the graph. The Laplacian matrix (He and Niyogi, 2003) of the graph is then defined by

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}; \quad D_{ii} = \sum_{j=1}^{n} S_{ij}, \quad \forall i. \tag{3}$$

In order for the bases in $\hat{\boldsymbol{W}}$ to provide discriminatory information, we would like the resulting coefficients of nearby $\boldsymbol{x}_i$ points to stay close to each other to group together when projected into $\hat{\boldsymbol{W}}$. This can be obtained by minimizing

$$\min_{\boldsymbol{w}\geq 0} F_2(\boldsymbol{W}) = \min_{\hat{\boldsymbol{w}}\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{n} \|\hat{\varphi}_i - \hat{\varphi}_j\|^2 \hat{S}_{ij} = \min_{\hat{\boldsymbol{w}}\geq 0} trace\{\hat{\boldsymbol{\Phi}}\hat{L}\hat{\boldsymbol{\Phi}}^T\}, \quad (4)$$

where $\hat{\boldsymbol{S}} = [\hat{S}_{ij}]$ is similarity matrix composed of the edge weights in the intrinsic graph of k nearest neighbors $\hat{G} = \{\boldsymbol{X}, \hat{\boldsymbol{S}}\}$ of the m-dimensional points $\boldsymbol{x}_i$, as illustrated in Fig. 5(b). The edge weight of neighbor points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined by

$$\hat{S}_{ij} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\hat{\sigma}^2}\right), \quad (5)$$

where $\hat{\sigma}$ is a scaling parameter. In this scheme, $\hat{S}_{ij}$ is non-zero, if and only if $\boldsymbol{x}_i$ is among the k-nearest-neighbors of $\boldsymbol{x}_j$ or vice versa, leading $\hat{\boldsymbol{S}}$ to be symmetric. According to Eq. (5), if datapoints $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close, their graph edge weight $\hat{S}_{ij}$ will be large, and therefore, the cost function $F_2(\boldsymbol{W})$ gets minimized only if the corresponding coefficients $\hat{\varphi}_i$ and $\hat{\varphi}_j$ remain close. It is worth noting that proper choice of sigma will highlight the impact of subjects whose connectivity matrices are better representative of their population (i.e. lie around the center of the population distribution). Such proper choices of sigma will also lower the effect of subjects whose connectivity points are outliers. This helps the optimization function to concentrate on the more population representative connectivity data. The choice of scaling parameters is explained in Section 3.5.

Similarly, in order to capture the space of developmental variations, we employ the developmental graph $\breve{G} = \{\boldsymbol{X}, \breve{\boldsymbol{S}}\}$, as shown in Fig. 5(c), in which points whose subjects are of similar ages are connected. Therefore, the space of developmental sub-network, $\breve{\boldsymbol{W}}$, grouping the coefficients of subjects with similar ages, is computed by minimizing

$$\min_{\boldsymbol{w}\geq 0} F_3(\boldsymbol{W}) = \min_{\breve{\boldsymbol{w}}\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{n} \|\breve{\varphi}_i - \breve{\boldsymbol{S}}_j\|^2 \breve{S} = \min_{\hat{\boldsymbol{w}}\geq 0} trace\{\breve{\boldsymbol{\Phi}}\breve{\mathbf{L}}\breve{\boldsymbol{\Phi}}^T\}. \quad (6)$$

The similarity matrix $\breve{\boldsymbol{S}}$ gets non-zero values for connected points in the developmental graph, whose elements are determined by

$$\breve{S}_{ij} = \exp\left(-\frac{\|age_i - age_j\|^2}{\breve{\sigma}^2}\right), \quad (7)$$

with $\breve{\sigma}$ being a scaling parameter and age_i is the age of subject i.

As explained earlier, we exploit the graph of k-farthest points $\tilde{G} = \{\boldsymbol{X}, \tilde{\boldsymbol{S}}\}$, as illustrated in Fig. 5(d), to impose the representative coefficients ($\tilde{\varphi}_i$) of the farthest points to remain as close as possible in the lower dimensional space when projected into the reconstructive set $\tilde{\boldsymbol{W}}$. This is performed by minimizing

$$\min_{\boldsymbol{w}\geq 0} F_4(\boldsymbol{W}) = \min_{\tilde{\boldsymbol{w}}\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{n} \|\tilde{\varphi}_i - \tilde{\varphi}_j\|^2 \tilde{S}_{ij} = \min_{\tilde{\boldsymbol{w}}\geq 0} trace\{\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{L}}\tilde{\boldsymbol{\Phi}}^T\}. \quad (8)$$

where the non-zeros elements of the similarity matrix forming the graph edges in $\tilde{G}$ is calculated by

$$\tilde{S}_{ij} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\tilde{\sigma}^2}\right). \quad (9)$$

It can be shown that minimizing Eq. (8) in conjunction with Eqs. (4) and (6) will help the sub-network components of $\hat{\boldsymbol{W}}$ maintain discriminative properties and components of $\breve{\boldsymbol{W}}$ be age-regressive.

3.3. Objective function

To achieve the above four objectives, the final objective function is formed by $F(\boldsymbol{W}) = F_1(\boldsymbol{W}) + \lambda(F_2(\boldsymbol{W}) + F_3(\boldsymbol{W}) + F_4(\boldsymbol{W}))$. According to the projective properties of the model, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T\boldsymbol{X}$, the final objective function can be modeled to minimize

$$\begin{aligned} F(\boldsymbol{W}) = & trace\{(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X})^T\} \\ & + \lambda(trace\{\hat{\boldsymbol{W}}^T\boldsymbol{X}\hat{\boldsymbol{L}}\boldsymbol{X}^T\hat{\boldsymbol{W}}\} + trace\{\breve{\boldsymbol{W}}^T\boldsymbol{X}\breve{\boldsymbol{L}}\boldsymbol{X}^T\breve{\boldsymbol{W}}\} \\ & + trace\{\tilde{\boldsymbol{W}}^T\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\}), \end{aligned} \quad (10)$$

where λ is a tunable parameter to balance the two terms of reconstruction error norm and graph embedding.

3.4. Optimization solution

Minimizing the objective function of Eq. (10) with non-negativity constraints of $\boldsymbol{W}$ yields the optimal projective bases among which q likeliest discriminative ones are obtained in $\hat{\boldsymbol{W}}$ and p developmental sub-networks are captured in $\breve{\boldsymbol{W}}$ while the rest of components play the major role in reconstructing the original networks. To minimize the final objective function, we use a gradient descent approach, i.e. updating $W_{ij} = W_{ij} - \eta_{ij}\frac{\partial F}{\partial W_{ij}}$ with a positive step-size η_{ij}, where

$$\frac{\partial F}{\partial \boldsymbol{W}} = -4(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}) + 2(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}) + 2(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}) + \lambda[2\boldsymbol{X}\hat{\boldsymbol{L}}\boldsymbol{X}^T\hat{\boldsymbol{W}}, 2\boldsymbol{X}\breve{\boldsymbol{L}}\boldsymbol{X}^T\breve{\boldsymbol{W}}, 2\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}]. \quad (11)$$

Regarding that $\hat{\boldsymbol{L}} = \hat{\boldsymbol{D}} - \hat{\boldsymbol{S}}$, $\breve{\boldsymbol{L}} = \breve{\boldsymbol{D}} - \breve{\boldsymbol{S}}$ and $\tilde{\boldsymbol{L}} = \tilde{\boldsymbol{D}} - \tilde{\boldsymbol{S}}$, and the fact that both $\boldsymbol{D}$ and $\boldsymbol{S}$ have non-negative elements, our non-negativity constraint is guaranteed by positive initialization of $\boldsymbol{W}$ and applying the step-size as follows:

$$\eta_{ij}=\frac{\frac{1}{2}W_{ij}}{(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}+\lambda[\boldsymbol{X}\hat{\boldsymbol{D}}\boldsymbol{X}^T\hat{\boldsymbol{W}},\boldsymbol{X}\breve{\boldsymbol{D}}\boldsymbol{X}^T\breve{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}])_{ij}}. \quad (12)$$

This results in the a multiplicative updating solution as

$$W_{ij}=W_{ij}\frac{\left(2\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\lambda[\boldsymbol{X}\hat{\boldsymbol{S}}\boldsymbol{X}^T\hat{\boldsymbol{W}},\boldsymbol{X}\breve{\boldsymbol{S}}\boldsymbol{X}^T\breve{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{S}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}]\right)_{ij}}{\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}+\lambda\left[\boldsymbol{X}\hat{\boldsymbol{D}}\boldsymbol{X}^T\hat{\boldsymbol{W}},\boldsymbol{X}\breve{\boldsymbol{D}}\boldsymbol{X}^T\breve{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}. \quad (13)$$

For the stability of convergence, at each iteration, each column of $\boldsymbol{W}$ is normalized by $\boldsymbol{w}_i=\frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|_2}$. Starting with initial random positive elements on $\boldsymbol{W}$, the iterative procedure will converge to the desired $\boldsymbol{W}=[\hat{\boldsymbol{W}},\boldsymbol{W},\breve{\boldsymbol{W}},\tilde{\boldsymbol{W}}]\geq 0$.

3.5. Graph edge weights

The parameters σ, $\hat{\sigma}$, $\breve{\sigma}$, and $\tilde{\sigma}$ in Eqs. (5), (7), and (9) are scaling measures of similarity between two points. Such scaling parameters are commonly set by trial and error, but this approach requires manual intervention and is time-consuming. A self-tuning local scaling parameter δ_i was proposed in (Manor and Perona, 2004) for each datapoint $\boldsymbol{x}_i$ and is used to calculate the scaling parameter separately for each edge of the graph as follows:

$$\sigma_{ij}^2=\delta_i\delta_j;\quad \delta_i=\|\boldsymbol{x}_i-\overleftrightarrow{\boldsymbol{x}}_{i,k}\|, \quad (14)$$

where $\boldsymbol{x}\overleftrightarrow{}_{i,k}$ is the most distant point among the k-nearest neighbors of $\boldsymbol{x}_i$. Here, δ_i becomes undesirably large for the outliers and small for those near the center of the data distribution. This would lead σ_{ij} in (14) to offset the distance effect in the computation of S_{ij} and consequently D_{ii}. This effect is illustrated in Fig. 6(b).

To avoid this problem, we propose to set the scaling parameter of the graph $\hat{G}$ by

$$\hat{\sigma}^2=\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i\right]^2;\quad \hat{\delta}_i=\|\boldsymbol{x}_i-\hat{\hat{\boldsymbol{x}}}_{i,k}\|, \quad (15)$$

where $\hat{\hat{\boldsymbol{x}}}_{i,k}$ is the most distant point among the k-nearest neighbors of $\boldsymbol{x}_i$. This is a suitable approach because $\hat{\delta}_i$ becomes large for the outliers and small for the points near the center of each distribution in the high dimensional space. The average of the $\hat{\delta}_i$s is dominated by the edges of the points around the center of population distribution, because the number of points around the distribution center exceeds the number of outliers. Synthetic 2-D points, distributed normally, were created to show that the scaling parameter of Eq. (15) outperforms Eq. (14), in constructing a weighted graph with center points having higher edge weights and outliers receiving lower edge weights in the graph. This is demonstrated in Fig. 6.

With the same rationale, the scaling parameter of the graphs $\check{G}_{i,k}$ and $\tilde{G}$ are set, respectively, by

$$\breve{\sigma}^2_{i,k}=\left[\frac{1}{n}\sum_{i=1}^{n}\breve{\delta}\right]^2;\quad \breve{\delta}_i=\|age_i-a\breve{g}\,e_{i,k}\|, \quad (16)$$

and

$$\tilde{\sigma}^2=\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{\delta}_i\right]^2;\quad \tilde{\delta}_i=\|\boldsymbol{x}_i-\tilde{\tilde{\boldsymbol{x}}}_{i,k}\|, \quad (17)$$

where $a\breve{g}\,e_{i,k}$ is the farthest age among the k nearest ages to subject i, and $\tilde{\tilde{\boldsymbol{x}}}_{i,k}$ is the least distant point among the k-farthest points to $\boldsymbol{x}_i$.

3.6. Group analysis model

As stated by Eq. (1), the n connectivity observations, i.e. $\boldsymbol{x}_i$ ($1 \le i \le n$), in the matrix $\boldsymbol{X}$ are approximated by

$$[\boldsymbol{x}_1,\ldots\boldsymbol{x}_n]=[\boldsymbol{w}_1,\ldots,\boldsymbol{w}_r]\begin{bmatrix}\Phi_{11} & \ldots & \Phi_{1n}\\ \vdots & & \vdots\\ \Phi_{r1} & \ldots & \Phi_{rn}\end{bmatrix}. \quad (18)$$

Each observation vector per subject i is thus, approximately reconstructed by

$$\boldsymbol{x}_i \approx \sum_{j=1}^{r}\Phi_{ji}\boldsymbol{w}_j=\sum_{j=1}^{r}(\boldsymbol{w}_j^T\boldsymbol{x}_i)\boldsymbol{w}_j;\quad 1\le i\le n. \quad (19)$$

Thereby, the presence of each component $\boldsymbol{w}_j$ in the corresponding connectivity vector of a subject $\boldsymbol{x}_i$ ($1 \le i \le n$), is characterized by the corresponding coefficients $\boldsymbol{\Phi}_{ji}$. Let us suppose, with no loss of generality, that the first n_1 elements are from the first group (e.g. population of patients) and the remaining $n_2 = n - n_1$ from the second group (e.g. controls). Therefore, the statistical significance between the set of $\{\boldsymbol{\Phi}_{ji} : 1 \le i \le n_1\}$ and $\{\boldsymbol{\Phi}_{ji} : n_1 + 1 \le i \le n\}$ describes the importance of the corresponding connectivity basis $\boldsymbol{w}_j$ in differentiating the two groups. Similarly, a developmental component $\boldsymbol{w}_j$ can be evaluated by correlating its coefficients $\boldsymbol{\Phi}_j$. with the age of the subjects.

4. Results

The proposed method provides a framework for extracting three sets of network components from the population. These three basis sets help us understand the primary global dominant networks as well as pattern-based discriminatory and developmental sub-networks characterizing population variations. The discriminatory and developmental set of components are expected to show localized sparse sub-networks which mostly capture the

changes related to pathology and developmental variations, but do not considerably contribute in the reconstruction of the original connectivity matrices. The reconstructive basis set consists of global networks of dominant connectivity patterns. The number of components in the basis is population dependent; however, we show that even with relatively small numbers, we can obtain stable group differences.

4.1. Connectivity component analysis

The 79 × 79 connectivity matrix of each subject was vectorized to its 3081 upper triangular elements. The vectorized connectivity of all subjects were stacked to create the observation matrix $\boldsymbol{X}$ with $m = 3081$ rows and $n = 83$ columns. We pooled both patient and control groups because performing decomposition on the pooled ASD and TDC populations yields a common set of network basis components facilitating statistical group comparison, whereas obtaining components separately from the two populations yields coefficients that are statistically incomparable as they do not share the same mapping space, thereby causing spurious group differences. Moreover, pooling the patient and control groups will provide the between-group variability in the data which will be captured by the basis sub-networks.

To compute the components, we first constructed the three graphs explained in Section 3.2. We set $k = 3$ for all three graphs to construct the three nearest neighbor graph for discriminative, three nearest age graph for developmental, and three farthest point graph for reconstructive components. The similarity matrix representing graph edge weights were all obtained using the Gaussian kernels explained in Eqs. (5), (7), and (9). The tuning parameter between dominant component analysis and graph embedding (in Eq. (10)) was set to $\lambda = 1$.

We used $q = 4$ discriminative, $p = 2$ developmental, and 6 reconstructive components. We suggest that for each discriminative or developmental component, a reconstructive component be considered, hence the number of reconstructive components is set equal to the total of other components. The iterative procedure of Eq. (13) was performed which yielded components shown in Fig. 7. The connectivity component sub-networks obtained were sparse and thresholded for binary visualization to show the dominant connections at each component.

The vectorized connectivity of each subject was projected into the discriminative and developmental bases to obtain the coefficients and subsequently perform statistics for group difference and age correlation. To measure the discriminability of the components, we performed a t-test between the coefficients of ASD and TDC subjects for each sub-network component, separately. The coefficients were also correlated with age to determine the effect of development that the component can capture. In addition, the 83 subjects were divided into three closely-balanced age groups of 6–10 years (25 subjects), 10–13.5 years (28 subjects) and 13.5–18 years old (30 subjects). A *t*-test was performed between the coefficients of the subjects in the three age groups to assess the ability of the developmental components in capturing the effect of age. Results are given in Table 1.

Table 1 shows the results of group-wise and age-based statistics on the coefficients of discriminatory (**a**–**d**) and developmental (**e**) and (**f**) components. The statistical group difference between ASD and TDC shows that the discriminative basis (**a**) is able to

differentiate the two groups with a high statistical significance $p = 0.002$. Inspection of this component in Fig. 7 shows distinct connectivity deficiencies in the thalamic network, and inter-hemispheric subcortical connections in children with ASD. The low correlation with age of this component (column 5, Table 1) is indicative of the fact that this component concentrates on the pathology-related patterns in the data. The developmental basis (**e**) does not show any group related differences, but has a high positive correlation with age. It demonstrates significant age increase ($p = 0.004$) between the second and third age ranges (age > 10). Its positive correlation with age suggests that the connections between (mainly) left frontal and its nearby frontal, temporal, and sub-cortical regions significantly develop with age (correlation = +0.55), likely capturing ongoing maturation of language and executive functioning. The second developmental basis (**f**) shows an independent sub-network whose weight decreases (based on its negative correlation = −0.54 with age) significantly with development, especially in first two age groups (age < 13.5, $p = 0.0002$). The behavior of the two components (**e**) and (**f**) are opposite with respect to age, as indicated by the sign of the correlation. Of interest is the discriminative component (**d**), that shows the second highest group difference, with several frontal regions compromised. However, this component also shows a relatively high correlation with age. The analysis of these components shows that our method is able to extract components that capture the changes due to pathology (**a**) and age (**e**); and there are components such as (**d**) and (**f**) that are representative of changes in both. This may be an indication of those aspects of pathology that are linked with age and cannot be completely separated, yet help in providing a comprehensive picture of the pattern of changes in the population.

The components are desired to be orthogonal between discriminative and developmental sub-networks for interpretability. In order to investigate the orthogonality of the discriminative and developmental sub-networks, we computed the inner product $\boldsymbol{w}_i^T \boldsymbol{w}_j$ between any one of discriminative and developmental components. Since they are normalized vectors as stated in Section 3.4, the inner product between them measures the non-orthogonality on a scale of 0 (orthogonal) to 1 (absolute non-orthogonal). Table 2 shows the inner product between pairs of discriminative and developmental components, confirming their orthogonality.

5. Discussion

We have proposed a new framework capable of capturing the heterogeneity of the population important for group-discriminative and regressional analyses. In this paper, we investigated brain structural connectivity in autism spectrum disorder (ASD). ASDs are a category of *neurobiological developmental* disorders characterized by social and communication impairments, as well as repetitive and restricted behaviors (APA, 1994, 2000). Research suggests that many ASD symptoms are associated with aberrant structural and functional brain connectivity (Vissers et al., 2012). In our analysis, we found discriminative sub-networks that differentiate autistic from typically developing brains. We also used age to find regressive sub-networks that characterize developmental aspects of the population as an important factor in ASD. The reconstructive sub-networks compensate for

what is not captured by the discriminative and developmental sub-network bases in reconstructing the original connectomes.

Fig. 7 displays the sets of discriminative, developmental, and reconstructive connectivity sub-networks obtained from 59 TDC subjects and 24 individuals having ASD. It is observed in this figure that the discriminatory and developmental components are quite sparse with localized patterns as expected, while the reconstructive network components ensure reconstructability of the original connectivity networks. The discriminative component (**a**) in Fig. 7 shows diminished inter-hemispheric connectivity in ASD ($p = 0.002$ as shown in first column of Table 1), mainly in the connections between sub-cortical regions (as shown in Fig. 7a). This is in line with many other studies in ASD suggesting that ASD brains suffer from reduced structural connectivity and altered diffusion measures like FA (Vissers et al., 2012). Several findings report reduced FA of white matter in sub-cortical regions in ASD (Catani et al., 2008; Noriuchi et al., 2010). In addition, recent findings demonstrate inter-hemispheric under-connectivity in adolescents with autism (Thomas et al. 2011; Lo et al., 2011). Also, the corpus callosum, as the largest tract bridging the two hemispheres, has shown reduced FA in ASD (Alexander et al., 2007; Jou et al., 2011; Aoki et al., 2013), indicating decreased inter-hemispheric connectivity in autism.

In addition, our experiments have shown that the reconstructive components do not show any significant age-group differences or age correlation, but two mild significances (components **h** and **j**, $0.01 < p < 0.05$) in ASD-TDC group difference; while interesting, for the purposes of the paper we have concentrated on the discriminative and developmental components. Also, we have observed that the average of the reconstructive coefficients is an order of magnitude larger than the discriminative and developmental basis coefficients. Thus, due to their relatively small coefficients, the discriminatory and developmental bases do not play a significant role in the reconstruction.

Our method is independent of the technique that is used to create the connectome. Streamline (Mori and Barker, 1999; Mori et al., 1999; Mori and Zijl, 2002) and probabilistic (Behrens et al., 2003a,b; Friman et al., 2006) tractography are among the many methods introduced in the literature for quantifying brain connectivity. It is debatable which technique is appropriate in more accurately measuring the brain anatomical connectivity in DTI (Descoteaux et al., 2009), and how DTI fiber tracts should be interpreted (Assaf and Pasternak, 2008; Jones et al., 2013). Although we used probabilistic tractography in this work, the proposed method here does not depend on it, and any method of tractography may be used based on user's choice, to form connectivity matrices to serve an input to our method.

An important advantage of this technique is the fact that, as opposed to ICA, the components obtained from this method can directly be interpreted as a brain structural sub-network and their corresponding coefficients as the weights of those sub-networks in each subject. ICA was originally developed to solve problems similar to the "cocktail party" problem where the speech signal of each individual was to be extracted from the mixed signals from multiple speakers, recorded by a microphone (Bell and Sejnowski, 1995). ICA has since then been successfully applied to medical imaging problems in studies of pathology, e.g. in

autism (Assaf et al., 2010) or ADHD (Hoekzema et al., 2013), mainly to extract the group-wise brain functional sub-networks such as default mode network (DMN) (Calhoun and Adali, 2012) in resting state fMRI (rs-fMRI) (Calhoun et al., 2008) or cognitive tasks (Calhoun et al., 2001), and in EEG (Chen et al., 2013). Moreover, there have been applications of ICA in recovering multiple tensors in voxels where crossing fibers occur (Singh et al., 2006), in extraction of eddy-current-induced components (Arfanakis et al., 2002), and in parcellation of tractograms of the thalamus (O'Muircheartaigh et al., 2011). Although application of PCA and ICA for finding sub-networks is computationally possible, the computed components are not meaningfully interpretable in terms of DTI structural connectomes. This is due to the possibility of obtaining negative components and coefficients which is not possible as negative connectivity is meaningless in DTI and so is a coefficient being negative, as that would indicate a negative contribution to the anatomical substrate. Disregarding positivity constraints may even lead to identification of inaccurate sub-networks, as we showed in the simulation experiment of Fig. 4.

It is notable that non-negativity constraints are sufficient for interpretability but not enough for having components that characterize discriminatory or developmental aspects of a population. For example, solving Eq. (2) by minimizing the reconstruction error yields non-negative components and coefficients that provide reconstructability but not necessarily discriminative and developmental components. To show this, we have computed a total of $r = 12$ sub-network components with the objective function lacking graph embedding terms (i.e. solving only $\boldsymbol{F}_1(\boldsymbol{W})$ as in Eq. (2), based on the algorithm explained in (Ghanbari et al., 2012)), and the results are shown in Fig. 8. It is seen that those discriminative and developmental components obtained with graph embedding (shown in Fig. 7) are not present in the set obtained with no graph embedding (shown in Fig. 8). Moreover, group-wise *t*-test between ASD-TDC coefficients of the components in Fig. 8 showed no significance except for component (**f**) with $p = 0.045$, and their correlations with age ranged between −0.37 and +0.24. This clearly demonstrates the advantage of graph embedding proposed in this work that has enabled the extraction of discriminatory and developmental sub-networks from the connectome of the ASD-TDC populations and a comprehensive capture of the pathology and age specific changes, which were the two major sources of variation in this population.

The number of components is population/application specific. While this continues to be a topic for future investigations, several methods are available in the literature to determine the optimal number of components, among which singular value decomposition (Hansen et al., 1999), minimum description length (Schöpf et al., 2011), Akaike and Bayasian information (Hui et al., 2011) are the most widely used techniques for such purposes.

As mentioned earlier, the coefficients resulting from projecting the connectivity matrices into the components are amenable to statistics. However, this requires the need to deal with the problem of multiple comparisons. Several techniques are available for multiple comparison correction including Bonferroni correction (Dunn, 1961), FDR procedure (Benjamini and Hochberg, 1995; Genovese et al., 2002), and extreme statistics (Blair and Karniski, 1993), yet analysis of brain connectivity with multiple comparison correction is challenging (Cheol et al., 2013). The advantage of the proposed method is its capability in alleviating the issue of multiple comparisons owing to the dimensionality reduction

obtained, leading to a considerable increase in the power of statistical significance while maintaining interpretability. For instance, in the connectivity dataset used here, the edge-wise statistics of the 79 × 79 symmetric connectivity matrices require 3081 comparisons while our dimensionality reduction led to 12 statistical comparisons.

This method proposed here is generalizable to any non-negative connectivity network in functional or structural connectivity analysis. We showed the feasibility of our method by applying it to diffusion-based structural connectivity matrices that are positive. Functional connectivity of brain is defined as the coherence of neural activation patterns between brain regions (Li et al., 2009; Martijn and Hilleke, 2010; Sakkalis, 2011). Functional connectivity in fMRI, EEG, and MEG is measured using model-based, such as cross-correlation (Cao and Worsley, 1999), coherence (Sun et al., 2004), and synchronization (Fell and Axmacher, 2011; Ghanbari et al., 2013), or data-driven-based methods, such as ICA (McKeown et al., 1998) and fuzzy clustering (Golay et al., 1998). The proposed method is capable of handling functional connectivity when positive measures, such as synchronization likelihood (Barttfeld et al., 2011), are employed for creating the connectome.

We have demonstrated the applicability of the proposed method in a study of autism via the interpretable dimensionality reduction of structural brain networks. Future work includes validation with other datasets, and its application for meaningful feature selection in group classification studies and the regression of continuous measures such as age or clinical measures (i.e. phenotypic scores such as SRS and executive functioning), and in the creation of biomarkers of a disease.

6. Conclusions

We have presented a novel technique for simultaneously extracting the discriminatory and regressional sub-networks of a population via graph embedding in NMF, enabling different kinds of statistical analyses. This also maps the connectivity patterns of the population onto a lower dimensional space to ease subsequent population statistics. Our method consists of a projective NMF basis learning scheme with locality preserving properties, and provides group-discriminatory as well as developmental network components. Application to a dataset of ASD subjects provided a discriminatory basis which revealed significant inter-hemisphere sub-cortical connectivity deficiencies. The developmental bases captured sub-networks which changed with age. The framework is generalizable to non-negative functional networks, as well as to modeling and identifying other forms of variation in the population.

Acknowledgments

This research was supported by the following grants from National Institutes of Health: MH092862 and MH098010 (PI: Ragini Verma), and grants from the Pennsylvania Department of Health: SAP # 4100042728 and SAP # 4100047863 (PI: Robert T. Schultz).

References

Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, Miller JN, Lu J, Jeong EK, McMahon WM, Bigler ED, Lainhart JE. Diffusion tensor imaging of the corpus callosum in Autism. Neuroimage. 2007; 34(1):61–73. [PubMed: 17023185]

Aljabar, P.; Wolz, R.; Srinivasan, L.; Counsell, S.; Boardman, JP.; Murgasova, M.; Doria, V.; Rutherford, MA.; Edwards, AD.; Hajnal, JV.; Rueckert, D. Combining Morphological Information in a Manifold Learning Framework: Application to Neonatal MRI. Medical Image Computing and Computer-Assisted Intervention – MICCAI; 2010; Beijing, China. 2010.

Aoki Y, Abe O, Nippashi Y, Yamasue H. Comparison of white matter integrity between autism spectrum disorder subjects and typically developing individuals: a meta-analysis of diffusion tensor imaging tractography studies. Molecular Autism. 2013; 4:25. [PubMed: 23876131]

APA. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. American Psychiatric Press; Washington, DC: 1994.

APA. DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders. 4. Washington, DC: American Psychiatric Association; 2000. text revision

Arfanakis K, Cordes D, Haughton VM, Carew JD, Meyerand EM. Independent component analysis applied to diffusion tensor MRI. Magn Reson Med. 2002; 47(2):354–363. [PubMed: 11810680]

Assaf Y, Pasternak O. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. J Mol Neurosci. 2008; 34(1):51–61. [PubMed: 18157658]

Assaf M, Jagannathan K, Calhoun VD, Miller L, Stevens MC, Sahl R, O'Boyle JG, Schultz RT, Pearlson GD. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. Neuroimage. 2010; 53(1):247–256. [PubMed: 20621638]

Barttfeld P, Wicker B, Cukier S, Navarta S, Lew S, Sigman M. A big-world network in ASD: dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. Neuropsychologia. 2011; 49(2):254–263. [PubMed: 21110988]

Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson, Ser B. 1996; 111:209–219. [PubMed: 8661285]

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. In vivo fiber tractography using DT-MRI data. Magn Reson Med. 2000; 44(4):625–632. [PubMed: 11025519]

Bassett DS, Brown JA, Deshpande V, Carlson JM, Grafton ST. Conserved and variable architecture of human white matter connectivity. Neuroimage. 2011; 54(2):1262–1279. [PubMed: 20850551]

Batmanghelich NK, Taskar B, Davatzikos C. Generative-discriminative basis learning for medical imaging. IEEE Trans Med Imag. 2012; 31(1):51–69.

Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci. 2003a; 6(7):750–757. [PubMed: 12808459]

Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med. 2003b; 50(5):1077–1088. [PubMed: 14587019]

Belkin, M.; Niyogi, P. Advances in Neural Information Processing Systems. Vol. 14. MIT Press; Cambridge, MA: 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering.

Bell AJ, Sejnowski TJ. An information maximization approach to blind separation and blind deconvolution. Neural Comput. 1995; 7:1129–1159. [PubMed: 7584893]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B Stat Methodol. 1995; 2:289–300.

Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ. Algorithms and applications for approximate nonnegative matrix factorization. Comput Stat. 2007; 52:155–173.

Bihan DL. Looking into the functional architecture of the brain with diffusion MRI. Nat Rev Neurosci. 2003; 4(6):469–480. [PubMed: 12778119]

Bihan DL, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology. 1986

Bihan DL, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: concepts and applications. J Magn Reson Imag. 2001; 13(4):534–546.

Blair RC, Karniski W. An alternative method for significance testing of waveform difference potentials. Psychophysiology. 1993; 30:518–524. [PubMed: 8416078]

Bozzali M, Parker GJM, Serra L, Embleton K, Gili T, Perri R, Caltagirone C, Cercignani MM. Anatomical connectivity mapping: a new tool to assess brain disconnection in Alzheimer's disease. NeuroImage. 2011; 54(3):2045–2051. [PubMed: 20828625]

Bullmore, EaOS. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci. 2009; 10(3):186–198. [PubMed: 19190637]

Calhoun VD, Adali T. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev Biomed Eng. 2012; 5:60–73. [PubMed: 23231989]

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Hum Brain Mapp. 2001; 14:140–151. [PubMed: 11559959]

Calhoun V, Kiehl K, Pearlson G. Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks. Hum Brain Mapp. 2008; 29(7):828–838. [PubMed: 18438867]

Cao J, Worsley KJ. The geometry of correlation fields, with an application to functional connectivity of the brain. Ann Appl Probab. 1999; 9:1021–1057.

Catani M, Jones DK, Daly E, Embiricos N, Deeley Q, Pugliese L, Curran S, Robertson D, Murphy DGM. Altered cerebellar feedback projections in Asperger syndrome. Neuroimage. 2008; 41(4): 1184–1191. [PubMed: 18495494]

Chen JL, Ros T, Gruzelier JH. Dynamic changes of ICA-derived EEG functional connectivity in the resting state. Hum Brain Mapp. 2013; 34(4):852–868. [PubMed: 22344782]

Cheol EH, Sang WY, Sang WS, Duk LN, Joon-Kyung S. Cluster-based statistics for brain connectivity in correlation with behavioral measures. PLoS one. 2013; 8(8):e72332. [PubMed: 23977281]

Clayden JD. Imaging connectivity: MRI and the structural networks of the brain. Funct Neurol. 2013; 28(3):197–203. [PubMed: 24139656]

Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp. 2012; 33(8):1914–1928. [PubMed: 21769991]

Dennis EL, Jahanshad N, McMahon KL, de Zubicaray GI, Martin NG, Hickie IB, Toga AW, Wright MJ, Thompson PM. Development of brain structural connectivity between ages 12 and 30: A 4-Tesla diffusion imaging study in 439 adolescents and adults. NeuroImage. 2013; 64:671–684. [PubMed: 22982357]

Descoteaux M, Deriche R, Knösche TR, Anwander A. Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans Med Imag. 2009; 28(2):269–286.

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006; 31(3):968–980. [PubMed: 16530430]

Dunn OJ. Multiple comparisons among means. J Am Stat Assoc. 1961; 56:52–64.

Fell J, Axmacher N. The role of phase synchronization in memory processes. Nat Rev Neurosci. 2011; 12:105–118. [PubMed: 21248789]

Friman O, Farneback G, Westin CF. A Bayesian approach for stochastic white matter tractography. IEEE Trans Med Imag. 2006; 25(8):965–978.

Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002; 15(4):870–878. [PubMed: 11906227]

Ghanbari, Y.; Bloy, L.; Batmanghelich, K.; Roberts, TPL.; Verma, R. In: Ayache, N.; Delingette, H.; Golland, P.; Mori, K., editors. Dominant component analysis of electrophysiological connectivity

networks; International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2012), Part III. LNCS; Nice, France. Heidelberg: Springer; 2012. p. 231-238.

Ghanbari Y, Bloy L, Edgar JC, Blaskey L, Verma R, Roberts TPL. Joint analysis of band-specific functional connectivity and signal complexity in autism. J Autism Develop Disord. 2013; 3:2.

Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn Reson Med. 1998; 40:249–260. [PubMed: 9702707]

Hansen LK, Larsen J, Nielsen FA, Strother SC, Rostrup E, Savoy R, Lange N, Sidtis J, Svarer C, Paulson O. Generalizable patterns in neuroimaging: how many principal components? NeuroImage. 1999; 9(5):534–544. [PubMed: 10329293]

He, X.; Niyogi, P. Advances in Neural Information Processing Systems (NIPS). MIT Press; Cambridge: 2003. Locality Preserving Projections.

Hoekzema E, Carmona S, Ramos-Quiroga JA, Richarte Fernández V, Bosch R, Soliva JC, Rovira M, Bulbena A, Tobeña A, Casas M, Vilarroya O. An independent components and functional connectivity analysis of resting state fMRI data points to neural network dysregulation in adult ADHD. Hum Brain Mapp. 2013 Epub ahead of print. 10.1002/hbm.22250

Hui M, Li J, Wen X, Yao L, Long Z. An empirical comparison of information-theoretic criteria in estimating the number of independent components of fMRI data. PLoS one. 2011; 6:e29274. [PubMed: 22216229]

Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw. 2000; 13(4–5):411–430. [PubMed: 10946390]

Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, Hakonarson H, Gur RE, Gur RC, Verma R. Sex differences in the structural connectome of the human brain. Proceedings of the National Academy of Sciences of the United States of America (PNAS). 2013 Published online.

Jones DK, Knösche TR, Turner R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. NeuroImage. 2013; 73:239–254. [PubMed: 22846632]

Jou RJ, Jackowski AP, Papademetris X, Rajeevan N, Staib LH, Volkmar FR. Diffusion tensor imaging in autism spectrum disorders: preliminary evidence of abnormal neural connectivity. J Psychiat. 2011; 45(2):153–162.

Lanyon, LJ. Neuroimaging – Methods, InTech, Chapter Diffusion Tensor Imaging: Structural Connectivity Insights, Limitations and Future Directions. 2012. p. 137-162.

Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999; 401:788–791. [PubMed: 10548103]

Li K, Guo L, Nie J, Li G, Liu T. Review of methods for functional brain connectivity detection using fMRI. Comput Med Imag Graph. 2009; 33(2):131–139.

Liu X, Tosun D, Weiner MW, Schuff N. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. NeuroImage. 2013; 83:148–157. [PubMed: 23792982]

Lo YC, Soong WT, Gau SS, Wu YY, Lai MC, Yeh FC, Chiang WY, Kuo LW, Jaw FS, Tseng WY. The loss of asymmetry and reduced interhemispheric connectivity in adolescents with autism: a study using diffusion spectrum imaging tractography. Psychiatry Res. 2011; 192:60–66. [PubMed: 21377337]

Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Develop Disord. 1994; 24(5):659–685.

Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Develop Disord. 2000; 30(3):205–223.

Manor, LZ.; Perona, P. Advances in Neural Information Processing Systems (NIPS, 2004). MIT Press; Cambridge, MA: 2004. Self-Tuning Spectral Clustering.

Martijn, PvH; Hilleke, EHP. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur Neuropsychopharmacol. 2010; 20(8):519–534. [PubMed: 20471808]

Matthews PM, Filippini N, Douaud G. Brain structural and functional connectivity and the progression of neuropathology in Alzheimer's disease. J Alzheimer's Dis. 2013; 33:S163–S172. [PubMed: 22669012]

McKeown M, Makeig S, Liddle PF, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. Hum Brain Mapp. 1998; 6:160–188. [PubMed: 9673671]

Mori S, Barker PB. Diffusion magnetic resonance imaging: its principle and applications. Anat Rec. 1999; 257(3):102–109. [PubMed: 10397783]

Mori S, van Zijl PC. Fiber tracking: principles and strategies – a technical review. NMR Biomed. 2002; 15(7–8):468–480. [PubMed: 12489096]

Mori S, Zijl PCM. Fiber tracking: principles and strategies – a technical review. NMR Biomed. 2002; 15(7–8):468–480. [PubMed: 12489096]

Mori S, Crain BJ, Chacko VP, van Zijl PC. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neurol. 1999; 45(2):265–269. [PubMed: 9989633]

Ng AY, Jordan MI, Weiss Y. On spectral clustering: analysis and an algorithm. Adv Neural Inform Process Syst. 2002; 14:849–856.

Noriuchi M, Kikuchi Y, Yoshiura T, Kira R, Shigeto H, Hara T, Tobimatsu S, Kamio Y. Altered white matter fractional anisotropy and social impairment in children with autism spectrum disorder. Brain Res. 2010; 1362:141–149. [PubMed: 20858472]

O'Muircheartaigh J, Vollmar C, Traynor C, Barker GJ, Kumari V, Symms MR, Thompson P, Duncan JS, Koepp MJ, Richardson MP. Clustering probabilistic tractograms using independent component analysis applied to the thalamus. NeuroImage. 2011; 54:2020–2032. [PubMed: 20884353]

Paul LK, Brown WS, Adolphs R, Tyszka JM, Richards LJ, Mukherjee P, Sherr EH. Agenesis of the corpus callosum: genetic, developmental and functional aspects of connectivity. Nat Rev Neurosci. 2007; 8(4):287–299. [PubMed: 17375041]

Price G, Cercignani M, Parker GJ, Altmann DR, Barnes TR, Barker GJ, Joyce EM, Ron MA. Abnormal brain connectivity in first-episode psychosis: a diffusion MRI tractography study of the corpus callosum. Neuroimage. 2007; 35(2):458–466. [PubMed: 17275337]

Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000; 290(5500):2323–2326. [PubMed: 11125150]

Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010; 52(3):1059–1069. [PubMed: 19819337]

Sakkalis V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Comput Biol Med. 2011; 41(12):1110–1117. [PubMed: 21794851]

Salvador R, Anguera M, Gomar JJ, Bullmore ET, Pomarol-Clotet E. Conditional mutual information maps as descriptors of net connectivity levels in the brain. Front Neuroinformatics. 2010; 4:111–118. Article 115.

Saul, LK.; Roweis, ST. Technical Report. AT&T Labs and Gatsby Computational Neuroscience Unit; 2000. An Introduction to Locally Linear Embedding.

Schöpf V, Windischberger C, Robinson S, Kasess CH, Fischmeister FP, Lanzenberger R, Albrecht J, Kleemann AM, Kopietz R, Wiesmann M, Moser E. Model-free fMRI group analysis using FENICA. NeuroImage. 2011; 55(1):185–193. [PubMed: 21078400]

Singh, M.; Wong, CW.; Jeong, JW. Simulation and experimental study of multiple-fibers per voxel detection by ICA in DTI tractography. Nuclear Science Symposium Conference Record; 2006. p. 3176-3179.

Skudlarski P, Jagannathan K, Anderson K, Stevens MC, Calhoun VD, Skudlarska BA, Pearlson G. Brain Connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. Biol Psychiat. 2010; 68(1):61–69. [PubMed: 20497901]

Sporns O. The Human Connectome: A Complex Network. Ann NY Acad Sci. 2011

Sun FT, Miller LM, D'Esposito M. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. NeuroImage. 2004; 21(2):647–658. [PubMed: 14980567]

Thomas C, Humphreys K, Jung KJ, Minshew N, Behrmann M. The anatomy of the callosal and visual-association pathways in high-functioning autism: A DTI tractography study. Cortex. 2011; 47(7): 863–873. [PubMed: 20832784]

van den Maaten L, Postma E, van den Herik J. Dimensionality reduction: a comparative review. J Mach Learn Res. 2009; 10:1–41.

Vissers ME, Cohen MX, Geurts HM. Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. Neurosci Biobehav Rev. 2012; 36(1):604–625. [PubMed: 21963441]

Wang C, Yan S, Zhang L, Zhang H. Non-negative semi-supervised learning. J Mach Learn Res. 2009; 5:575–582.

Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S. Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell. 2007; 29(1):40–51. [PubMed: 17108382]

Yang Z, Oja E. Linear and nonlinear projective nonnegative matrix factorization. IEEE Trans Neural Netw. 2010; 21(5):1734–1749.

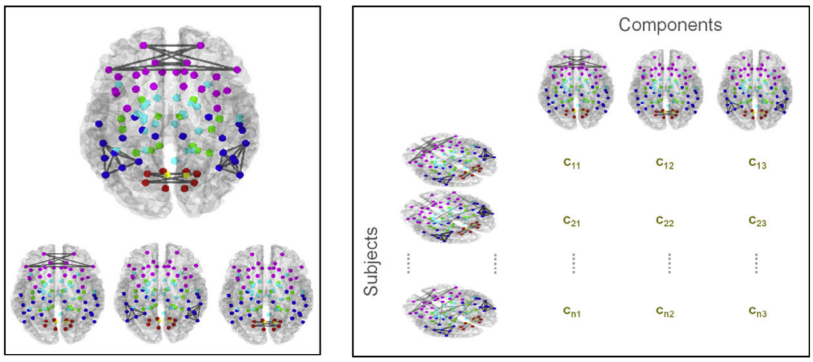

Fig. 1.

Illustration of connectivity network decomposition into connectivity components. (a) A sample brain connectivity (top) composed of frontal (bottom left), temporal (bottom middle), and occipital (bottom right) sub-network modules. (b) A set of connectivity networks from *n* subjects (left) with three main components (top). The weight of each component in each subject is quantified by the corresponding coefficient c_{ij}.

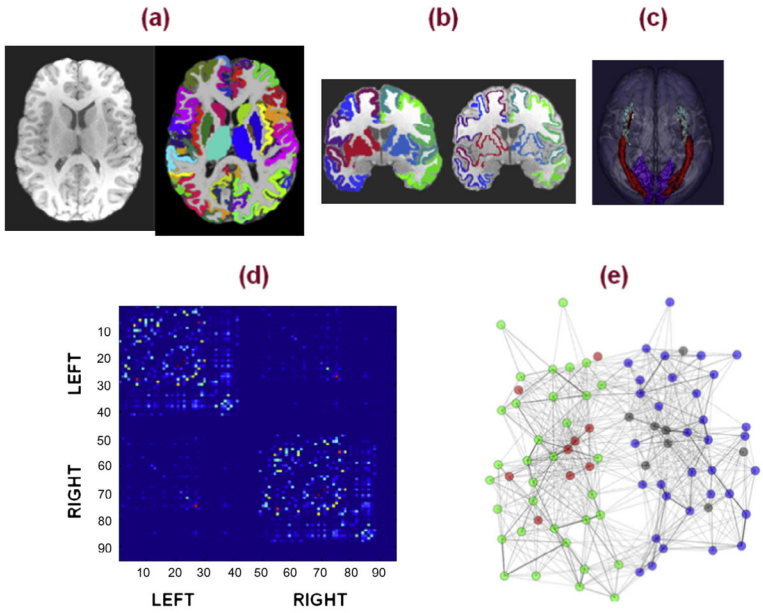

Fig. 2.

Creating the connectome: (a) Parcellation of T1 images into 95 cortical and sub-cortical regions. (b) Transfer of region labels to diffusion space and computing the GMWM boundary. (c) Probabilistic fiber tracking from each seed ROI *i* to target ROI *j*. (d) Connectivity quantification between each ROI pair (*i*, *j*) computed from conditional probability of a pathway between two regions. (e) Construction of a weighted structural connectivity network.

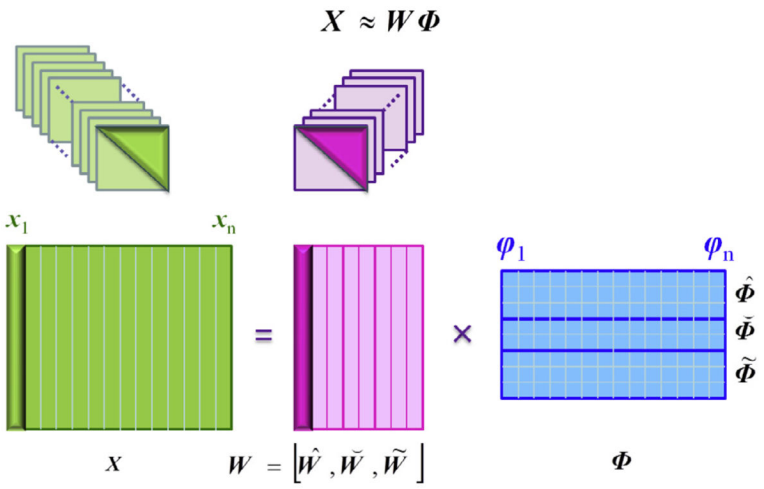

Fig. 3.

The process of matrix decomposition of a set of connectivity matrix whose upper triangular elements are stacked into the matrix *X*. Each of the connectivity sub-network components are represented by the columns of the matrix *W* which is reconstructed back into a connectivity component matrix by reordering those elements into the upper triangular elements of their matrix. The three subsets of *W* represent the discriminative, developmental, and reconstructive components.

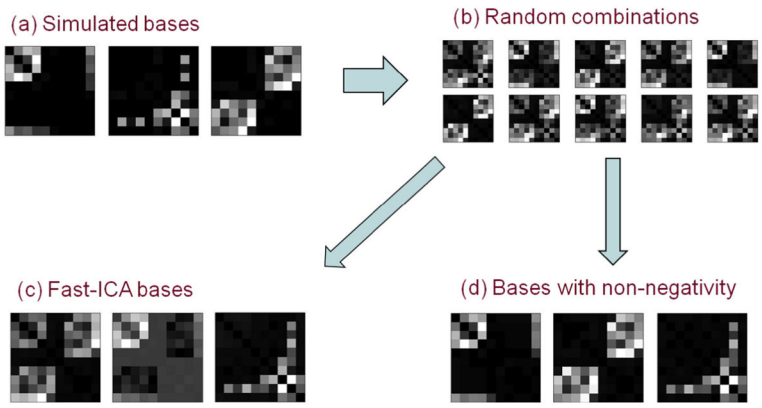

Fig. 4.

The effect of non-negativity constraints in identifying the components are shown using simulated symmetric connectivity matrices visualized by gray-scale images. (a) The three simulated basis components. (b) Ten linear combinations of the three components by random positive coefficients. (c) The components obtained from the fast-ICA algorithm. (d) The components obtained from the projective non-negative basis learning method described in Section 3.1.

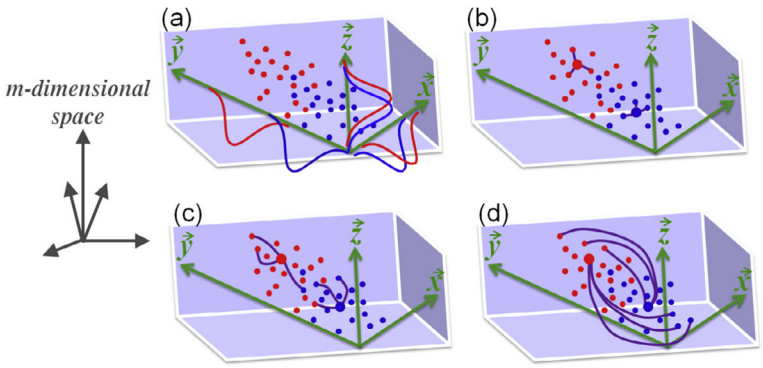

Fig. 5.

Illustration of a two-group *m*-dimensional points lying on a 3-D manifold, shown as a cube in the *m*-dimensional space. (a) Point distributions when projected into $\boldsymbol{x}, \vec{\boldsymbol{y}}, \vec{\text{or}}\ \boldsymbol{z}\ \vec{\text{v}}$ectors. (b) The 3-nearest-neighbor graph $\hat{G}$ of two selected magnified points. (c) The graph $\check{\boldsymbol{G}}$ with edges connecting points whose subjects are of similar ages. (d) The 3-farthest-point graph $\tilde{G}$ of the same two points.

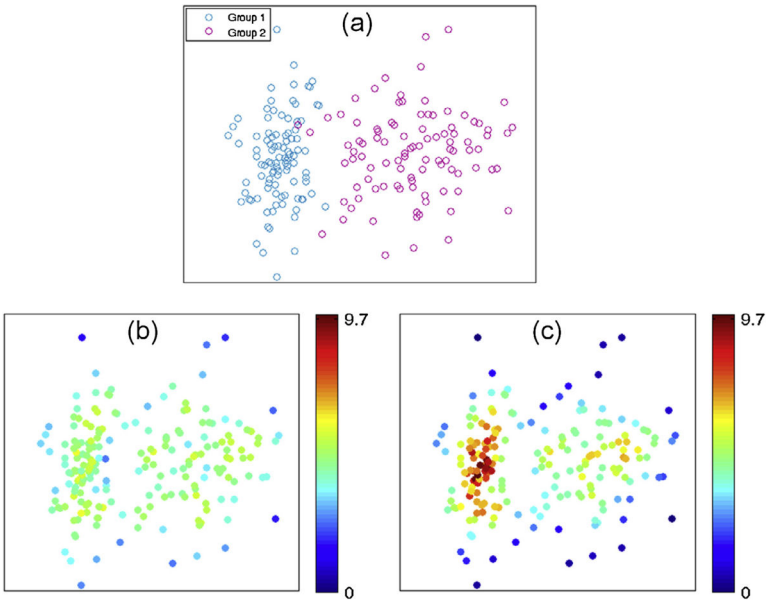

Fig. 6.

The scaling parameter effect on neighborhood graph construction. (a) Two synthetic clusters composed of normally distributed points in the 2-D space. (b) The node strength (elements of $\hat{D_{ii}}$) of the graph constructed using the scaling parameter given in Eq. (14). (c) The node strength (elements of $\hat{D_{ii}}$) of the graph constructed using our proposed scaling parameter in Eq. (15). It is clear that our method provides stronger connection around the center of the distributions as well as weaker connection for the outliers.

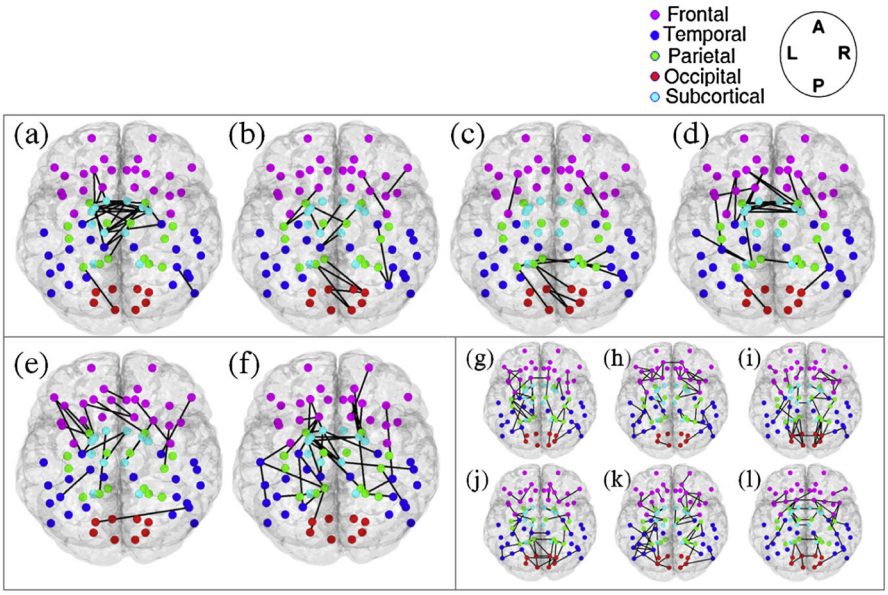

Fig. 7.

The $r = 12$ connectivity sub-networks obtained from the population of 24 ASD and 59 TDC connectivity matrices. (a)–(d) display the $q = 4$ discriminative sub-network components. (e and f) show the sub-network of $p = 2$ developmental components. (g)–(l) are the set of 6 reconstructive sub-networks.

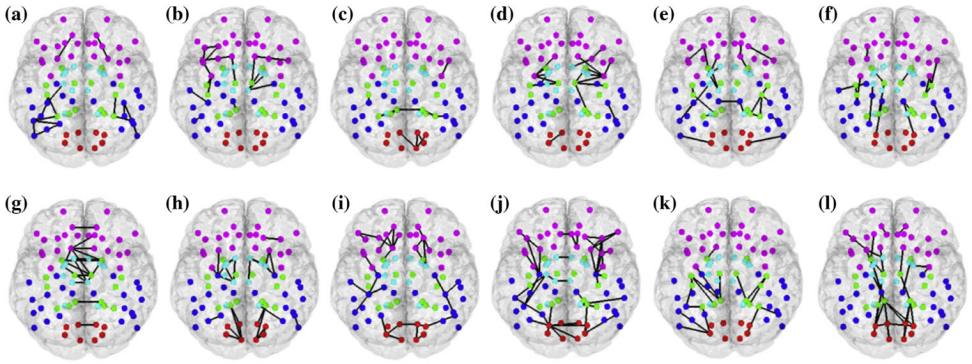

Fig. 8.

Twelve connectivity sub-networks obtained from solving the minimization of Eq. (2) based on the algorithm explained in (Ghanbari et al., 2012). These results do not involve the graph embedding term, and were obtained by only solving for projective NMF.

Table 1

The statistical analysis of the components' coefficients including patient-control group difference, age group difference, and age correlations. The components' IDs are according to those given in Fig. 7.

Component ID	ASD-TDC group difference $t(p)$-value	Age group II vs. age group I $t(p)$-value	Age group III vs. age group II $t(p)$-value	Coefficients' correlation with age
(**a**)	−3.2 (0.002)	−1.6 (0.1)	−0.7 (0.5)	−0.24
(**b**)	−1.2 (0.2)	−0.6 (0.6)	−0.3 (0.8)	−0.14
(**c**)	+0.5 (0.6)	−1.5 (0.1)	−1.3 (0.2)	−0.32
(**d**)	−2.3 (0.02)	−2.1 (0.04)	−2.3 (0.02)	−0.48
(**e**)	+0.7 (0.5)	+1.9 (0.06)	+3.0 (0.004)	+0.55
(**f**)	−3.1 (0.004)	−4.0 (0.0002)	−1.7 (0.09)	−0.54

Table 2

The inner product measure between discriminative (**a**)–(**d**) and developmental (**e**) and (**f**) components as well as the angle between the component vectors given in the parentheses. The inner product is a measure of non-orthogonality that is between 0 (orthogonal vectors with 90-degree angle) and 1 (absolute non-orthogonal with zerodegree angle). The components' IDs are according to Fig. 7.

Component ID	(e)	(f)
(**a**)	0.00 (89.9°)	0.01 (89.2°)
(**b**)	0.04 (87.8°)	0.05 (87.1°)
(**c**)	0.02 (88.9°)	0.01 (89.7°)
(**d**)	0.01 (89.5°)	0.03 (88.0°)