IDENTIFYING STRUCTURAL BRAIN NETWORKS FROM FUNCTIONAL CONNECTIVITY: A NETWORK DECONVOLUTION APPROACH

Yang Li and Gonzalo Mateos

Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

ABSTRACT

We address the problem of identifying structural brain networks from signals measured by resting-state functional magnetic resonance imaging (fMRI). To this end, we model functional brain activity as graph signals generated through a linear diffusion process on the unknown structural network. While this is admittedly an oversimplification of the complex mechanisms at work in the brain, recent studies have shown it is an accurate generative model for the second-order statistics of functional signals. We show the diffusion model implies that the signal covariance matrix (a.k.a. functional connectivity) is an unknown polynomial function of the structural network's adjacency matrix. Accordingly, we advocate a network deconvolution approach whereby we: (i) use the fMRI signals to estimate the eigenvectors of the structural network from those of the empirical covariance; and (ii) solve a convex, sparsity-regularized inverse problem to recover the eigenvalues that were obscured by diffusion. The inferred structural networks capture some key patterns that match known pathology in attention deficit/hyper activity disorder. We also offer preliminary evidence supporting their role as potential biomarkers for subject diagnosis and classification.

***Index Terms*—** Brain network inference, functional signals, graph signal processing, diffusion process, network deconvolution.

1. INTRODUCTION

Understanding brain function represents one of the most fundamental and pressing scientific challenges of the 21st century. Driven by advances in neuroimaging technology, brain data have increased in volume and complexity, and accordingly graph-centric tools and methods of network science have become indispensable for mapping and modeling brain structure [1, 2], as well as function [3].

Brain connectivity broadly consists of networks of brain regions connected by functional associations (functional connectivity) [4] or anatomical tracts (structural connectivity) [5]. Structural connectivity can be extracted from tractography algorithms applied to diffusion magnetic resonance imaging (MRI) or diffusion tensor imaging (DTI). Functional networks representing pairwise correlation structure between activation signals in various brain regions are measured by functional MRI (fMRI) or positron emission tomography (PET). The study of brain activity patterns expressed on brain networks is a timely application domain [6], where it is possible but costly to measure structural and functional networks separately due to different spatio-temporal resolutions, running time, and acquisition methods [7]. Consequently, the relationship between structural and functional connectivity of brain networks is of great importance and a very active area of research [8–12].

In this paper, a network deconvolution framework is put forth for identifying the topology of the structural brain network from functional brain signals measured by resting-state fMRI. The basic assumption of our model (Section 2) is that communications among brain regions depend crucially on the neuronal pathways [13], and the observed brain signals are generated through a network diffusion process on the said structural brain graph [12–15]. Fundamental to the method proposed here, recent studies have shown that linear diffusion dynamics can accurately model the relationship between structural and functional brain connectivity networks [14]. While this is admittedly an oversimplification of the complex mechanisms at work in the brain, results in [14] suggest it can be an accurate generative model for the second-order statistics of functional signals (i.e., the functional connectivity graph). In Section 3 we show the diffusion model implies that the blood-oxygen-level dependent (BOLD) signal covariance matrix is an unknown polynomial function of the structural network's adjacency matrix, and hence both matrices have the same eigenvectors [16]. Using the graph signal processing (GSP) parlance, this means that functional signals are approximately stationary on the structural connectivity network [17]. Accordingly, in Section 3.1 we advocate a network deconvolution approach [16] whereby we: (i) use the fMRI signals to estimate the eigenvectors of the structural network from those of the empirical covariance matrix; and (ii) solve a convex, sparsity-regularized inverse problem to recover the eigenvalues that were obscured by diffusion. Different from the linear model inversion approach in [14], here we exploit sparsity that is a cardinal property of structural connectivity [7], and do not require prior knowledge on the sought graph's degree distribution. Our algorithm is also computationally more efficient than traditional approaches which rely on large-scale simulations of nonlinear cortical activity models, and then extrapolate the results to the entire brain following inter-region couplings dictated by the structural connectivity [11].

In Section 4 we corroborate the effectiveness of the proposed approach in recovering structural connectivity using both simulated brain signals (on real structural graphs) and real fMRI data. Moreover, networks inferred from the attention deficit/hyper activity disorder (ADHD)-preprocessed dataset [18] are used for group-wise and subject-wise analysis to identify significant differences between patients and controls. ADHD is the most commonly diagnosed neurodevelopmental disorder in children [19], and previous works have used actual structural networks to detect brain anomalies due to ADHD [20,21]. Our approach identifies some significant differences consistent with known pathology, and leverages the inferred structural networks to competitively classify subjects from each group. It has the distinctive advantage of directly using functional signals for structural analysis, circumventing structural network acquisition methods such as DTI. Concluding remarks are given in Section 5.

2. PRELIMINARIES AND PROBLEM STATEMENT

We first formalize the notion of functional brain signals supported on the structural brain network. Then we introduce the overarching diffusion model for the BOLD signals, which relates the structural and functional connectivities and makes topology inference feasible. We close by stating the problem of inferring structural connectivity from fMRI signals.

Work in this paper was supported in part by the NSF award CCF-1750428. Emails: yli131@ur.rochester.edu, gmateosb@ece.rochester.edu.

2.1. Structural brain networks and functional signals

Structural brain networks represent anatomical connectivity patterns between brain regions [6, 7]. They are often mathematically modeled via a weighted, undirected graph $\mathcal{G} := (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is a set of N nodes corresponding to brain regions (according to standard parcellations [6]), and $\mathbf{A}$ is the symmetric adjacency matrix with $A_{ij} \geq 0$ representing the strength of axonal connections between regions i and j. Structural brain networks are typically sparse, meaning that the number of non-zero weights A_{ij} is significantly smaller than $\binom{N}{2}$. With $\mathbf{\Lambda} := \text{diag}(\lambda_1, \ldots, \lambda_N)$ denoting the diagonal matrix of eigenvalues and $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ the orthonormal matrix of eigenvectors, one can decompose the symmetric adjacency matrix as $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$.

Besides structural connectivity, it is also possible to acquire brain activity signals $\mathbf{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$, where the value of the ith component x_i quantifies the level of neuronal activity in brain region $i \in \mathcal{V}$. Here we focus on resting-state fMRI readings collected over $t = 1, \ldots, T$ time points. Accordingly, we arrange the BOLD signals acquired for all the N studied brain regions over T successive time points in the matrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_T] \in \mathbb{R}^{N \times T}$, where the column vector $\mathbf{x}_t \in \mathbb{R}^N$ represents the (centered) functional brain signal at time t. Likewise, the row vector $\mathbf{x}_i^\top \in \mathbb{R}^T$ represents the BOLD time series at brain region $i \in \mathcal{V}$. The functional connectivity network is given by the correlation matrix $\mathbf{\Sigma} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] \in \mathbb{R}^{N \times N}$ [7], which can be estimated from the signals in $\mathbf{X}$ via sample averaging, namely, $\hat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t^\top$.

Recently there has been growing interest in adopting GSP tools and models [22] to carry out principled analyses of brain activity from neuroimaging data; see [6] for a recent tutorial treatment. In fact, GSP offers a natural framework to exploit the underlying pattern of structural connectivity that couples the signal values at different brain regions [15, 23, 24]. In this direction, next we introduce a generative model for the brain activity signals $\mathbf{x}$ supported on $\mathcal{G}$.

2.2. Diffusion process model of functional brain signals

Here we postulate a simple network diffusion model for the functional signal $\mathbf{x}$, that establishes a useful link between functional connectivity $\mathbf{\Sigma}$ and the structural network $\mathbf{A}$. Consider a zero-mean white input signal $\mathbf{w}$ with covariance matrix $\mathbb{E}\left[\mathbf{w}\mathbf{w}^\top\right] = \mathbf{I}$. We say that $\mathcal{G}$ represents the structure of the BOLD signal $\mathbf{x}$ if there exists a diffusion process in $\mathbf{A}$ that generates the signal $\mathbf{x}$ from $\mathbf{w}$ [16], i.e.,

$$\mathbf{x} = \alpha_0 \prod_{l=1}^{\infty}(\mathbf{I} - \alpha_l\mathbf{A})\mathbf{w} = \sum_{l=0}^{\infty} \beta_l \mathbf{A}^l\, \mathbf{w}. \tag{1}$$

While the adjacency matrix $\mathbf{A}$ only encodes one-hop interactions among brain regions, each successive application of $\mathbf{A}$ (when viewed as an operator acting on brain signals) in (1) percolates $\mathbf{w}$ over the entire $\mathcal{G}$ [25]. The justification for (1) is that we can think of the edges of $\mathcal{G}$, i.e. the non-zero entries in $\mathbf{A}$, as direct (one-hop, anatomical) relations between brain regions. The diffusion process modifies the original correlation by inducing multi-hop relations, thus capturing the indirect interactions and shaping up the functional connectivity structure [26]. In the next section we show how (1) induces a simple polynomial (convolutive) relationship between $\mathbf{\Sigma}$ and $\mathbf{A}$. Thus our goal is to derive a so-termed *network deconvolution* estimator of the structural connectivity $\mathbf{A}$, as described in the following formal problem statement.

Problem: Given resting-state fMRI readings $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_T] \in \mathbb{R}^{N \times T}$ generated by diffusion (1) in the network $\mathcal{G}$, estimate the structural connectivity encoded in the sparse adjacency matrix $\mathbf{A}$.

With regards to the scope and validity of the model, linear diffusion has been validated as a tenable mechanism to describe the structural-functional connectivity relationship [12–14]. While admittedly simplistic (it is known that complex non-linear neural processes are prevalent in the brain), a linear model like (1) can be accurate to describe the functional structure in $\mathbf{\Sigma}$ [14]. Moreover, there is evidence that functional links tend to form where there is no or little structural connection [26], a characteristic naturally captured by (1).

3. NETWORK DECONVOLUTION

The diffusion expressions in (1) are polynomials on $\mathbf{A}$ of possibly infinite degree, yet the Cayley-Hamilton theorem asserts they are equivalent to polynomials of degree smaller than N. By defining a coefficient vector $\mathbf{h} := [h_0, \ldots, h_{L-1}]^\top \in \mathbb{R}^L$ and the graph filter $\mathbf{H} := \sum_{l=0}^{L-1} h_l\mathbf{A}^l \in \mathbb{R}^{N \times N}$ [22, 27], (1) can be rewritten as

$$\mathbf{x} = \left(\sum_{l=0}^{L-1} h_l\mathbf{A}^l\right)\mathbf{w} = \mathbf{H}\mathbf{w}, \tag{2}$$

for some particular $\mathbf{h}$ and $L \leq N$. Since the graph filter $\mathbf{H}$ is a polynomial on $\mathbf{A}$, it is a linear graph-signal operator that has the same eigenvectors $\mathbf{V}$ as the adjacency matrix. More important for the approach adopted here, the convolutive signal model $\mathbf{x} = \mathbf{H}\mathbf{w}$ can be used to show that *the eigenvectors of* $\mathbf{A}$ *are also eigenvectors of the functional connectivity matrix* $\mathbf{\Sigma}$. To that end, recall that $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ to decompose the filter as $\mathbf{H} = \sum_{l=0}^{L-1} h_l(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l\mathbf{\Lambda}^l)\mathbf{V}^\top$. Also using that $\mathbb{E}\left[\mathbf{w}\mathbf{w}^\top\right] = \mathbf{I}$ one can write

$$\mathbf{\Sigma} = \mathbb{E}\left[\mathbf{H}\mathbf{w}(\mathbf{H}\mathbf{w})^\top\right] = \mathbf{H}\mathbf{H}^\top = \mathbf{V}(\textstyle\sum_{l=0}^{L-1} h_l\mathbf{\Lambda}^l)^2\mathbf{V}^\top. \tag{3}$$

The eigenvectors of the structural graph $\mathbf{A}$ and the functional connectivity matrix $\mathbf{\Sigma}$ are the same. Equivalently, (3) implies $\mathbf{\Sigma} = \phi(\mathbf{A}) = \mathbf{V}\phi(\mathbf{\Lambda})\mathbf{V}^\top$ for some analytic (i.e., polynomial) matrix function ϕ that depends on $\mathbf{h}$ and L. The only difference between $\mathbf{\Sigma}$, which includes indirect relationships between signal components, and $\mathbf{A}$, which contains only direct anatomical connections, is on their eigenvalues. These observations motivate a general two-step network topology inference approach [16] whereby we: (i) use the fMRI signals in $\mathbf{X}$ to estimate the eigenvectors $\hat{\mathbf{V}}$ of $\mathbf{A}$ from the empirical covariance $\hat{\mathbf{\Sigma}}$; and (ii) rely on these spectral templates to recover $\mathbf{A}$ by estimating its eigenvalues – the subject dealt with next.

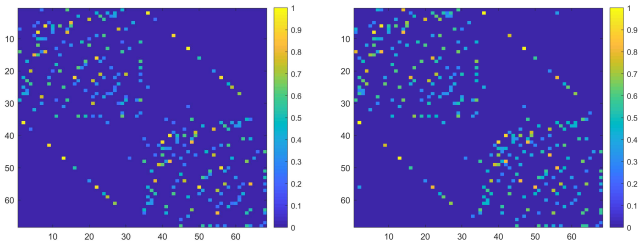
3.1. Recovering structural graphs from functional connectivity

Identity (3) also shows that the deconvolution problem of identifying the structural brain network from functional connectivity is underdetermined. As long as the matrices $\mathbf{\Sigma}$ and $\mathbf{A}$ have the same eigenvectors, there exist filter coefficients $\mathbf{h}$ that generate $\mathbf{x}$ through a diffusion process on $\mathcal{G}$ [cf. (1)]. To sort out this ambiguity, which amounts to selecting the eigenvalues of $\mathbf{A}$ in step (ii), we assume that the adjacency matrix of interest is optimal in some sense. At the same time, it is prudent to account for the (finite sample, noise-induced) discrepancies between $\hat{\mathbf{V}}$ and the actual eigenvectors of $\mathbf{A}$. Accordingly, we build on [16] and seek to estimate the graph topology $\mathbf{A}$ that: (a) is optimal with respect to convex criteria $f(\mathbf{A}, \mathbf{\Lambda})$; (b) belongs to a convex set $\mathcal{S}$ that specifies admissible adjacency matrices; and (c) is close to $\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T$ as measured by a convex matrix distance $d(\cdot, \cdot)$. Formally, our idea is to solve the inverse problem

$$\hat{\mathbf{A}} := \underset{\mathbf{\Lambda}, \mathbf{A} \in \mathcal{S}}{\text{argmin}}\ f(\mathbf{A}, \mathbf{\Lambda}), \quad \text{s. to } d(\mathbf{A}, \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T) \leq \epsilon, \tag{4}$$

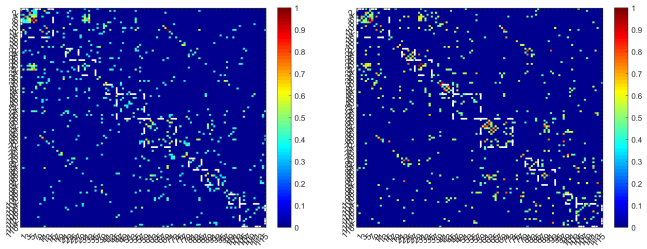
where ϵ is chosen based on a priori information on the noise level.

Within the scope of the model (1), the formulation (4) entails a general class of network topology inference problems [16]. For instance, the selection of $f(\mathbf{A}, \mathbf{\Lambda})$ allows to seamlessly incorporate



(a) (b)

Fig. 1: (a) Ground-truth structural network with $N = 68$. (b) Recovered topology after solving (5). The rows and columns (indexing brain regions) of $\mathbf{A}$ have been rearranged so that the 34 regions in the left brain hemisphere are shown first. The sparser pattern of inter-hemisphere connections is apparent.



(a) (b)

Fig. 2: Study of ADHD-200 data. Recovered structural network for the (a) control group; and (b) ADHD patient group. Blocks along the diagonal correspond to the Frontal, Occipital, Parietal, Temporal and Cerebellum region on the left and right brain hemispheres, respectively.

physical characteristics of the desired graph. Structural brain networks are sparse and there are fewer inter-hemisphere connections in the brain; see [7] and Fig. 1a. Accordingly, here we choose a sparsity-promoting criterion $f(\mathbf{A}, \mathbf{\Lambda}) = \|\mathbf{W} \circ \mathbf{A}\|_1$, where $\circ$ stands for (entry-wise) Hadamard product and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is a weight matrix to impose non-uniform sparsity priors across candidate edges. In the numerical tests of Section 4, we set $W_{ij} = 1$ for edges joining regions i and j in different hemispheres, and $W_{ij} = 0.5$ otherwise.

Adopting a Frobenius-norm distance in (4), we estimate

$$\hat{\mathbf{A}} := \underset{\mathbf{\Lambda}, \mathbf{A} \in \mathcal{S}}{\operatorname{argmin}} \ \|\mathbf{W} \circ \mathbf{A}\|_1, \quad \text{s. to } \|\mathbf{A} - \hat{\mathbf{V}} \mathbf{\Lambda} \hat{\mathbf{V}}^T\|_F^2 \leq \epsilon, \quad (5)$$

where $\mathcal{S} := \{\mathbf{A} \,|\, A_{ij} = A_{ji} \geq 0, A_{ii} = 0, \sum_j A_{j1} = 1\}$ since we aim to recover the adjacency matrix of an undirected graph, with no self-loops and non-negative weights representing the strength of axonal connections. The last condition in $\mathcal{S}$ fixes the scale of the admissible graphs by setting the weighted degree of the first node to 1 (without loss of generality), ruling out the trivial solution $\mathbf{A} = \mathbf{0}$.

There are two major tasks to consider when it comes to quantifying the computational complexity of the proposed network deconvolution approach. First, computing the eigenvectors of $\hat{\mathbf{\Sigma}}$ incurs $\mathcal{O}(N^3)$ complexity. Second, solving iteratively the convex optimization problem (5) costs $\mathcal{O}(N^3)$ per iteration using e.g., off-the-shelf sparse minimization solvers. The overall procedure scales well to standard brain parcellations or atlases that comprise (at most) a few hundred brain regions, and is more efficient than traditional approaches based on non-linear simulations [11]. The numerical tests in Section 4 were carried out using the CVX package for Matlab.

4. NUMERICAL TEST CASES

In this section, we test the structural network inference framework on both simulated and real fMRI signals from public ADHD datasets. Throughout, we set ϵ in (5) via a binary search in $[0.5, 1.5]$ to find the smallest value that gives a feasible solution $\hat{\mathbf{A}} \in \mathcal{S}$.

4.1. Simulated fMRI signals on real structural brain network

We first validate our approach on simulated signals, using a ground-truth preprocessed structural network with $N = 68$ brain regions estimated via DTI [28]; see Fig. 1a. We synthetically generate fMRI signals $\mathbf{x}_t$ adhering to the diffusion model in (1), with standard Gaussian inputs $\mathbf{w}_t$, for $t = 1, \ldots, 10^4$. We form the sample covariance matrix $\hat{\mathbf{\Sigma}}$ and find its eigenvectors. By solving (5) we estimate the structural connectivity network shown in Fig. 1b.

A clear correspondence between the recovered structural network and the ground truth is apparent. To improve visualization, edge weights in both networks have been normalized and thresholded with the largest value yielding a connected graph. Besides recovering connections between brain regions (the support of $\mathbf{A}$), our model can also provide an accurate estimate of the connection strengths (i.e. edge weights). Specifically, we obtained a relative recovery error of 11.1%, averaged over 10 Monte Carlo realizations of the experiment.

4.2. Real ADHD data

To validate our approach with real data, we first consider the ADHD-200 dataset [18]. This public repository was put together through a collaboration of 8 international imaging sites that aggregate and openly share neuroimaging data. In our experiment, we obtain the preprocessed BOLD timeseries from fMRI scans of 182 healthy subjects (92 male, 90 female, age = 12.18 ± 3.35) and 107 ADHD type-1 patients (82 male, 25 female, age = 11.65 ± 3.33) from KKI, NYU and NeuroImaging datasets acquired from 3 different imaging sites.

Brain fMRI signals are aggregated and registered on the standard AAL-116 brain atlas, hence $N = 116$. As a result, for each region we have available a BOLD timeseries of duration $T = \{120, 257, 172\}$, respectively the 3 aforementioned imaging sites. We concatenate the brain signals of all subjects in each group and form a matrix $\mathbf{X}_c \in \mathbb{R}^{116 \times 182T}$ collecting the brain signals of the control group, and likewise $\mathbf{X}_p \in \mathbb{R}^{116 \times 107T}$ for the patient group. After running the network deconvolution pipeline proposed in Section 3 for both groups, Fig. 2 depicts the recovered structural networks. To avoid hindering visualization, both graphs have been thresholded to keep only edges with the top 10% weights. In both cases, notice the denser pattern of connections within hemispheres.

For the patient group network shown in Fig. 2b, edges are more clustered and concentrated within the 10 blocks along the main diagonal corresponding to the Frontal, Occipital, Parietal, Temporal and Cerebellum regions of the left and right brain hemispheres. In the control group however, connections are visibly less structured and no brain region has a significant enhanced activity. This difference indicates higher inter-connectivity in the 10 aforementioned regions, consistent with the findings in [29]. Accordingly, the group representative connectivity matrices provide a visual representation of some important qualitative differences between the control and ADHD patient group.

Graph theory has been proven useful for brain network analysis [30]. To quantitatively analyze and compare the representative structural network for each group, we evaluate several graph-based measures and identify significant differences between groups

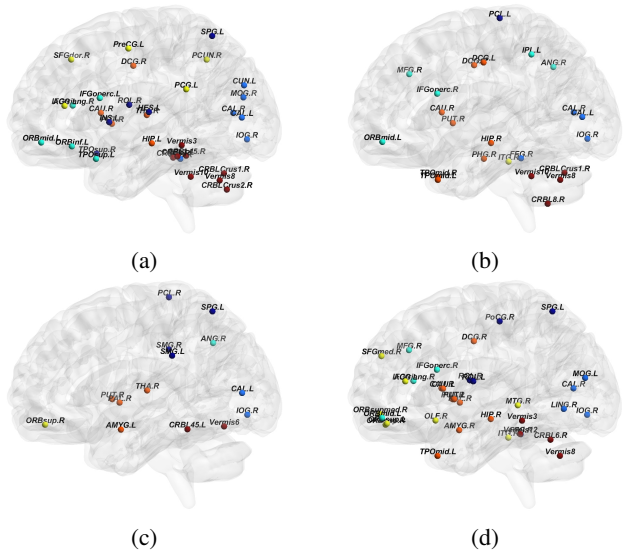


Fig. 3: Regions with statistically significant (positive) differences in ADHD patients based on: (a) clustering coefficient; (b) degree; (c) local efficiency; and (d) closeness centrality. Most brain regions identified are in the Frontal, Occipital, Parietal and Temporal areas, suggesting increased connectivity for the ADHD group. This is consistent with the findings in [19, 29].

($p = 0.01$). Some noteworthy findings follow, while details on the statistical tests are omitted due to lack of space.

Loss of long-range connection. We compute the physical distance between brain regions connected by edges in the networks in Fig. 2. A significant decrease in connection length is found for the ADHD group. This matches well the findings in [29], citing a decrease in global, long-range anatomical connections for ADHD patients.

Increased local connection. The clustering coefficient and local efficiency measures quantify the density of sub-networks centered at each region (a.k.a. egonets). The significantly larger clustering coefficient and local efficiency in Fig. 2b shows that ADHD brain regions are more locally connected with neighboring regions [29].

Highly inter-connected subnetwork. A sub-network of stronger connectivity encompassing the Frontostriatal, Occipital, Temporal and Parietal regions is found in ADHD subjects [29]. To corroborate this finding, we extract such sub-network for both groups and calculate node-level subgraph centrality (a weighted sum of closed walks of different lengths in the sub-network). We also find the ADHD group has significantly higher subgraph centrality than control.

Regions with most difference in graph measures. We calculate node-level graph measures (degree, clustering coefficient, closeness centrality, local efficiency; see [5, Table A1]) and single out the brain regions where the values of the aforementioned graph measures are significantly larger for subjects in the ADHD group. The results are summarized in Fig. 3. Most brain regions identified are in the Frontal, Occipital, Parietal and Temporal areas, once more suggesting that the connectivity in these areas is stronger for ADHD patients than controls. Moreover, brain areas such as Hippocampus, Basal Ganglia (caudate, putamen, pallidum), Thalamus, Amygdala are also detected. The results are consistent with known ADHD pathology [19, 29], and serve to identify brain regions with high discrimination power between groups.

4.3. Network-based subject-level classification

Here we use the recovered structural networks for control-patient classification. Data are preprocessed BOLD timeseries from 30 controls (all male, age $= 11.367 \pm 1.91$) and 29 patients (1 female, age $= 11.81 \pm 1.75$) from the Peking dataset. We adopt a different (now single imaging site) dataset to test the generality and flexibility of the topology inference approach. For each subject, we recover a structural network from the given fMRI signals and compute the graph measures (degree, clustering coefficient, closeness centrality, local efficiency) of each of the 116 nodes. We concatenate all these graph measure values, thus each subject has a 464×1 feature vector.

In our previous study, we identified 23, 14, 34, 34 brain regions that are significantly different between groups based on degree, local efficiency, clustering coefficient and closeness centrality, respectively; see Fig. 3. We only focus on these important regions to reduce the dimensionality of the per subject features to 105. We then apply the sequential forward feature selection method to further reduce the number of features to 6. The retained features correspond to the degree of left paracentral lobule, right inferior frontal gyrus and right fusiform, local efficiency of right paracentral lobule, closeness of right fusiform and right caudate.

To assess the discrimination power of the 105 regions identified in our subject-level analysis, we also use the original 464×1 feature vector (with sequential forward feature selection) to classify the subjects. We compare with the approach in [31], that can as well be used to infer structural brain graphs from functional connectivity. However, [31] adopts a very specific diffusion model to infer direct relations (structural network) from indirect ones (functional connectivity). Table 1 summarizes the best classification performance of each framework, across different methods such as KNN and SVM.

	ACC	AUC	TPR	TNR
105 features with selection	0.774	0.836	0.767	0.782
464 features with selection	0.678	0.737	0.733	0.621
Method in [31]	0.492	0.493	0.522	0.459

Table 1: Subject classification performance. ACC: accuracy; AUC: area under curve; TPR: true positive rate; TNR: true negative rate. All experiments were carried out using Matlab's Classification Learner with 10-fold cross-validation.

The results show that those 105 regions identified previously can improve classification performance, which implicitly corroborates the ability of our framework to accurately infer structural networks and capture key brain patterns. In addition, our approach outperforms [31] by modeling brain activity via a more general diffusion process and promoting edge sparsity. Our results are superior to previous tests on ADHD-200 data [21,32], and are also competitive with the best results of the ADHD-200 global competition[1]. Subject-level network inference and disease diagnosis are generally more challenging than identifying significant group differences, so our future research agenda will be steered towards this exciting direction.

5. CONCLUSIONS

This paper puts forth a network deconvolution framework to identify structural brain networks from fMRI signals. Based on a simplifying diffusion model assumption, we estimated both group-level and subject-level structural networks and identified patterns that match known pathology. ADHD patient classification was also carried out to (indirectly) validate the discriminative power of the brain regions identified, which could serve as potential biomarkers for patient diagnosis. To the best of our knowledge, this is the first time that structural connectivity is estimated from functional signals for disease-related analysis. Future research will focus on GSP-based brain signal analysis in the graph frequency domain, for improved denoising and unveiling of additional discriminative (spectral) features.

[1]To view the competition results, see: `http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html`

6. REFERENCES

[1] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of Brain Network Analysis*, Academic Press, 2016.

[2] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, pp. e159, 2008.

[3] J. Power, A. Cohen, S. Nelson, G. Wig, K. Barnes, J. Church, A. Vogel, T. Laumann, F. Miezin, B. Schlaggar, et al., "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, 2011.

[4] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville, "Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 58–70, 2013.

[5] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186, 2009.

[6] W. Huang, T. A. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proc. IEEE*, vol. 106, no. 5, 2018.

[7] O. Sporns, *Networks of the Brain*, MIT Press, 2010.

[8] P. Skudlarski, K. Jagannathan, V. D. Calhoun, M. Hampson, B. A. Skudlarska, and G. Pearlson, "Measuring brain connectivity: Diffusion tensor imaging validates resting state temporal correlations," *Neuroimage*, vol. 43, no. 3, pp. 554–561, 2008.

[9] R. F. Galán, "On how network architecture determines the dominant patterns of spontaneous neural activity," *PLoS One*, vol. 3, no. 5, pp. e2148, 2008.

[10] C. J. Honey, J.-P. Thivierge, and O. Sporns, "Can structure predict function in the human brain?," *Neuroimage*, vol. 52, no. 3, pp. 766–776, 2010.

[11] C. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 6, pp. 2035–2040, 2009.

[12] J. Goñi, M. P. van den Heuvel, A. Avena-Koenigsberger, N. V. de Mendizabal, R. F. Betzel, A. Griffa, P. Hagmann, B. Corominas-Murtra, J.-P. Thiran, and O. Sporns, "Resting-brain functional connectivity predicted by analytic measures of network communication," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 2, pp. 833–838, 2014.

[13] A. Raj, A. Kuceyeski, and M. Weiner, "A network diffusion model of disease progression in dementia," *Neuron*, vol. 73, no. 6, pp. 1204–1215, 2012.

[14] F. Abdelnour, H. U. Voss, and A. Raj, "Network diffusion accurately models the relationship between structural and functional brain connectivity networks," *Neuroimage*, vol. 90, pp. 335–347, 2014.

[15] C. Hu, X. Hua, J. Ying, P. M. Thompson, G. E. Fakhri, and Q. Li, "Localizing sources of brain disease progression with network diffusion model," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1214–1225, Oct. 2016.

[16] S. Segarra, A. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Aug. 2017.

[17] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Aug. 2017.

[18] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The neuro bureau ADHD-200 preprocessed repository," *Neuroimage*, vol. 144, pp. 275–286, 2017.

[19] A. De La Fuente, S. Xia, C. Branch, and X. Li, "A review of attention-deficit/hyperactivity disorder from the perspective of brain networks," *Front. Hum. Neurosci.*, vol. 7, pp. 192, 2013.

[20] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural mri data," *PLoS One*, vol. 8, no. 11, pp. e79476, 2013.

[21] C.-W. Chang, C.-C. Ho, and J.-H. Chen, "ADHD classification by a texture analysis of anatomical brain mri data," *Front. Syst. Neurosci.*, vol. 6, pp. 66, 2012.

[22] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[23] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals.," *J. Sel. Topics Signal Processing*, vol. 10, no. 7, pp. 1189–1203, 2016.

[24] M. Ménoret, N. Farrugia, B. Pasdeloup, and V. Gripon, "Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2017, pp. 618–622.

[25] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, 2017.

[26] J. S. Damoiseaux and M. D. Greicius, "Greater than the sum of its parts: A review of studies combining structural connectivity and resting-state functional connectivity," *Brain Struct. Func.*, vol. 213, no. 6, pp. 525–533, 2009.

[27] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[28] J. A. Brown, J. D. Rudie, A. Bandrowski, J. D. Van Horn, and S. Y. Bookheimer, "The UCLA multimodal connectivity database: A web-based platform for brain connectivity matrix sharing and analysis," *Front. Neuroinform.*, vol. 6, pp. 28, 2012.

[29] R. Beare, C. Adamson, M. A. Bellgrove, V. Vilgis, A. Vance, M. L. Seal, and T. J. Silk, "Altered structural connectivity in ADHD: A network based analysis," *Brain Imaging Behav.*, vol. 11, no. 3, pp. 846–858, 2017.

[30] Q. Yu, Y. Du, J. Chen, J. Sui, T. Adalē, G. D. Pearlson, and V. D. Calhoun, "Application of graph theory to assess static and dynamic brain connectivity: Approaches for building brain graphs," *Proc. IEEE*, vol. 106, no. 5, pp. 886–906, 2018.

[31] S. Feizi, D. Marbach, M. Médard, and M. Kellis, "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nat. Biotechnol.*, vol. 31, no. 8, pp. 726, 2013.

[32] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Front. Syst. Neurosci.*, vol. 6, pp. 63, 2012.