Connectivity Subnetwork Learning for Pathology and Developmental Variations

Yasser Ghanbari[1], Alex R. Smith[1], Robert T. Schultz[2], and Ragini Verma[1,⋆]

[1] Section of Biomedical Image Analysis, University of Pennsylvania, Philadelphia, PA
{Yasser.Ghanbari,Alex.Smith,Ragini.Verma}@uphs.upenn.edu
[2] Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, PA
schultzrt@mail.chop.edu

Abstract. 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 networks demands methods that are not only able to extract the patterns that highlight these sources of variation, but describe them individually. In this paper, we present a unified framework for learning subnetwork patterns of connectivity by their projective non-negative decomposition into a reconstructive basis set, as well as, additional basis sets representing development and group discrimination. In order to obtain these components, we exploit the geometrical distribution of the population in the connectivity space by using a graph-theoretical scheme that imposes locality-preserving properties. In addition, the projection of the subject networks into the basis set provides a low dimensional representation of it, that teases apart the different sources of variation in the sample, facilitating variation-specific statistical analysis. The proposed framework is applied to a study of diffusion-based connectivity in subjects with autism.

Keywords: Connectivity analysis, non-negative matrix factorization, locality preserving projections, graph embedding, population difference.

1 Introduction

Analysis of brain connectivity networks created from DTI, MEG/EEG, or fMRI data has provided a novel insight into brain changes arising from pathology, development or aging [1, 2]. The high dimensionality of these networks has necessitated the development of methods that can extract the connectivity patterns of the population and separate components pertaining to each source of variation. In this paper, we provide such a method of extracting the basis of patterns using a locality-preserving basis learning, which helps provide a low dimensional representation of the subject networks that can then be used to study group differences, or train pathology specific or developmental biomarkers. Traditional analysis techniques, such as PCA and ICA, provide dimensionality reduction in

⋆ Authors acknowledge support from grants NIH-MH092862 (PI: R. Verma), Pennsylvania Department of Health (SAP#4100042728, SAP#4100047863, PI: R. Schultz).

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 90–97, 2013.

brain network investigations [3] but may lack the physiological interpretability of non-negative connectivity matrices. Recently, non-negative matrix factorization (NMF) methods [4] have provided interpretable sparse bases characterizing multivariate data [5–8]. The non-negativity constraints, facilitate interpretability of the components as sparse connectivity matrices [6], but are unable to separate patterns that represent the different sources of variability in the population.

This paper presents a framework for learning sparse subnetwork patterns of non-negative connectivity matrices by their projective non-negative decomposition into sets of i) *discriminative* or pathology-specific, ii) *developmental* (age related), and iii) *reconstructive* components. The decomposition maintains the interpretation of each component as a connectivity matrix and their associated coefficients as the weight of the subnetwork, while providing a succinct low dimensional representation of the population amenable to statistical analysis.

While the method is generalizable to any type of non-negative connectivity matrix, we have demonstrated the applicability of the framework to DTI-based structural connectivity networks for a population of subjects with autism spectrum disorder (ASD). Our method is able to extract components that describe the underlying patterns of pathology related variability, as well as patterns of developmental (age) variation. In addition, the corresponding low dimensional representation of the subjects in these basis components is able to identify group differences in pathology, and quantify the age related variations.

2 Methods

The connectivity network of each subject can be modeled as a linear combination of several components that act as the building blocks of the brain network system [6]. Due to their symmetry, the upper triangular elements of a connectivity matrix are represented by $\boldsymbol{x}_i$ for the subject i. To compute the r connectivity components whose mixture constructs the original non-negative connectivity matrices, a matrix factorization is used as $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{\Phi}$, where columns of $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$, i.e. $\boldsymbol{x}_i$, represent the connectivity matrices, and columns of $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_r] \in \mathbb{R}^{m \times r}$, i.e. $\boldsymbol{w}_j$, are representative of the normalized basis connectivity components. These components $\boldsymbol{w}_j$ are then mixed by each column of the loading matrix $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_n] \in \mathbb{R}^{r \times n}$ to approximate the corresponding column of $\boldsymbol{X}$ [8].

2.1 Projective Non-negative Basis Learning

Inspired by [6], we assume that $\boldsymbol{\Phi}$ is the projection of $\boldsymbol{X}$ onto $\boldsymbol{W}$, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T\boldsymbol{X}$. The projective properties of the NMF coefficients ($\boldsymbol{\Phi}$) help imbibe orthogonality properties into components [8]. Hence, the reconstruction of original non-negative connectivity matrices can be obtained by minimizing the cost function $F_1(\boldsymbol{W}) = \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\|_F^2$ with respect to $\boldsymbol{W} \geq 0$. This can be denoted by

$$\min_{\boldsymbol{W} \geq 0} F_1(\boldsymbol{W}) = \min_{\boldsymbol{W} \geq 0} trace\left\{ \left(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\right) \left(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\right)^T \right\}. \quad (1)$$

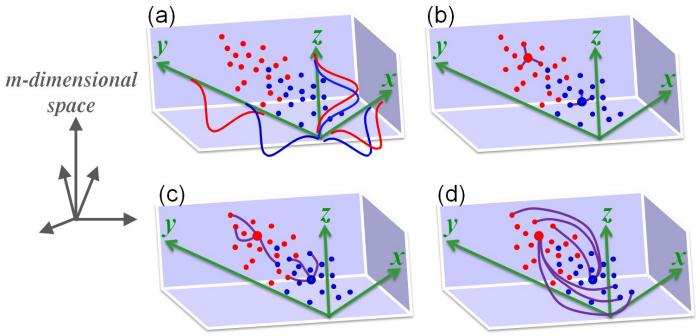

Fig. 1. Illustration of a two-group m-dimensional points lying on a 3-D manifold, shown as a cube in the m-dimensional space. (a) Point distributions when projected into $\vec{\boldsymbol{x}}$, $\vec{\boldsymbol{y}}$ or $\vec{\boldsymbol{z}}$. (b) The 3-nearest-neighbor graph $\hat{G}$ of two selected magnified points. (c) The graph $\breve{G}$ with edges connecting points whose subjects are of similar ages. (d) The 3-farthest-point graph $\tilde{G}$ of the same two points.

2.2 Locality Preserving Bases with Graph Embedding

We split the set of r components into three subsets $\boldsymbol{W} = [\hat{\boldsymbol{W}}, \breve{\boldsymbol{W}}, \tilde{\boldsymbol{W}}]$ where $\hat{\boldsymbol{W}} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q]$ are the *discriminative* basis components, $\breve{\boldsymbol{W}} = [\boldsymbol{w}_{q+1}, \ldots, \boldsymbol{w}_{q+p}]$ are the *developmental* basis components, and $\tilde{\boldsymbol{W}} = [\boldsymbol{w}_{q+p+1}, \ldots, \boldsymbol{w}_r]$ is the complementary space containing the *reconstructive* basis components which minimizes the reconstruction error together with $\hat{\boldsymbol{W}}$ and $\breve{\boldsymbol{W}}$. Thus, the coefficient matrices are also split into $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{W}}^T \boldsymbol{X}$, $\breve{\boldsymbol{\Phi}} = \breve{\boldsymbol{W}}^T \boldsymbol{X}$, 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 $\breve{\boldsymbol{W}}$.

To clarify our mathematical modeling, suppose that the m-dimensional connectivity points ($\boldsymbol{x}_i$) of two groups lie on a 3-D manifold, as illustrated in Fig. 1(a), and are to be projected into an $r = 3$ dimensional subspace with $q = 1$ discriminative (say $\vec{\boldsymbol{y}}$), $p = 1$ developmental (say $\vec{\boldsymbol{z}}$), and $r - q - p = 1$ reconstructive components (say $\vec{\boldsymbol{x}}$). In order to achieve this, we construct three separate graphs made up of these m-dimensional points as their vertices. The first graph is an intrinsic k-nearest-neighbor graph [5, 9] which connects point i to j if point j is among the k nearest neighbors of point i (as illustrated in Fig. 1(b)). This graph is used for obtaining the discriminative component $\hat{\boldsymbol{W}}$. The second graph connects point i to point j if point j is among the k farthest points to point i (illustrated in Fig. 1(d)), which is used for obtaining the reconstructive component $\tilde{\boldsymbol{W}}$. The third graph, used for obtaining the developmental component $\breve{\boldsymbol{W}}$, connects point i to j if subject j's age is among the k nearest ages to subject i, irrespective of the two points' distance (illustrated in Fig. 1(c)).

There are a variety of approaches that can characterize separability of multivariate data-points. Most of such techniques can be unified in the framework of graph embedding [5, 9]. Let $G = \{\boldsymbol{X}, \boldsymbol{S}\}$ be an undirected weighted graph

of n vertices, i.e. data points $\boldsymbol{x}_i$, with a symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ with non-negative elements, within the range of 0 to 1, corresponding to the edge weight of the graph. The Laplacian matrix $\boldsymbol{L}$ of the graph is then defined by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$ where $\boldsymbol{D}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} S_{ij}$ [9].

In order for the bases in $\hat{\boldsymbol{W}}$ to provide discriminatory information, we would like the resulting coefficients ($\hat{\boldsymbol{\varphi}}_i$) 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

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

where $\hat{\boldsymbol{S}} = [\hat{S}_{ij}]$ is the similarity matrix composed of the edge weights in the k-nearest-neighbor graph $\hat{G} = \{\boldsymbol{X}, \hat{\boldsymbol{S}}\}$ of the m-dimensional points $\boldsymbol{x}_i$, as illustrated in Fig. 1(b). According to the equation (2), if data-points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close, their edge weight S_{ij} will be large, and therefore, the cost function $F_2(\hat{\boldsymbol{W}})$ gets minimized only if the corresponding coefficients $\hat{\boldsymbol{\varphi}}_i$ and $\hat{\boldsymbol{\varphi}}_j$ remain close.

Similarly, in order to capture the connectivity space of developmental variations, we form a developmental graph $\breve{G} = \{\boldsymbol{X}, \breve{\boldsymbol{S}}\}$, as in Fig. 1(c), in which $\boldsymbol{x}_i$ is connected with an edge to $\boldsymbol{x}_j$ if subject i is within the k nearest age to subject j, and vice versa. Therefore, the space of developmental variations, $\breve{\boldsymbol{W}}$, grouping the coefficients ($\breve{\boldsymbol{\varphi}}_i$) of subjects with similar ages, is computed by minimizing $F_3(\boldsymbol{W}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \|\breve{\boldsymbol{\varphi}}_i - \breve{\boldsymbol{\varphi}}_j\|^2 \breve{S}_{ij} = trace\left\{\breve{\boldsymbol{\Phi}} \breve{\boldsymbol{L}} \breve{\boldsymbol{\Phi}}^T\right\}$ when $\breve{\boldsymbol{W}} \geq 0$.

As explained earlier, we exploit the graph of k-farthest points $\tilde{G} = \{\boldsymbol{X}, \tilde{\boldsymbol{S}}\}$ (as illustrated in Fig. 1(d)), to impose the representative coefficients ($\tilde{\boldsymbol{\varphi}}_i$) 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 $F_4(\boldsymbol{W}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \|\tilde{\boldsymbol{\varphi}}_i - \tilde{\boldsymbol{\varphi}}_j\|^2 \tilde{S}_{ij} = trace\left\{\tilde{\boldsymbol{\Phi}} \tilde{\boldsymbol{L}} \tilde{\boldsymbol{\Phi}}^T\right\}$ subject to $\tilde{\boldsymbol{W}} \geq 0$.

2.3 Objective Function and Optimization Solution

To achieve the above objectives, and according to the projective properties of the model, i.e. $\boldsymbol{\Phi} = \boldsymbol{W}^T \boldsymbol{X}$, the final objective function is modeled to minimize

$$F(\boldsymbol{W}) = trace\left\{\left(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\right)\left(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\right)^T\right\} +$$
$$\lambda\left(trace\left\{\hat{\boldsymbol{W}}^T \boldsymbol{X} \hat{\boldsymbol{L}} \boldsymbol{X}^T \hat{\boldsymbol{W}}\right\} + trace\left\{\breve{\boldsymbol{W}}^T \boldsymbol{X} \breve{\boldsymbol{L}} \boldsymbol{X}^T \breve{\boldsymbol{W}}\right\} + trace\left\{\tilde{\boldsymbol{W}}^T \boldsymbol{X} \tilde{\boldsymbol{L}} \boldsymbol{X}^T \tilde{\boldsymbol{W}}\right\}\right), \qquad (3)$$

where λ is a tunable parameter to balance the two terms of reconstruction error norm and graph embedding. To minimize (3) with $\boldsymbol{W} \geq 0$, we use a gradient descent approach, updating $W_{ij} = W_{ij} - \eta_{ij} \frac{\partial F}{\partial W_{ij}}$ with step-sizes $\eta_{ij} \geq 0$, where

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

Regarding that $\hat{\boldsymbol{L}} = \hat{\boldsymbol{D}} - \hat{\boldsymbol{S}}$, $\check{\boldsymbol{L}} = \check{\boldsymbol{D}} - \check{\boldsymbol{S}}$ 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 $\eta_{ij} = \frac{\frac{1}{2}W_{ij}}{\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}+\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}+\lambda\left[\boldsymbol{X}\hat{\boldsymbol{D}}\boldsymbol{X}^T\hat{\boldsymbol{W}},\boldsymbol{X}\check{\boldsymbol{D}}\boldsymbol{X}^T\check{\boldsymbol{W}},\boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}$. This results in the following multiplicative updating solution

$$W_{ij} = W_{ij}\frac{\left(2\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W} + \lambda\left[\boldsymbol{X}\hat{\boldsymbol{S}}\boldsymbol{X}^T\hat{\boldsymbol{W}}\ ,\ \boldsymbol{X}\check{\boldsymbol{S}}\boldsymbol{X}^T\check{\boldsymbol{W}},\ \boldsymbol{X}\tilde{\boldsymbol{S}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}{\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W} + \boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W} + \lambda\left[\boldsymbol{X}\hat{\boldsymbol{D}}\boldsymbol{X}^T\hat{\boldsymbol{W}}, \boldsymbol{X}\check{\boldsymbol{D}}\boldsymbol{X}^T\check{\boldsymbol{W}}, \boldsymbol{X}\tilde{\boldsymbol{D}}\boldsymbol{X}^T\tilde{\boldsymbol{W}}\right]\right)_{ij}}. \quad (5)$$

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 $\boldsymbol{W} = [\hat{\boldsymbol{W}}, \check{\boldsymbol{W}}, \tilde{\boldsymbol{W}}] \geq 0$.

3 Results

The proposed method provides a framework for extracting three sets of network components from the population. 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. The reconstructive basis set consists of global networks of dominant connectivity patterns. The number of bases is population dependent; however, we show that even with relatively small numbers, we can obtain stable group differences.

Dataset Demographics and Connectivity Measures. We study the applicability of our method on a dataset of DTI connectivity matrices computed for a population of ASD subjects and typically developing controls (TDCs). Our dataset consisted of 83 male children, 24 ASD and 59 TDCs, aged 6-18 years (mean=12.9yrs, SD=3.0 in ASD, and mean=11.6yrs, SD=3.2 in TDC, no significant group difference in age). DTI was acquired for each subject (Siemens 3T Verio, 32 channel head coil, single shot spin echo sequence, TR/TE = 11000/76 ms, b = 1000 s/mm^2, 30 gradient directions). 79 ROIs from the Desikan atlas were extracted to represent the nodes of the structural network. Probabilistic tractography [10] 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 represents the conditional probability of a pathway between regions, normalized by the active surface area of the seed ROI.

Connectivity Component Analysis. The 79×79 connectivity matrix of each subject was vectorized to its $m = 3081$ upper triangular elements. To compute

the components, we set $\lambda = 1$ and used $k = 3$ to construct the three graphs, and correspondingly calculated their graph edge weights ($\boldsymbol{S}$) using a Gaussian kernel [9]. We used $q = 4$ discriminative, $p = 2$ developmental, and 6 reconstructive components (we suggest the number of reconstructive components be equal to the total of other components). The iterative procedure of equation (5) yielded components shown in Fig. 2. The connectivity components shown were sparse and thresholded for binary visualization to show the dominant edges. Also, in regards to the orthogonality of the discriminative and developmental components, the non-orthogonality between any two of these components was measured by computing their inner products (or angle) which were all less than 0.04 (between 88 and 90 degrees), in a scale of 0 (90-degree) to 1 (0-degree).

The vectorized connectivity of each subject was projected into the discriminative and developmental components to obtain the coefficients for subsequent statistical t-test to validate the discriminative bases and compare between the ASD and TDC groups. We divided the 83 subjects into three closely-balanced age groups of 6–10 (age group I, 25 subjects), 10–13.5 (age group II, 28 subjects), and 13.5–18 years old (age group III, 30 subjects). A t-test was performed between the coefficients of the subjects in these age groups to validate the ability of the developmental components in capturing the effect of age. We then correlated the coefficients of all subjects with their age. Results are given in Table 1.

It is observed that the discriminatory and developmental bases obtained are quite sparse with localized patterns as expected, while the reconstructive network

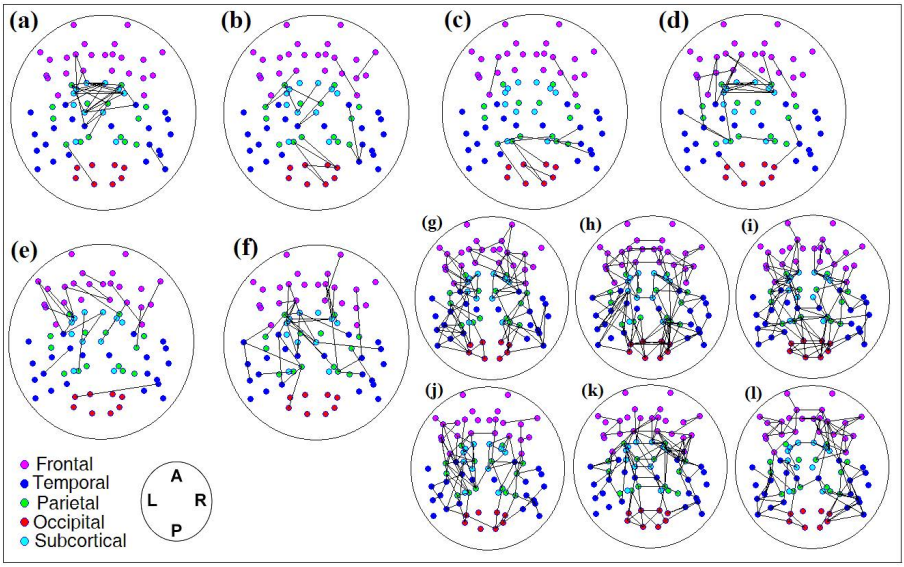

Fig. 2. The r=12 connectivity basis components learned by the proposed method. (a)–(d) are the q=4 discriminative bases, (e)–(f) are the p=2 developmental components, and (g)–(l) compose the set of the 6 reconstructive bases.

Table 1. Statistical group analysis of the coefficients of the ASD-TDC bases

Component label	ASD-TDC group $t(p)$–values	Age group II vs. group I $t(p)$–values	Age group III vs. group II $t(p)$–values	Basis coeffs' 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

components are globally connected. Table 1 shows the results of group-wise and age-based statistics on the coefficients of discriminatory (a-d) and developmental (e-f) components. Group-based analysis shows that the discriminative basis (a) is able to differentiate ASD and TDC groups with a high statistical significance ($p = 0.002$). Inspection of this component in Fig. 2 shows distinct connectivity deficiencies in the thalamic network, and inter-hemespheric 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 subcortical 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 that diminishes (based on its negative correlation=-0.54 with age) significantly with development, especially during the younger ages (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 the components shows that our method is able to extract components that capture the changes due to pathology (a) and age (e); 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.

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, $p > 0.01$) 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 are 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, and therefore, would not have been captured by solving only for the reconstruction components as has been done in the literature. However, the graph-embedded modeling proposed in this work has been able to extract them from the connectivity matrices of the two populations and comprehensively capture the pathology and age specific changes, which were the two major sources of variation in this population.

4 Conclusion

We have presented a novel technique for simultaneously extracting the discriminatory and developmental sub-networks of a population via graph embedding. Our method consists of an 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 subcortical connectivity deficiencies. The developmental bases captured subnetworks 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.

References

1. Vissers, M., et al.: Brain connectivity and high functioning autism: a promising path of research that needs refined models, methodological convergence, and stronger behavioral links. Neurosci. Biobehav. Rev. 36(1), 604–625 (2012)
2. Dennis, E.L., Jahanshad, N., et al.: Development of brain structural connectivity between ages 12 and 30: A 4-tesla diffusion imaging study in 439 adolescents and adults. Neuroimage 64, 671–684 (2013)
3. Calhoun, V., et al.: Modulation of temporally coherent brain networks estimated using ica at rest and during cognitive tasks. Hum. Brain Mapp. 29(7), 828–838 (2008)
4. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791 (1999)
5. Yan, S., et al.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Patt. Anal. Mach. Intell. 29(1), 40–51 (2007)
6. Ghanbari, Y., Bloy, L., Batmanghelich, K., Roberts, T.P.L., Verma, R.: Dominant component analysis of electrophysiological connectivity networks. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 231–238. Springer, Heidelberg (2012)
7. Berry, M.W., et al.: Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat. Data Anal. 52, 155–173 (2007)
8. Yang, Z., Oja, E.: Linear and nonlinear projective nonnegative matrix factorization. IEEE Trans. Neural Netw. 21(5), 1734–1749 (2010)
9. He, X., Niyogi, P.: Locality preserving projections. In: Advances in neural information processing systems (NIPS), vol. 16, pp. 153–160 (2004)
10. Behrens, T., et al.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. 6(7), 750–757 (2003)