Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes

Colin J. Brown[1], Steven P. Miller[2], Brian G. Booth[1], Jill G. Zwicker[3], Ruth E. Grunau[3], Anne R. Synnes[3], Vann Chau[2], and Ghassan Hamarneh[1]

[1]Medical Image Analysis Lab, Simon Fraser University, BC, Canada, [2]The Hospital for Sick Children and The University of Toronto, ON, Canada, [3]University of British Columbia and Child and Family Research Institute, BC, Canada

Abstract. We present a new method to identify anatomical subnetworks of the human white matter connectome that are predictive of neurodevelopmental outcomes. We employ our method on a dataset of 168 preterm infant connectomes, generated from diffusion tensor images (DTI) taken shortly after birth, to discover subnetworks that predict scores of cognitive and motor development at 18 months. Predictive subnetworks are extracted via sparse linear regression with weights on each connectome edge. By enforcing novel backbone network and connectivity based priors, along with a non-negativity constraint, the learned subnetworks are simultaneously anatomically plausible, well connected, positively weighted and reasonably sparse. Compared to other state-of-the-art subnetwork extraction methods, we found that our approach extracts subnetworks that are more integrated, have fewer noisy edges and that are also better predictive of neurodevelopmental outcomes.

1   Introduction

Preterm birth is a world-wide health challenge, affecting millions of children every year [1]. Very preterm birth ($\leq$ 32 weeks post-menstrual age, PMA) affects brain development and puts a child at a high risk for delayed, or altered, cognitive and motor neurodevelopment. It is known from studies of diffusion MR images, that the development of white matter plays a critical role in the function of a child's brain, and that white matter injury is associated with poorer outcomes [2–5]. Recently, Ziv et al. and Brown et al. showed that by representing the set of white matter connections as a network (i.e., connectome), features of network topology could be used to predict abnormal general neurological function and neuromotor function respectively [4, 6].

Representing a diffusion tensor image (DTI) of the brain as a network defined between regions of interest (ROIs) allows an anatomically informed reduction of dimensionality from millions of tensor-valued voxels down to thousands of connections (edges). However, for the purposes of prediction, thousands of features may still be too many and cause over-fitting when limited numbers (e.g. only hundreds) of scans are available [7]. Furthermore, region of interest based studies

suggest that structural abnormalities related to poor neurodevelopmental outcomes are not spread evenly across the entire brain, but instead are localized to particular anatomy [3]. Thus, there is motivation to discover which particular subnetworks (group of connections or edges) in the brain network best predict different brain functions.

Some previous works have explored the use of brain subnetworks for predicting outcomes [8, 9, 7, 10]. For instance, Zhu et al. used t-tests at each edge in a dataset of functional connectomes for group discriminance, followed by correlation-based feature selection and training of a support vector machine (SVM), to find subnetworks that were predictive of schizophrenia [8]. This multistage feature selection and model training is not ideal, however, because it precludes simultaneous optimization of all model parameters. Munsell et al. used an Elastic-Net based subnetwork selection for predicting the presence of temporal lobe epilepsy and the success of corrective surgery in adults [7]. This method encourages sparse selection of stable features, useful for identifying those edges most important for prediction [11], but fails to leverage the underlying structure of the brain networks that might inform the importance or the relationships between edges. In order to capture dependencies between neighbouring edges, Li et al. employed a Laplacian-based regularizer (in a framework similar to GraphNet [11]) that encouraged their subnetwork weights to smoothly vary between neighbouring edges [10]. However, this smoothing may reduce sparsity by promoting many small weights and blur discontinuities between the weights of neighbouring edges that should be preserved. An ideal regularizer would encourage a well connected subnetwork while preserving sparsity and discontinuities. Ghanbari et al. used non-negative matrix factorization to find a sparse set of non-negative basis subnetworks in structural connectomes [9]. However, rather than trying to predict specific outcomes (as we propose below), Ghanbari et al. introduced age-regressive, group-discriminative, and reconstructive regularization terms on groups of subnetworks, encouraging each group to covary with a particular factor. They argued that non-negative subnetwork edge weights are more anatomically interpretable, especially in the case of structural connectomes which have only non-negative edge feature values.

In this paper, we present our novel approach to identifying anatomical subnetworks of the human white-matter connectome that are optimally predictive of a preterm infant's cognitive and motor neurodevelopmental scores assessed at 18 months of age, adjusted for prematurity. Similar to Munsell et al., our method is based on a regularized linear regression on the outcome score of choice. Here, however, we introduce a constraint that ensures the non-negativity of subnetwork edge weights. We further propose two novel informed priors designed to find predictive edges that are both anatomically plausible and well integrated into a connected subnetwork. We demonstrate that these priors effectuate the desired effect on the learned subnetworks and that, consequently, our method outperforms a variety of other competing methods on this very challenging outcome prediction task. Finally, we discuss the structure of the learned subnetworks in the context of the underlying neuroanatomy.

2 Method

2.1 Preterm Data: Our dataset contains 168 scans taken between 27 and 45 weeks PMA from a cohort of 115 preterm infants (nearly half of the infants were scanned twice), born between 24 and 32 weeks PMA. Connectomes were generated for each scan by aligning an infant atlas of 90 anatomical brain regions with each DTI. Full-brain streamline tractography was then performed in order to count the number of tracts (i.e., edge strength) connecting each pair of regions. Our previous works provide details on the scanning and connectome construction processes [6] and a discussion on interpreting infant connectomes [5]. Cognitive and neuromotor function of each infant was assessed at 18 months of age, corrected for prematurity, using the Bayley Scales of Infant and Toddler Development 3rd edition [12]. The scores are normalized to 100 ± 15; adverse outcomes are those with scores at or below 85 (i.e., ≤ -1 std.).

Our dataset is imbalanced, containing few scans of infants with high and low outcome scores. In order to flatten this distribution, the number of connectomes in each training set was doubled by synthesizing instances with high and low outcome scores, using the synthetic minority over-sampling technique [13].

2.2 Subnetwork Extraction: Given a set of preterm infant connectomes, our goal is to find a subnetwork that is: **a)** predictive (i.e., contains edges that accurately predict a neurodevelopmental outcome), **b)** anatomically plausible (i.e., edges correspond to valid axon bundles), **c)** well connected (i.e., high network integration [5]), **d)** reasonably sparse and **e)** non-negative.

Each connectome is represented as a graph $G(V, E)$ comprising a set of 90 vertices, V, and $M = 90 \times 89/2 = 4005$ edges, E. The tract counts associated with the edges are represented as a single feature vector $\mathbf{x} \in \mathbb{R}^{1 \times M}$ and the entire training set of N subjects is represented as $X \in \mathbb{R}^{N \times M}$ with outcome scores $\mathbf{y} \in \mathbb{R}^{N \times 1}$. To find a subnetwork that fits the above criteria, we optimize an objective function over a vector of subnetwork edge weights, $\mathbf{w} \in \mathbb{R}^{M \times 1}$:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} ||\mathbf{y} - X\mathbf{w}||^2 + \lambda_{L1}||\mathbf{w}||_1 + \lambda_B(\mathbf{w}^T B \mathbf{w}) + \lambda_C(\mathbf{w}^T C \mathbf{w}) \quad (1)$$

$$\text{such that } \mathbf{w} \geq 0, \quad (2)$$

where $||\mathbf{w}||_1$ is a sparsity regularization term, B is the network backbone prior matrix (see Section 2.3), and C is the connectivity prior matrix (see Section 2.4). Hyper-parameters, λ_B, λ_C and λ_{L1} are used to weight each of the regularization terms. Given a set of learned weights, $\mathbf{w}^*$, the outcome score of a novel infant connectome, $\mathbf{x}_{new}$ can be predicted as $y_{pred} = \mathbf{x}_{new}\mathbf{w}^*$.

Note that since X is non-negative and since $\mathbf{w}$ is required to be non-negative, we also require $\mathbf{y}$ to be non-negative, as they should since the true Bayley scores range between 45 and 155. To perform this optimization we used the method (and software) of Schmidt et al. [14].

2.3 Network Backbone Prior: Many of the 4005 possible connectome edges are anatomically unlikely (i.e., between regions not connected by white matter fibers) but may be non-zero in certain scans due to imaging noise and accumulated pipeline error (i.e. due to atlas registration, tractography, and tract counting) [15]. With many more edges than training samples, some edges may appear discriminative by pure chance, when in fact they are just noise. Therefore, we propose a network backbone prior term that encodes a penalty discouraging the subnetwork from including edges with a low signal-to-noise ratio (SNR) in the training data. The SNR of the j-th edge can be computed as the ratio $\mathrm{MEAN}(X_{:,j})/\mathrm{SD}(X_{:,j})$. However, this may falsely declare an edge as noisy when the variability (c.f. denominator) in the edge value is not due to noise but rather due to the edges values changing in a manner that correlates with the outcome of the subject. To counteract this problem, we divide the scans into two classes: scans with normal outcomes, H, and scans with adverse outcomes, U. The SNR is then computed separately for each class. Let X_Ω represent a matrix with a subset of the rows in X where $\Omega \in \{U, H\}$. The SNR for each edge, j, in each class, Ω, is computed as $\mathrm{SNR}(X_{\Omega,j}) = \frac{\mathrm{MEAN}(X_{\Omega,j})}{\mathrm{SD}(X_{\Omega,j})}$. In order not to favour the strongest fiber bundles over weak yet important bundles, we threshold the SNR at each edge conservatively, to exclude only the least anatomically likely edges. An edge, j, is only penalized if both $\mathrm{SNR}(X_{U,j})$ and $\mathrm{SNR}(X_{H,j})$ are less than or equal to 1 (i.e., signal is weaker than noise in both classes). In particular, B is an $M \times M$ diagonal matrix, such that,

$$B_{j,j} = \begin{cases} 1, & \text{if } \mathrm{SNR}(X_{H,j}) \leq 1 \text{ and } \mathrm{SNR}(X_{U,j}) \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

So $\mathbf{w}^T B \mathbf{w}$ only penalizes edges that do not pass the SNR threshold among either instances with normal outcomes or abnormal outcomes, and thus are likely noisy. Fig. 1 shows an example of B. Note that, especially for infant connectomes, even edges with high SNR may not represent white matter fibers but instead high FA from other causes [5]. Nevertheless, such high-SNR edges are not likely due to noise but instead to some real effect and thus may aid prediction.

2.4 Connectivity Prior: We also want to encourage the subnetwork to be highly integrated as opposed to being a set of scattered, disconnected edges. This is motivated by the fact that functional brain network activity is generally constrained to white matter structure [16] and white matter structure is organized into well connected link communities [17]. Thus, we do not expect there to be many, disconnected sub-parts of the brain that are all highly responsible for any particular neurodevelopmental outcome type. To embed this prior, we incentivize pairs of edges in the target subnetwork to share common nodes. For edge $e_{i,j}$, between nodes i and j, and edge $e_{p,q}$ between nodes p and q, we construct the matrix,

$$C(e_{i,j}, e_{p,q}) = \begin{cases} -1, & \text{if } i = p \text{ or } i = q \text{ or } j = p \text{ or } j = q \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$
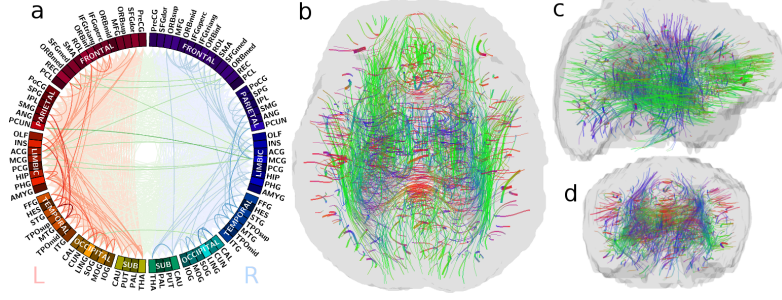
Fig. 1. a) A sample backbone prior network (i.e., all edges where $B_{j,j} = 0$) mapped on to a Circos ideogram (http://circos.ca/). Inter-hemispherical connections are in green and intra-hemispherical connections are in red (left) and blue (right). Opacity of each link is computed as $\text{SNR}(X_{U,j}) \times \text{SNR}(X_{H,j})$. **b)** Axial, **c)** sagittal and **d)** coronal views of the same network rendered as curves representing the mean shape of all tracts between those connected regions (from one infant's scan).

such that the term $\mathbf{w}^T C \mathbf{w}$ becomes smaller (i.e., more optimal) for each pair of non-zero weighted subnetwork edges sharing a node. This term places a priority on retaining edges in the subnetwork that are connected to hub nodes. This is desirable since subnetwork hub nodes indicate regions that join many connections (i.e., edges) predictive of outcome. In contrast to a Laplacian based regularizer which would encourage subnetwork weights to become locally similar, reducing sparsity, our proposed term simply rewards subnetworks with stronger hubs.

3   Results

We compare the proposed subnetwork-driven predicted outcomes for the preterm infant cohort (N=168) with competing outcome prediction techniques. Methods are evaluated using i) Pearson's correlation between ground truth and predicted scores, and ii) the area over the regression error characteristic curve (AOC), which provides an estimate of regression error [18]. Some previous studies have focused on predicting a binary abnormality label instead of predicting actual scalar outcome scores [6, 4]. Thus, to compare more directly to these works, we also evaluate the accuracy of our models as a binary classifier for predicting scores above or below 85. Similar to Brown et al., an SVM was used to classify normal from abnormal instances as it was found to perform better than thresholding the predicted scores at 85. SVM learns a max-margin threshold for the predicted scores (i.e., one input feature), optimal for classification over the training set.

For each method (both proposed and competing), coarse grid searches were performed in powers of two over the method's hyper-parameters to find the best performance for both cognitive and motor outcomes independently. For the proposed method, this search was over $\lambda_{L1}, \lambda_C, \lambda_B \in \{2^0, ..., 2^9\}$. A finer grid search was not performed to avoid over-fitting to the dataset. For each setting of the parameters, a leave-2-out, 1000-round cross validation test was performed. If two scans were of the same infant, those scans were not split between test and

	Motor			Cognitive		
Method	r	AOC	acc.	r	AOC	acc.
Zhu et al. [8]	0.159	27.39	45.1	0.021	28.05	49.7
Elastic-Net [7]	0.270	24.58	58.8	0.207	24.83	54.8
Brown et al. [6]	-	-	62.9	-	-	52.6
Linear Regression	0.270	24.78	58.8	0.245	24.72	55.2
+ L1 regularization	0.314	18.55	64.0	0.244	24.75	55.2
+ Non-neg. Constraint	0.433	14.53	68.8	0.317	17.73	57.7
+ Backbone Prior	0.436	14.47	68.6	0.327	17.82	58.5
+ Connectivity Prior (Ours)	**0.442**	**14.25**	**70.8**	**0.343**	**17.38**	**59.5**

Table 1: Correlation (r) between ground-truth and predicted scores, area over REC curve (AOC) values and classification accuracy of scores at or below 85 (acc.) for each model, assessed via 1000 rounds of leave-2-out cross validation. Note that Brown et al.'s method [6] performs binary classification only.

training sets. Table 1 shows a comparison of the different methods tested on the preterm infant connectomes for prediction of motor and cognitive scores.

Our proposed method with backbone and connectivity priors achieved the highest correlations, lowest AOCs and best 2-class classification accuracies for both motor and cognitive scores (for parameter settings, $[\lambda_{L1}, \lambda_C, \lambda_B]$ of $[2^2, 2^1, 2^6]$ and $[2^5, 2^2, 2^5]$, respectively). For 2-class classification in particular, our method outperformed Brown et al.'s method by 7.4%, Elastic-Net [7] by 8.4% and Zhu et al.'s method [8] by 17.6% higher accuracy on average. Using a two-proportion z-test, we found all these differences to be statistically significant ($p < 0.05$). Also, note that, beginning with standard linear regression, the correlation values improved as each regularization term was added. All tested methods had statistically significant ($p < 0.05$) correlations since, for $1000 \times 2 = 2000$ total predictions, the threshold for 95% significance is $r \geq 0.0439$.

Fig. 2 displays the predictive subnetworks learned by our proposed method (averaged over all rounds of cross validation). Subnetworks were stable across rounds: 93.6% of all edges were consistently in or out of the subnetwork 95% of the time. We examined the structure of the selected subnetworks to analyse the effect of the proposed regularization terms. By including the L1 regularization term, the learned subnetworks were very sparse, having an average of 71.6% and 98.2% of edge weights set to zero for motor and cognitive scores, respectively, up from only 6.7% (for either score) without the L1 term. Adding the backbone network prior reduced the number of low SNR edges (i.e., $B_{j,j} = 1$) by 18.6% percent for motor score prediction and 11.2% for cognitive score prediction. Adding the connectivity prior improved subnetwork efficiencies (a measure of network integration [5]) by a factor of 6.8 (from 0.0059 to 0.0403) and 2.2 (from 0.2807 to 0.6215) for subnetworks predictive of motor and cognitive scores, respectively.

As expected, the predictive motor subnetwork clearly includes the cortico-spinal tracts (Fig. 2a.i). The predictive cognitive subnetwork was more sparse and had generally lower weights than the motor subnetwork (as visualized by less dense, more transparent streamlines), due to the larger L1 weight used for best prediction of the cognitive scores. However, the left and right medial superior frontal gyri (SFGmed) and the connection between these two regions
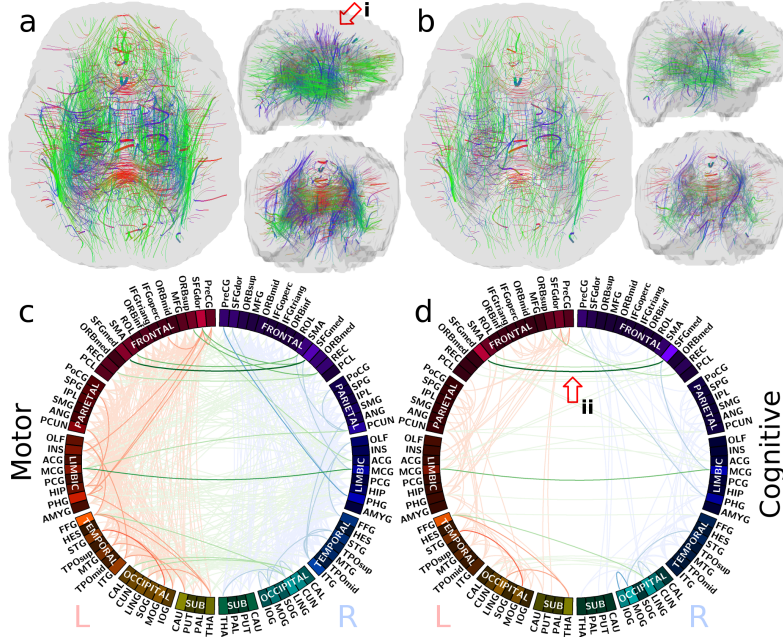
Fig. 2. (Top) Optimal weighted subnetworks for prediction of **a)** motor and **b)** cognitive outcomes. Stronger edge weights are represented with more opaque streamlines. **(Bottom)** Circos ideograms for the **c)** motor and **d)** cognitive subnetworks.

that had stronger weights (factor of 2.1) in the cognitive network than in the motor network, (Fig. 2d.ii). This is not surprising as these regions contain the presupplementary motor area which is thought to be responsible for a range of cognitive functions [19].

4 Conclusions

To better understand neurodevelopment and to allow for early intervention when poor outcomes are predicted, we proposed a framework for learning subnetworks of structural connectomes that are predictive of neurodevelopmental outcomes for infants born very preterm. We found that by introducing our novel network backbone prior, the learned subnetworks were more robust to noise by including fewer edges with low SNR weights. By including our connectivity prior, the subnetworks became more highly integrated, a property we expect for subnetworks pertinent to specific functions. Compared to other methods, our approach achieved the best accuracies for predicting both cognitive and motor scores of preterm infants, 18 months into the future.

References

1. World Health Organization. Preterm birth fact sheet no. 363. http://www.who.int/mediacentre/factsheets/fs363/en/. Accessed 2015-03-03.
2. S A Back and S P Miller. Brain injury in premature neonates: A primary cerebral dysmaturation disorder? *Annals of Neurology*, 75(4):469–486, 2014.
3. V Chau, A Synnes, R E Grunau, K J Poskitt, R Brant, and S P Miller. Abnormal brain maturation in preterm neonates associated with adverse developmental outcomes. *Neurology*, 81(24):2082–2089, 2013.
4. E Ziv, O Tymofiyeva, D M Ferriero, A J Barkovich, C P Hess, and D Xu. A machine learning approach to automated structural network analysis: application to neonatal encephalopathy. *PloS one*, 8(11):e78824, 2013.
5. C J Brown, S P Miller, B G Booth, S Andrews, V Chau, K J Poskitt, and G Hamarneh. Structural network analysis of brain development in young preterm neonates. *NeuroImage*, 101:667–680, 2014.
6. C J Brown, S P Miller, B G Booth, K J Poskitt, V Chau, A R Synnes, J G Zwicker, R E Grunau, and G Hamarneh. Prediction of motor function in very preterm infants using connectome features and LSI. *MICCAI*, 2015.
7. B C Munsell, C-Y Wee, S S Keller, B Weber, C Elger, L A T da Silva, T Nesland, M Styner, D Shen, and L Bonilha. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *NeuroImage*, 118:219–230, 2015.
8. D Zhu, D Shen, X Jiang, and T Liu. Connectomics signature for characterizaton of MCI and schizophrenia. In *ISBI*, pages 325–328. IEEE, 2014.
9. Y Ghanbari, A R Smith, R T Schultz, and R Verma. Identifying group discriminative and age regressive sub-nets from DTI-based connectivity via a unified framework of NMF and graph embedding. *MIA*, 18(8):1337–1348, 2014.
10. H Li, Z Xue, T M Ellmore, R E Frye, and S T Wong. Identification of faulty DTI-based sub-networks in autism using network regularized SVM. *Proceedings of ISBI*, 6:550–553, 2012.
11. L Grosenick, B Klingenberg, K Katovich, B Knutson, and J E Taylor. Interpretable whole-brain prediction analysis with GraphNet. *NeuroImage*, 72(2):304–321, 2013.
12. N. Bayley. *Manual for the Bayley Scales of Infant Development*. Harcourt, San Antonio, 3rd edition, 2006.
13. N V Chawla, K W Bowyer, L O Hall, and W P Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of AI Research*, 16(1):321–357, 2002.
14. M Schmidt. *Graphical model structure learning with l1-regularization*. PhD thesis, University of British Columbia (Vancouver), 2010.
15. H Cheng, Y Wang, J Sheng, W G Kronenberger, V P Mathews, T A Hummer, and A J Saykin. Characteristics and variability of structural networks derived from diffusion tensor imaging. *NeuroImage*, 61(4):1153–1164, 2012.
16. C J 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. USA*, 106(6):2035–40, 2009.
17. M A de Reus, V M Saenger, R S Kahn, and M P van den Heuvel. An edge-centric perspective on the human connectome: link communities in the brain. *Phil. Trans. R. Soc. B*, 369(1653):20130527–, 2014.
18. J Bi and K P. Bennett. Regression Error Characteristic Curves. *Proceedings of ICML-2003*, pages 43–50, 2003.
19. S Zhang, J S Ide, and C S R Li. Resting-state functional connectivity of the medial superior frontal cortex. *Cerebral Cortex*, 22(1):99–111, 2012.