

# A network-based analysis of the preterm adolescent brain using PCA and graph theory

Hassna Irzan <sup>1,2</sup>, Michael Hütel <sup>2</sup>, Carla Semedo <sup>1,2</sup>, Helen O'Reilly <sup>3</sup>, Manisha Sahota <sup>2</sup>, Sebastien Ourselin <sup>2,1</sup>, Neil Marlow <sup>3</sup>, and Andrew Melbourne <sup>2,1</sup>

<sup>1</sup>Dept. Medical Physics and Biomedical Engineering, University College London,

<sup>2</sup>Biomedical Engineering and Imaging Sciences, Kings College London,

<sup>3</sup>Institute for Women's Health, University College London

**Abstract.** The global increase in the rate of premature birth is of great concern since it is associated with an increase in a wide spectrum of neurologic and cognitive disorders. Neuroimaging analyses have been focused on white matter alterations in preterm subjects and findings have linked neurodevelopment impairment to white matter damage linked to premature birth. However, the trajectory of brain development into childhood and adolescence is less well described. Neuroimaging studies of extremely preterm born subjects in their adulthood are now available to investigate the long-term structural alterations of disrupted neurodevelopment. In this paper, we examine white matter pathways in the preterm adolescent brain by combining state-of-the-art diffusion techniques with graph theory and principal component analysis (PCA). Our results suggest that the pattern of connectivity is altered and differences in connectivity patterns result in more vulnerable premature brain network.

**Keywords:** Diffusion MRI · PCA · Graph Theory · Prematurity · Brain Efficiency · Brain Vulnerability.

## 1 Introduction

Neonatal care has transformed our ability to treat babies born at the lowest gestational ages, although increased survival has not been associated with as strong a decrease in neonatal morbidity [7]. Emerging evidence suggests that the preterm infant brain endures anatomical, microstructural and functional disruption, it is less clear, however, how these alterations extend into adulthood. This is of great concern, since the number of subjects with a wide spectrum of neurologic and cognitive disorders is increasing. This has an associated burden for society, health care systems and education. Neuroimaging studies have documented brain volume alteration [8] and impaired white matter (WM) connectivity and correlated this with low cognitive performance following preterm birth [1]. Recent studies have analysed the structural integrity of WM in adult preterm brain; the findings revealed preservation of the global connectivity at the expense of peripheral connectivity in infants [6] and adults [2]. The primary aim of this study is to identify brain sub-networks in which term

and preterm are separable and to examine the properties of such sub-networks and how critical they are to the overall functioning of the brain. The role of such sub-networks can be estimated by examining the impact on the global efficiency when inflicting damage to these sub-networks. We hypothesise that extremely premature birth affects the structure of the adolescent preterm brain and that this structural alteration negatively influences measures of brain communication efficiency and network vulnerability to attack. Generally, the WM tract architecture has been characterised by diffusion tensor imaging (DTI); more robust models are now available, which overcome some of the inherent limitations of the diffusion tensor. For example fiber tracking with constraints has improved the biological plausibility of reconstructed WM bundles [5]. Comparing all the WM pathways without an *a priori* hypothesis destroys the statistical power of the analysis because of multiple comparison. We use PCA to reduce the redundancy of the data, which produces deterministic and reproducible low-dimensional representations of the data in which variance is preserved. The resulting networks are amenable to graph theory analysis [9], which can be used to summarise complex networks and aid interpretation. Specifically, we use PCA to detect the network properties in which term and preterm are distinguishable. We further examine the networks by investigating network vulnerability to attack. The outline of the analysis is presented in Fig.1.

## 2 Methods

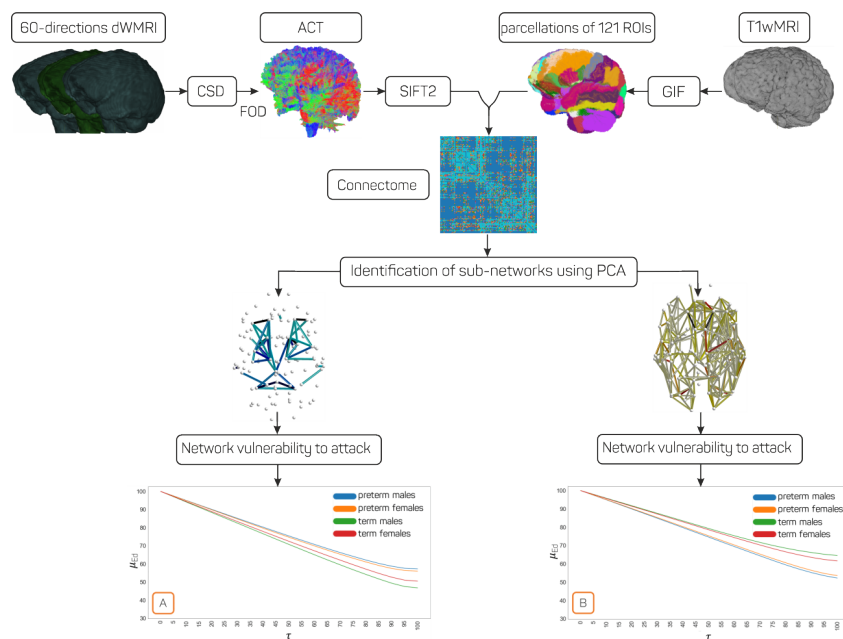
We perform data preprocessing and estimate WM tracts in section 2.1. We analyse the effect of the confounding variables on the global efficiency of the networks in section 2.2 and identify the tracts with highest variance between term born and preterm born subjects in section 2.3. To investigate the role of the sub-networks derived from PCA analysis we examine the vulnerability to attack of term and preterm brain in section 2.4.

### 2.1 Data

*Data acquisition:* T1-weighted (T1w) MRI acquisitions are acquired for a cohort of adolescents born extremely preterm (<26 weeks gestation): 39/12 females/males, term born: 20/12 females/males on a 3T Philips Achieva at TR=6.93 ms, TE=3.14 ms and 1 mm isotropic resolution. Diffusion weighted MRI (dWMRI) volumes are acquired at  $2.5 \times 2.5 \times 3$  mm resolution across  $b$ -values of (0, 300, 700, 2000)  $s/mm^2$ ,  $n=4, 8, 16, 32$  directions and TE= 70 ms. On average preterm subjects were born at  $24.9 \pm 0.8$  weeks of gestation and their average birth weight is  $728.4 \pm 126.3$  g. The age at scan of all subjects is 19 years of age.

*Data preprocessing:* We apply N4ITK algorithm for bias field correction of T1-weighted images. Subject motion in the dWMRI volumes is addressed by affinely registering them to the mean- $b_0$  volume.

*Tractography and networks extraction:*



**Fig. 1.** The presented analysis is organised in three main steps. Top: for each individual subject a structural brain network is computed. We applied Constrained Spherical Deconvolution (CSD) algorithm to estimate the Fiber Orientation Distribution (FOD) in each voxel and performed Anatomically Constrained Tractography (ACT). From Geodesic Information Flow template we consider 121 unique regions as nodes of each connectome. The connection between each node region is quantified after applying Spherical-Deconvolution Informed Filtering of Tracks (SIFT2). Middle: principal Component Analysis (PCA) is applied and discriminative sub-networks between preterm and term subjects are identified. Bottom: network vulnerability analysis is conducted by causing damage to the identified networks and quantifying the change in network global efficiency.

All analysis is performed in the original anatomical space of each subject to minimise anatomical bias in the subsequent analysis. Brain parcellations are obtained using Geodesic Information Flow (GIF) [3]. Fibre orientation distribution (FOD) is derived by means of constrained spherical deconvolution (CSD) [12]. We normalise the FODs intensity such that they are comparable across subjects and perform anatomically constrained tractography (ACT) [10] on multi-shell data [5]. We consider 121 regions-of-interest (ROIs) of a parcellation that include: neocortex, subcortical structures, cerebellum, pons and brainstem. For each subject  $i$ , we define a network  $G_i$ , in which  $J$  nodes are the individual ROIs and  $K$  edges are connections between each pair of ROIs equivalent to streamline count weighted by the cross-sectional area of the fibre (SIFT2) [11]. We remove the effect of network strength on global efficiency by normalising each network  $G_i$  by its total strength.

## 2.2 Statistical analysis

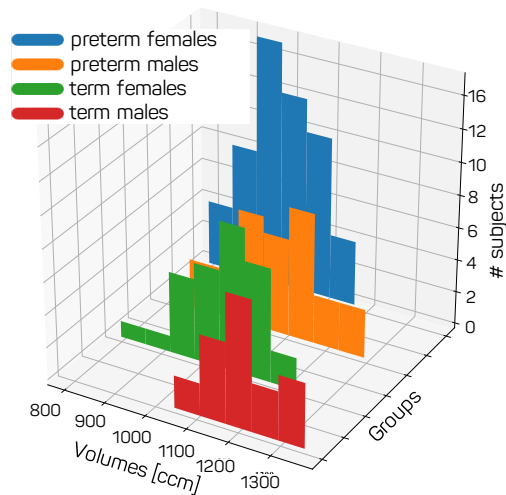
Group, gender and brain size may present a confounding effect in the analysis. There is correlation between brain efficiency and weeks of gestation [2]. In addition, as illustrated in Fig.2, there is difference in brain size between females and males as well as between preterm and term born subjects [4]. As such, to establish the influence of brain size, gender and group in the analysis, we test their statistical influence in the network efficiency. The statistical model is  $E = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ , where  $\beta_0$  is the intercept,  $X_1$ ,  $X_2$  are regressors encoding for group and gender membership respectively,  $X_3$  the brain size values and  $E$  the network global efficiency.

## 2.3 Identification of sub-networks using PCA

We apply PCA to the set of edges of  $N$  subjects networks  $\{G_i, i = 1 \dots N\}$  to obtain the first  $C$  eigenvectors  $\mathbf{E} \in R^{K \times C}$  that approximate 90% of the variance in all edges  $\mathbf{X} \in R^{N \times K}$ .  $\mathbf{X}$  is projected onto the  $C$  eigenvectors to obtain the reduced data  $\mathbf{X}_{\text{red}} = \mathbf{X}\mathbf{E}$ . We then investigate the capability of each eigenvector  $E_c$  to separate term from preterm subjects. We test the statistical significance of this by running the Kruskal-Wallis test because it has no assumptions about the distribution of the data. To compensate for multiple comparisons, we test each individual null hypothesis at a significance level of  $\alpha/C$  (Bonferroni correction), with  $\alpha = 0.05$ .

## 2.4 Network vulnerability to attack

We investigate how damaging some nodes or edges impacts the appropriate functioning of a network, which we model as the capacity of information transfer. We associate the network performance with its global efficiency  $E_G$ . We define the efficiency decay as  $E_d = E_{Gd}/E_G$  with  $E_G$  and  $E_{Gd}$  the global efficiency of the intact and damaged network, respectively. The damage to network  $G_d$  is carried out by setting the weights of the targeted connection  $w_{ij}$  to



**Fig. 2.** Brain volume distribution for preterm born females, preterm born males, term born females and term born males.

$w_{ij}^d = w_{ij} \times (1 - \tau/100)$  [13], where  $\tau$  is the inflicted damage ranging from a minimum of 0 (no damage) to a maximum of 100 (full damage). We test the brain network vulnerability by performing targeted attacks to the sub-network that best separates preterm from term subjects. We carry vulnerability analysis for each network  $G_i$  and evaluate the mean  $E_d$  ( $\mu_{E_d}$ ) value for each group. The main aim of these tests is not only to identify the most critical structures or connections for the network performance, but to investigate if the preterm born network is more or less vulnerable to attack and which brain regions or network are more critically impacted.

### 3 Results

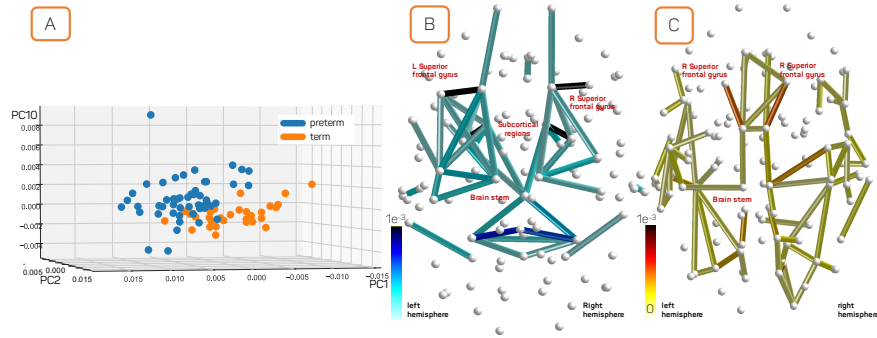
#### 3.1 Statistical analysis

With respect to the network global efficiency  $E$ , the statistical analysis indicated that there was not significant statistical influence of gender ( $p = 0.253$ ) and brain size ( $p = 0.064$ ). The group membership was statistically significant ( $p = 0.010$ ).

#### 3.2 Identification of sub-networks using PCA

Forty principal components (PCs) explained 90% of the variation in the dataset. The PCs with discriminatory power between term and preterm were 1<sup>st</sup> ( $p = 2.02e^{-6}$ ), 2<sup>nd</sup> ( $p = 5.82e^{-4}$ ) and 10<sup>th</sup> ( $p = 2.93e^{-5}$ ). All PCs are statistically significant after Bonferroni correction. Fig. 3.A captures the separation between term and preterm born subjects along these PCs. The discriminatory sub-network between the preterm and term group is obtained as  $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_{10} v_{10}$ .

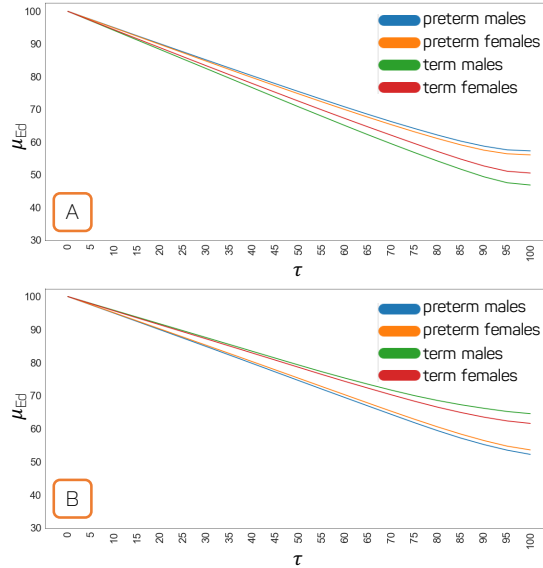
Where, the linear combination of the eigenvectors ( $v_1, v_2, v_{10}$ ) corresponding to the 1<sup>st</sup>, 2<sup>nd</sup>, and 10<sup>th</sup> PCs is weighted by the mean contribution of the preterm subjects in each PC ( $\lambda_1, \lambda_2, \lambda_{10}$ ). Fig. 3.B illustrates the top 25% negative values and Fig. 3.C the top 25% positive values. The highest weights are found in connections of bilateral middle and superior frontal gyrus, bilateral precentral gyrus, bilateral thalamus and putamen, bilateral caudate, right supplementary motor cortex. The connectivity strength in the sub-network Fig.3.B is significantly reduced ( $p = 4.68e^{-5}$ ) in the preterm subjects ( $\mu = 2.14e^{-2} \pm 5.9e^{-3}$ ) compared to term subjects ( $\mu = 2.80e^{-2} \pm 5.7e^{-3}$ ). Preterm group ( $\mu = 2.33e^{-2} \pm 5.41e^{-3}$ ) have a comparable ( $p = 0.11$ ) connectivity strength in the sub-network in Fig.3.C with term born group ( $\mu = 2.15e^{-2} \pm 5.0e^{-3}$ ).



**Fig. 3.** (A) The three PCs with discriminatory power between term born (females and males) and preterm born (females and males) subjects. The corresponding eigenvectors reveal the underlying sub-networks: (B) sub-network from the lowest 25% values shows connections from the fronto-parietal areas to the deep grey matter regions and (C) sub-network from the highest 25% values displays a widespread connectivity.

### 3.3 Network vulnerability to attack

The effect of targeted attack to the sub-networks in Fig.3 (B and C) is shown in Fig.4 (A and B respectively) from null (0) to full (100) damage. Fig.4.A shows that targeted attack to the connections in sub-network in Fig.3.B results in a slight reduction of global efficiency for the full term groups more than the preterm groups; this is more pronounced in the full term males. The decay of global efficiency when damaging the sub-network in Fig. 3.C shows higher vulnerability in the preterm groups (males and females) with respect to the control group (about 12.8% less).



**Fig. 4.** The effect on the brain of progressively increasing the inflicted damage from a minimum of no damage  $\tau = 0$  to a maximum of full damage  $\tau = 100$ : (A) The effect on the brain network of damaging sub-network in Fig.3.B and (B) sub-network in Fig.3.C.

## 4 Discussion

This study investigates differences in structural connectivity between 19 year-old adolescents born extremely preterm (born before 26 weeks completed gestation) and their socio-economically matched term-born peers.

The main finding is that the connectivity in the preterm brain is altered leading to a more vulnerable network. The strength is reduced in the network in (Fig. 3.B), while it is preserved in the connections in Fig. 3.C. Furthermore, the vulnerability analysis suggests that the full term subjects are relatively more vulnerable in the sub-network in between fronto-parietal and subcortical structures (Fig. 3.B). In this respect, it is more plausible that this sub-network is preserved in preterm subjects. However, when damaging the network in Fig. 3.C, the preterm subjects showed higher vulnerability regardless of gender. This supports the hypothesis of a widespread vulnerability to attack in the preterm brain network suggesting an overall fragile network.

Brain connectivity is affected by brain volume, however the functional relationship between brain size and streamline count is not known [4]. In our cohort, there are differences in average brain volume between groups (Fig.2). Therefore, we analysed the influence of gender, group and brain size on the brain global efficiency; the results showed that only prematurity condition affects the global efficiency ( $p = 0.010$ ). Hence, we compared term born (females and males) and

preterm born (females and male) subjects in the analysis.

The interpretation of graph metrics makes assumptions about the network. For example, the network global efficiency definition assumes that the signals propagate in the network along the shortest paths between network elements. Although this assumption is sensible for artificial networks, there is no tangible evidence that such hypothesis holds for the human brain which may have different optimisation criteria. Graph theory analysis of networks is also sensitive to the parcellation scheme; a more detailed analysis could result from a higher resolution network. Here, the PCA analysis is biased towards the average preterm network as there are more preterm born subjects. To mitigate this effect we ran the same analysis with matched sample size of term and preterm born subjects and observed the same results.

This work has shown differences in the connectivity of the adolescent brain network after preterm birth; such abnormalities resulted in a network more vulnerable to network damage. Future work will aim to detect the nature of such alterations by analysing white matter microstructure at the voxel level and within individual brain regions linked to the sub-networks we have identified (hubs and peripheral regions). In addition, we aim to investigate the neuropsychological correlates of the sub-networks that we have established.

*Acknowledgements:* We acknowledge the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1), the National Institute for Health Research (NIHR) and the MRC (MR/J01107X/1).

## References

1. Gareth Ball et al. Thalamocortical connectivity predicts cognition in children born preterm. *Cerebral cortex (New York, N.Y. : 1991)*, 25(11):4310–4318, 11 2015.
2. Dafnis Bataille et al. Early development of structural networks and the impact of prematurity on brain connectivity. *NeuroImage*, 149:379 – 392, 2017.
3. M. J. Cardoso et al. Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. *IEEE Transactions on Medical Imaging*, 34(9):1976–1988, Sep. 2015.
4. Jürgen Hänggi et al. The hypothesis of neuronal interconnectivity as a function of brain size—a general organization principle of the human connectome. *Frontiers in human neuroscience*, 8:915; 915–915, 11 2014.
5. Ben Jeurissen et al. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage*, 103:411 – 426, 2014.
6. Vyacheslav R Karolis et al. Reinforcement of the brain’s rich-club architecture following early neurodevelopmental disruption caused by very preterm birth. *Cerebral cortex (New York, N.Y. : 1991)*, 26(3):1322–1335, 03 2016.
7. Tamanna Moore et al. Neurological and developmental outcome in extremely preterm children born in england in 1995 and 2006: the epicure studies. *BMJ : British Medical Journal*, 345:e7961, 12 2012.
8. Chiara Nosarti et al. Adolescents who were born very preterm have decreased brain volumes. *Brain*, 125(7):1616–1623, 07 2002.
9. Mikail Rubinov et al. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52:1059–1069, 2010.



10. Robert E. Smith et al. Anatomically-constrained tractography: Improved diffusion mri streamlines tractography through effective use of anatomical information. *NeuroImage*, 62(3):1924 – 1938, 2012.
11. Robert E. Smith et al. Sift2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage*, 119:338 – 351, 2015.
12. J-Donald Tournier et al. Robust determination of the fibre orientation distribution in diffusion mri: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459 – 1472, 2007.
13. Martijn P. van den Heuvel et al. Aberrant frontal and temporal complex network structure in schizophrenia: A graph theoretical analysis. *The Journal of Neuroscience*, 30(47):15915, 11 2010.