

Min-cut Max-flow for Network Abnormality Detection: Application to Preterm Birth

Hassna Irzan^{1,2}, Lucas Fidon², Tom Vercauteren², Sebastien Ourselin², Neil Marlow³, and Andrew Melbourne^{2,1}

¹ Dept. Medical Physics and Biomedical Engineering, University College London

² School of Biomedical Engineering and Imaging Sciences, Kings College London

³ Institute for Women's Health, University College London

Abstract. Neuroimaging studies of structural connectomes typically average the data from many subjects and analyse the average properties of the resulting network. We propose a new framework for individual brain-network structural abnormality detection. The framework uses a graph-based anomaly detection algorithm that allows to detect abnormal structural connectivity on a subject level. The proposed method is generic and can be adapted for a broad range of network abnormality detection problems. In this study, we apply our method to investigate the integrity of white matter tracts of 19-year-old extremely preterm born individuals. We show the feasibility to cast the network abnormality detection problem into a min-cut max-flow problem, and identify consistent abnormal white matter tracts in extremely preterm subjects, including a shared network involving the bilateral thalamus and frontal gyri.

Keywords: Prematurity · Brain-network abnormality detection · Diffusion-weighted MRI · Min-cut Max-flow · Grey matter

1 Introduction

The neuropsychological outcome and neuroimaging phenotype of preterm-born children and infants is greatly influenced by premature exposure to the extrauterine environment [2,3,9]. Studies have shown anatomical [2], micro-structural [3] and a range of neuropsychological differences [9] linked to prematurity. The majority of neuroimaging studies on preterm populations analyse the average properties of the preterm group compared to that of the controls. Such population-level studies usually register all images of a population into a common space. Due to the anatomical abnormalities associated with the preterm brain, such as ventriculomegaly, this step can lead to misleading results. Generally, analysis that contrast two populations inherently ignore considerable inter-subject heterogeneity in each group. In the present work, we acknowledge that besides the normal individual variability, preterm birth has a broad range of effects on the brain. The most-reported impairments in the preterm born population are negative neuropsychological outcome [9] and white matter (WM) differences [5] with a varied pattern of severity.

Researchers’ efforts in investigating WM led to developing brain structural connectomes (or networks) as a promising tool to investigate major brain pathways and examine essential circuits [12].

The connectome can be studied as a graph where the nodes are the brain regions, and the edges quantify inter-connectivity between those areas. The main aim of the present paper is to detect abnormal WM connectivity. Labels for WM abnormalities are difficult to obtain. However, we hypothesise that abnormalities in edge connectivity are unlikely to be isolated and that conversely, anomalies are likely to be contiguous. Graph-based methods allow us to model this anatomical hypothesis and to detect abnormalities even in the absence of ground-truth labels.

The min-cut max-flow framework is applied to investigate the integrity of WM tracts of 19-year-old extremely preterm individuals (born before 27 weeks completed gestation). We analyse WM connectivity in the individual subject space by mapping the tracts into structural connectomes; then we separate aberrant connectivity from the aged-matched control-group connectivity using the min-cut max-flow framework. The framework takes into account the anatomical information of the WM tracts and brain regions to which they are connected. We use a graph similarity measure based on a Laplacian matrix to measure the global differences in structural connectivity between a reference connectivity matrix and the connectivity matrix under investigation. While the distance matrix estimates a global measure of divergence, the min-cut max-flow framework localises the abnormality. We experimentally show that the proposed framework can detect consistent abnormal WM tracts across the subjects, and the abnormal WM tracts identified for each subject correlate with the changes in structural connectivity as measured by the graph similarity measure.

2 Methods

We describe the data in section 2.1 and the steps to perform tractography and network extraction in section 2.2. Section 2.3 describes the measure we use to quantify the distance between two brain-networks, while section 2.4 describes the min-cut max-flow formulation to detect brain-network abnormalities. Figure 1 illustrates the main steps of the pipeline.

2.1 Data

Diffusion weighted MRI and T1-weighted MRI acquisitions were performed on a 3T Philips Achieva system for $N_p = 80$ (49/31 females/males) extremely preterm born 19-year-old individuals and $N_c = 36$ (19/17 females/males) full-term born age-matched peers. T1-weighted MRI images were acquired at TR=6.93 ms, TE=3.14 ms and 1 mm isotropic resolution. EPI-SE volumes of dWMRI were 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, TE: 70 ms, TR: 3500 ms, FOV: $(240 \times 240 \times 150)$ mm, flip angle: 90° , and SENSE factor of 1.

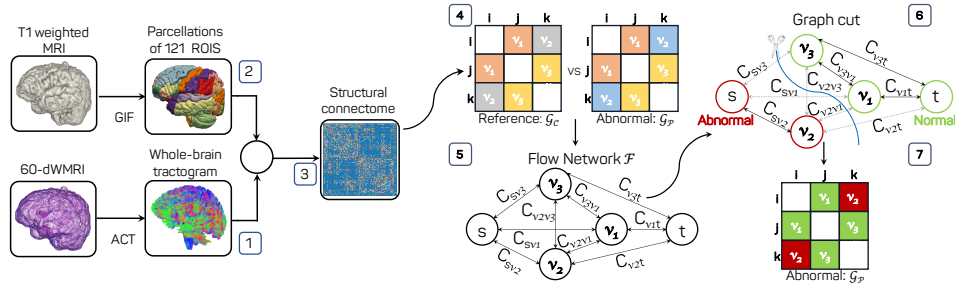


Fig. 1. Outline of the methodology to estimate structural networks and find abnormal structural connections. We performed Anatomically Constrained Tractography (ACT) (1) [10] to estimate the white matter streamlines. We estimate the structural connectome (3) by quantifying the connectivity between brain regions obtained from Geodesic Information Flow (GIF) [4] parcellations (2). We compare the structural connectivity matrix of each extremely preterm born subject \mathcal{G}_P to the average structural connectivity matrix of the full-term born subjects \mathcal{G}_C (4); similar edges have the same colour, the colour of the abnormal edge in \mathcal{G}_P is different to the corresponding edges in \mathcal{G}_C (4). We cast the problem into a min-cut max-flow framework (5). The detection of structural brain abnormality (7) results from the graph cut after solving the min-cut max-flow problem (6).

T1-weighted images were bias-corrected using N4ITK algorithm [14]. Diffusion-weighted MRI volumes were corrected for thermal noise [15], Gibbs-ringing artefacts [7], eddy current-induced distortion and subject movements artefacts [1].

The median gestational age at birth for extremely preterm born individuals is 25.14 (CI 95% 22.14 – 25.86) weeks of gestation. The full term born subjects were born after 37 weeks of gestation. All the subjects had MRI assessment at 19 years of age.

2.2 Tractography and networks extraction

As shown in step 1 of figure 1, we generate a whole-brain tractogram for each subject. A multi-shell multi-tissue approach [6] was used to estimate the response function for each tissue type. The fibre orientation distribution (FOD) was first calculated in each voxel using constrained spherical deconvolution (CSD) [13] and then normalised for inter-subject comparisons. Anatomically Constrained Tractography (ACT) was performed using dynamic seeding and backtrack re-tracking algorithms [10]. To account for the fact that the density of the estimated fibres is not representative of the density of the underlying white matter fibres, the ten million streamlines generated per subject were filtered using the spherical-deconvolution informed filtering of tracks (SIFT2) procedure [11]. SIFT2 determines an adequate cross-sectional area for each estimated streamline, such that the estimated streamlines densities throughout the white matter are reflective of the fibre densities computed using the spherical deconvolution

model [11]. As illustrated in figure 1, in step 2, tissue parcellations of the corrected T1-weighted volumes were obtained using Geodesic Information Flow (GIF) [4]. Brain regions of interest (ROI) were defined based on the GIF labelling protocol [4]. The grey matter areas (121 brain regions) formed the nodes for the brain network derivation. For each subject $n \in N$, a network $\mathcal{G}_n = (Q, Z)$ is defined, in which each node corresponds to an ROI, and Z is the set of edges connecting the ROIs. In the network \mathcal{G}_n , for all of edges (i, j) , we denote g_{ij} the strength of the connectivity between i and j defined by the weighted (SIFT2) [11] contribution of each streamline connecting i and j . This is illustrated in step 3 of figure 1.

2.3 Graph similarity measure

We aim to evaluate the divergence of the connectivity matrix of each extremely preterm born subject from the normality. The graph similarity is quantified using the spectral distance (SD) [16] of the normalised Laplacian. The eigenvalues of the normalised Laplacian describe aspects of the global network structure. The difference between the spectra of normalised Laplacians can be used to quantify the similarity between networks. The normalised Laplacian \mathcal{L} of a graph \mathcal{G} with edge weights g_{ij} is defined as $\mathcal{L} = \mathcal{I} - \mathcal{D}^{-\frac{1}{2}}\mathcal{G}\mathcal{D}^{-\frac{1}{2}}$ [16], where \mathcal{I} is the identity matrix and \mathcal{D} is a diagonal matrix such that $\mathcal{D} = \text{diag}(d_i)$ with $\forall i d_i = \sum_{j \in \mathcal{Q}} g_{ij}$. To avoid the use of arbitrary control connectivity matrices, we consider the mean connectivity matrix \mathcal{G}_C of the full-term born subjects and the corresponding normalised Laplacian \mathcal{L}_C . Let \mathcal{L}_P be the Laplacian of the connectivity matrix \mathcal{G}_P of the extremely preterm born subject, the spectral distance $SD(\mathcal{L}_C, \mathcal{L}_P)$ is defined as the Euclidean distance between the eigenvalues of \mathcal{L}_C and \mathcal{L}_P [16]

$$SD(\mathcal{L}_C, \mathcal{L}_P) = \sum_u \sqrt{(\lambda_u^C - \lambda_u^P)^2} \quad (1)$$

Therefore we aim to measure the normality of \mathcal{G}_P using $SD(\mathcal{L}_C, \mathcal{L}_P)$.

2.4 Graph cut optimisation for the detection of abnormal connectivity

This section shows how the problem of detecting anomalies in subject's connectivity network can be cast as a min-cut max-flow problem.

Min-cut max-flow framework: Given the group-level reference connectivity matrix \mathcal{G}_C and the subject-level abnormal connectivity matrix \mathcal{G}_P (as illustrated in step 4 of figure 1), we aim to identify the abnormal connectivity in \mathcal{G}_P with respect to \mathcal{G}_C . As each edge (i, j) connects brain regions i and j with strength of connectivity g_{ij} , by comparing the strength of connectivity g_{ij} in the reference connectivity matrix \mathcal{G}_C and the abnormal connectivity matrix \mathcal{G}_P , we aim to separate abnormal edges from normal edges.

We cast this problem into a min-cut max-flow framework [8]. A max-flow framework involves a fully connected bi-directed graph $\mathcal{F} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}|$ nodes and $|\mathcal{E}|$ directed edges connecting them. An edge $(i, j) \in \mathcal{Z}$ in the connectivity matrix \mathcal{G}_n is a node $v \in \mathcal{V}$ in the graph \mathcal{F} as illustrated in step 5 of figure 1. The graph \mathcal{F} has two additional nodes: the source node $s \in \mathcal{V}$ and the sink node $t \in \mathcal{V}$. Each edge of \mathcal{F} has a fixed and non-negative capacity C which is the maximum flow that edge can handle. The graph \mathcal{F} has three types of edges. Namely: 1) the edges that connect the source node $s \in \mathcal{V}$ to the nodes $v \in \mathcal{V}$ with capacity $C_{s,v}$, the edges that link the sink node $t \in \mathcal{V}$ to the nodes $v \in \mathcal{V}$ with capacity $C_{t,v}$ and the edges that connect the nodes $v \in \mathcal{V}$ between each other with capacity $C_{v1,v2}$. The source node s and the sink node t are not directly connected. According to the max-flow min-cut theorem, the maximum flow from the source node $s \in \mathcal{V}$ to the sink node $t \in \mathcal{V}$ corresponds to the minimum total capacities of the edges, which, if removed, would partition the graph \mathcal{F} into two subsets: the abnormal nodes set \mathcal{S} and the normal nodes set \mathcal{T} .

In order to reflect the similarity between the edge $(i, j) \in \mathcal{G}_C$ and the corresponding edge $(i, j) \in \mathcal{G}_P$, we define the capacity $C_{t,v}$ as the Gaussian similarity function between the edge weights $g_{ij}^{\mathcal{G}_C}$ and $g_{ij}^{\mathcal{G}_P}$

$$C_{t,v}(g_{ij}^{\mathcal{G}_C}, g_{ij}^{\mathcal{G}_P}) = K \cdot \exp\left(\frac{-\left(g_{ij}^{\mathcal{G}_C} - g_{ij}^{\mathcal{G}_P}\right)^2}{2\sigma^2}\right) \quad (2)$$

where K is an arbitrary multiplicative constant. The capacity $C_{s,v}$ of the edges connecting the source node $s \in \mathcal{V}$ with the nodes $v \in \mathcal{V}$ is set to $K - C_{t,v}(g_{ij}^{\mathcal{G}_C}, g_{ij}^{\mathcal{G}_P})$. The capacity $C_{s,v}$ reflects the extent to which the strength of connectivity associated with the edge $(i, j) \in \mathcal{G}_P$ is abnormal while the capacity $C_{t,v}$ reflects the degree to which the strength of connectivity associated with the edge $(i, j) \in \mathcal{G}_P$ is normal.

In addition, if one brain region has an abnormal edge connection (i, j) , then the likelihood it has other abnormal edge connections is high. Sporadic abnormal connections are more likely to be due to noise or error in the streamline reconstruction. Pair of edges of the form (i, j) and (i, y) that are connected to the same brain region i are considered as neighbours in \mathcal{G}_n . To account for that in the graph \mathcal{F} , the capacity of the edges $C_{v1,v2}$ is set to a positive constant value M if the two nodes represent two neighbouring edges in the connectivity matrices \mathcal{G}_n and zero otherwise. Therefore, partitioning the graph into sets \mathcal{S} and \mathcal{T} maximises

$$E = \sum_{v \in \mathcal{S}} K - C_{t,v}(g_{ij}^{\mathcal{G}_C}, g_{ij}^{\mathcal{G}_P}) + \sum_{v \in \mathcal{T}} C_{t,v}(g_{ij}^{\mathcal{G}_C}, g_{ij}^{\mathcal{G}_P}) - \sum_{v_1 \in \mathcal{S}, v_2 \in \mathcal{T}} C_{v1,v2} \quad (3)$$

which is solved using highest-label preflow-push algorithm.

Parameter tuning: The average full-term connectivity matrix is the reference connectivity matrix \mathcal{G}_C . The parameters of the graph \mathcal{F} are K, σ and M . As

K and M are dependent, we fix $K = 10^4$ and determine σ and M using a grid search such that when the reference connectivity matrix \mathcal{G}_C is compared to the connectivity matrix of each full-term born subject, the graph-cut framework identifies minimal abnormal edges E_c ; when the reference connectivity matrix \mathcal{G}_C is compared to the connectivity matrix of each extremely preterm subjects, the graph-cut identifies the maximum number of abnormal edges E_p . Therefore, the best σ and M maximise the quantity $E_p - E_c$.

Framework evaluation: To assess the performance of the min-cut max-flow framework, we evaluate two aspects: 1) the consistency of the identified abnormal edges across the subjects and 2) the consistency of the number of the identified edges E_p with respect to the graph similarity measure $SD(\mathcal{L}_C, \mathcal{L}_P)$. We expect the number of the identified abnormal edges to correlate with the similarity measure $SD(\mathcal{L}_C, \mathcal{L}_P)$, as higher $SD(\mathcal{L}_C, \mathcal{L}_P)$ indicates stronger structural deviation from the reference matrix \mathcal{G}_C . The identified edges constitute a sub-network for each subject. It is expected that the identified sub-networks show two characteristics: 1) a general pattern that is shared between the extremely preterm subjects as being born extremely preterm might induce similar brain abnormalities, and 2) a distinctive one that is characteristic to individual subjects as result of individual variability. To investigate how the identified sub-networks cluster across the extremely preterm subjects, we use principal component analysis (PCA) to derive a low-dimensional set of features X_{red} that represent the original abnormal sub-networks X . We apply PCA to the set of edges $X \in R^{N_p \times |Z|}$ in which for each extremely preterm subject, we set the edges that have been identified as abnormal to 1 and 0 otherwise.

Simulation: The weight of each edge is proportional to the WM connecting the corresponding brain regions. A WM abnormality is a reduction or increase in the weight of the edges with respect to the reference group (full term born subjects). Since reduction in WM connectivity is a characteristic of the preterm brain phenotype [5], we simulate abnormalities of WM connectivity by reducing weights in \mathcal{G}_P with respect to \mathcal{G}_C . To show the ground-truth link between SD and the percentage of edges identified abnormal edges. We consider \mathcal{G}_C as a 15×15 zero-diagonal symmetric matrix of ones, and consider \mathcal{G}_P as a 15×15 matrix with synthetic abnormalities. In the beginning, \mathcal{G}_P is identical to \mathcal{G}_C , then the abnormalities are induced by randomly reducing the weights associated with the edges connected to the same node in \mathcal{G}_P . In each iteration an additional node is reduced at random by 90% to 100% until 99% of all the edges are reduced. This simulation has been carried out on 6 pairs of \mathcal{G}_C and \mathcal{G}_P . We used the same K , σ and M parameters as in real data.

3 Results

The grid search over the parameters σ and M that minimise the number of abnormal edges E_c detected in the full-term born subjects and maximise the

number of abnormal edges E_p identified in the extremely preterm born group shows that the best values are $\sigma = 0.01$ and $M = 51$.

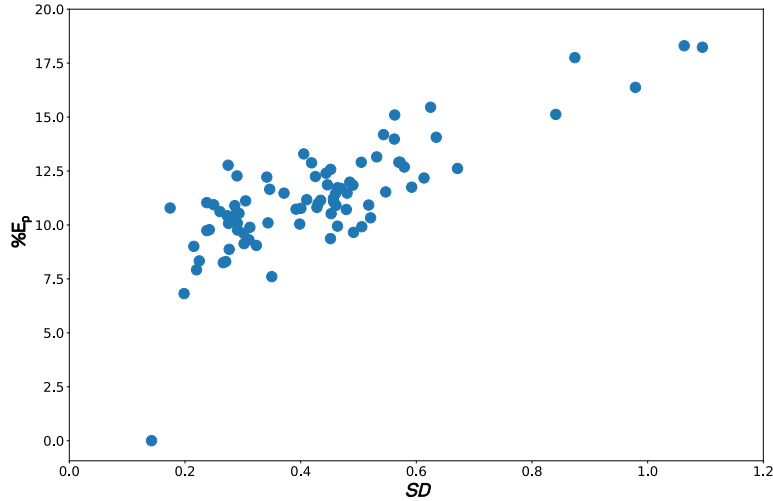


Fig. 2. Visualisation of the relationship between the spectral distance $SD(\mathcal{L}_C, \mathcal{L}_P)$ and proportion of abnormal edges $\%E_P$ identified for each extremely preterm subject. The correlation between $SD(\mathcal{L}_C, \mathcal{L}_P)$ and the proportion of abnormal edges is $\rho = 0.71$ with p -value = 10^{-14} .

Figure 2 shows the relationship between SD and the ratio of abnormal to normal edges in each extremely preterm subject. The correlation between SD and the total number of abnormal edges E_p is statistically significant (p -value = 10^{-14}) with a correlation coefficient of $\rho = 0.71$. On average $11.3\% \pm 2.5\%$ of the total edges in extremely preterm subjects have been identified as abnormal. Figure 3 shows the most identified edges (in 98% of the extremely preterm subject). These edges form a sub-network related to WM connecting mainly frontal cortex and deep grey matter regions such as bilateral thalamus and bilateral frontal gyrus. Moreover, the pattern of the identified sub-network shows hemispheric symmetry. Figure 4 shows a plot of the identified abnormal sub-networks with respect to the first and second principal components. In addition, the subjects have been colour-coded with respect to the percentage of abnormal edges that were identified for that subject. Figure 4 shows that the subjects with higher percentage of abnormal edges form different clusters.

Simulation: The figure in the supplementary material shows the results for the relationship between SD and the proportion of abnormal edges that have been detected in the connectivity matrix with synthetic abnormalities. The figure displays the results for 6 pairs of \mathcal{G}_P and \mathcal{G}_C . The mean correlation coefficient

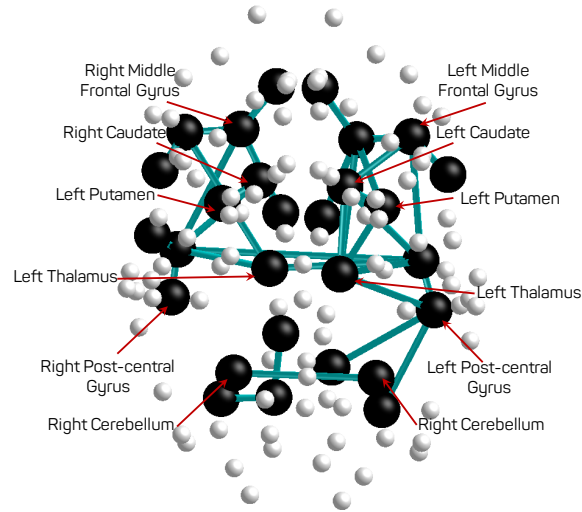


Fig. 3. Visualisation of the most common identified edges in the extremely preterm subjects. The connections in green are the abnormal edges, the brain regions in black are connected by abnormal edges while brain regions in white do not have abnormal edges.

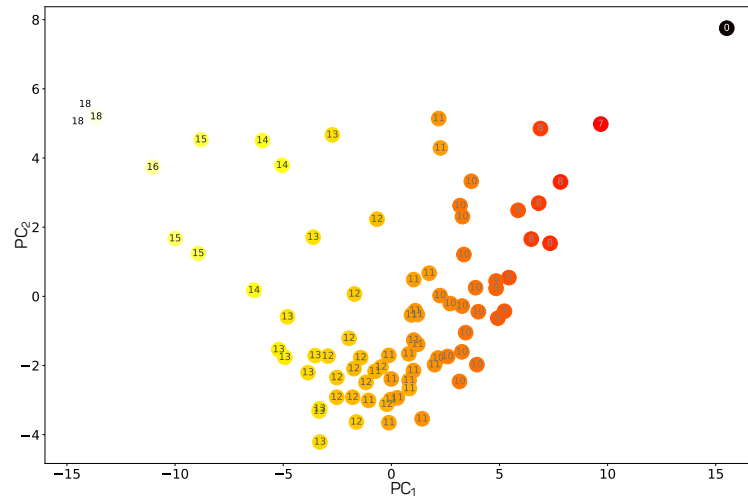


Fig. 4. Visualisation of the PCA results. The reduced sub-network data is plotted with respect to the first and second principal component. To visualise the data, an annotated colour scale is used to represent the percentages of abnormal edges found in each subject, with black being the least and white being the most.

between the SD and the proportion of abnormal edges is $\rho = 0.98$ (p -value = 4×10^{-9}). This suggests that the SD metric summarised the amount of local abnormalities present in the connectivity matrix.

4 Discussion

We propose a new framework for brain-network structural abnormality detection. The framework is based on a min-cut max-flow algorithm and aims to detect abnormal structural connectivity at the subject level.

The results show agreement between the graph similarity measure (SD) and the number of abnormal edges E_p both on the real data (figure 2) and on the data with simulated abnormalities (figure in the supplementary material). This indicates that the framework can detect the number of abnormal edges that is proportional to how different is the individual connectivity matrix with respect to the reference connectivity matrix. In the case of real data, the most identified sub-network across the extremely preterm born subjects (figure 3) demonstrates that there is consistency across the subjects as the abnormal WM connectivity is distributed between the deep grey matter regions and the frontal cortex. These results are consistent with previous findings in extremely preterm neonatal population [3] and in extremely preterm adolescents [5]. The agreement between these findings suggests that WM connectivity in these brain areas is vulnerable to extreme preterm exposure to the extra-uterine environment. Moreover, it appears that the extremely preterm brain at adolescence does not recover from early-life WM injury. However, it is still unclear whether the WM alteration represents a developmental delay or permanent damage. Further analysis of older preterm samples needs to be performed.

Figure 4 shows clusters of identified sub-networks suggesting that the identified sub-networks have a variable pattern. The subjects with similar abnormal sub-networks form clusters. The clustering along the first principal component might be driven by the number of abnormal edges, while the clustering along the second principal component is more subtle. In general, it seems that the degree to which each extremely preterm born subject has been affected by extremely preterm exposure to the extrauterine environment is variable. In the future, it would be interesting to analyse how this translates to the varied cognitive outcome of these subjects [9].

In this study, we demonstrated for the first time the feasibility of casting the network abnormality detection problem into a min-cut max-flow problem. This method is able to detect abnormal connectivity at an individual level, compared to conventional group-wise comparisons. Although the method was employed to analyse abnormal structural connectivity in extremely preterm subjects, it can be extended to detect abnormal functional connectivity. This could be of great relevance to other conditions such as dementia or autism. Moreover, the proposed framework could be applied to a broad range of network abnormality detection beyond the proposed medical application.

Acknowledgements: This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1). We would like to acknowledge the MRC (MR/J01107X/1) and the National Institute for Health Research (NIHR).

References

1. Jesper L. R. Andersson and Stamatios N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *NeuroImage*, 125:1063–1078, 2016.
2. Gareth Ball et al. The effect of preterm birth on thalamic and cortical development. *Cerebral cortex (New York, N.Y. : 1991)*, 22(5):1016–1024, 05 2012.
3. 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.
4. 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.
5. H. Irzan, M. Hütel, C. Semedo, H. O’Reilly, M. SahotaS, S. Ourselin, N. Marlow, and A. Melbourne. A network-based analysis of the preterm adolescent brain using pca and graph theory. In *2019 MICCAI International Workshop on Computational Diffusion MRI (CDMRI)*, 2019.
6. Ben Jeurissen et al. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage*, 103:411 – 426, 2014.
7. Elias Kellner et al. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*, 76(5):1574–1581, 2019/09/10 2016.
8. D. R. Fulkerson L. R. Ford. *Flows in networks*. Rand Corporation Research Studies Series. PUP, 1962.
9. Helen O’Reilly et al. Neuropsychological outcomes at 19 years of age following extremely preterm birth. *Pediatrics*, 145(2):e20192087, 02 2020.
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. Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: A structural description of the human brain. *PLoS computational biology*, 1(4):e42–e42, 09 2005.
13. 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.
14. N. J. Tustison et al. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.
15. Jelle Veraart et al. Denoising of diffusion mri using random matrix theory. *NeuroImage*, 142:394–406, 2016.
16. Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.