1   Summary of SWU4 Study

The *SWU4* dataset is a collection of neurological images (anatomical, resting state fMRI, and DTI) of 227 students attending Southwest China Normal University. Participants were recruited from the SWU campus and were imaged twice roughly one year apart. The data was made public by the Consortium for Reliability and Reproducibility [5] and later expanded as part of the Southwest University Longitudinal Imagining Multimodal (SLIM) dataset. The data used in this report was accessed through Nuerodata's *MRI-Cloud* project[1] which produces connectomes- network representations of brain structure and function- from neurological imaging data using NeuroData's MRI to Graph (*NDMG*) pipeline. *NDMG* was constructed in an attempt to remove the need to construct custom pipelines for processing imagining data which consequently may jeopardize the reproducibility of studies that rely on these custom tools. We refer the reader to Appendix B of Kiar et al.[2] for a complete description of the methods used in each stage of this pipeline.

In this report, we restrict our analysis to DTI (diffusion tensor imaging) images. DTI is a particular type of diffusion Magnetic Resonance Imaging (dMRI) that has been used to map the diffusion of water molecules within neural tissue. By identifying these diffusion patters, DTI can be used in identifying systems, called *nerve tracts*, that connect different nuclei of the central nervous system[4]. Vertices in this area are defined according to a practitioner specified *parcellation* - a mapping that assigns regions, called *voxels*, of the brain to vertices of the network based on functional, anatomical, or spatial features. Here, we use the CC200 parcellation [1] which defines a vertex set V of size $N_V = 200$. Edge weights are then defined according to the number of tracts that connect these regions. In particular, the edge weight between vertices v_i and v_j is given by

$$w(v_i, v_j) = \sum_{u \in N(v_i)} \sum_{w \in N(v_j)} I(F_{u,w}) \tag{1}$$

where $N(v) = \{\text{voxels in } v\}$ is the set of voxels assigned to vertex v by the parcellation and where $F_{u,w}$ is true if there is a tract that connects voxel u and voxel w. In this definition, we implicitly define the edge set E to be the pair vertices that share tracts between voxels in their respective neighborhoods. The *NDMG* pipeline provides these network summaries of the DTI images for each of the 227 students at two imaging sessions, resulting in 454 network graphs. In addition, the *SWU 4* dataset provides phenotype information on the participants (e.g. Age, Sex) and the CC200 parcellation provides vertex attributes (e.g. vertex coordinates in the stereotaxic-coordinate system, region of the brain, amount of tissue found in each $N(v_i)$). For an example of one such network produced by the *NDMG* pipeline, see Figure 1.

2   Research Objectives

Oftentimes, it is of interest to attribute different functional tasks to particular anatomical regions of the brain. Indeed, work has shown that the left and right side of the brain undertake very different functional roles [7]. While functional MRI data attempts to capture the functional organization of the brain, DTI more readily captures structural organization. In order to support these functional differences, one may hypothesize that this left-right division may also play a role in the structural organization of the brain. Indeed, I hypothesize that the induced subgraph of vertices in the left and right side of the brain will be more cohesive than the graph-wide cohesion.

There are several measures for network cohesion available to test this hypothesis. In particular, we are interested a statistic which captures the propensity of vertices to collect more within anatomical region than between anatomical regions while also incorporating the weighted nature of these network graphs. One

---
[1]http://neurodata.io/

such statistic that satisfies these requirements is Barrat's clustering coefficient [6]. Roughly, this clustering coefficient measures the probability adjacent vertices of a vertex of interest are also connected while weighting by the edge weighes in the graph. The statistic is given by

$$C^w = \frac{1}{N_v} \sum_{i=1}^{N_v} \left\{ \frac{1}{s_i(k_i - 1)} \sum_{j,h \neq i} \left(\frac{w_{ij} + w_{ih}}{2} a_{ij} a_{jh} a_{ih} \right) \right\} \tag{2}$$

where w_{ij} is the edge weight between vertex i and j, a_{ij} is the binary variable indicating an edge between vertex i and j, k_i is the degree of vertex i, and $s_i = \sum_{j \neq i} w_{ij} a_{ij}$ is the *strength* of vertex i. As in the standard clustering coefficient, each summand counts the number of closed triangles within the neighborhood of vertex i but weights this value by the average weight found on incident edges. The normalization term $s_i(k_i - 1)$ is the maximum number of triangles vertex i can be in times the maximum edge weight which ensures that each summand falls within the unit interval. Lastly, we take the mean over full network to measure average clustering within the network.

If our hypothesis - that the left and right brain are more likely to have collections of vertices that cluster together to support functional behavior - is true, than we expect to see higher clustering coefficients for the induced subgraph of vertices in the left and right brain than across the full brain. If we only consider a single brain network, this will produce a simple three number summary. In order to more robustly explore if there is a difference in the clustering coefficient for different regions of the brain, one may construct a generative probability model in order to construct an appropriate reference distribution to compare this three number summary. Constructing a generative model that mimics the structure of the brain is a daunting task. Instead, we will collect this three number summary across all 454 networks in the *SWU 4* study and compare the distributions of the observed clustering coefficients.
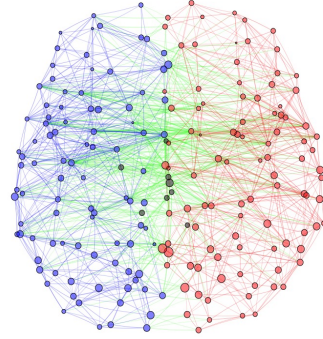

Figure 1: Subject 0025629's first session brain network graph as produced by the *NDMG* pipeline using DTI imaging. Vertices are colored according to the anatomical region (left - blue, right - red, unknow - grey). Edges are colored by anatomical region if both incident nodes fall in the same region and are colored green otherwise. Size of vertices are proportional to the amount of tissue found in each $N(v_i)$.
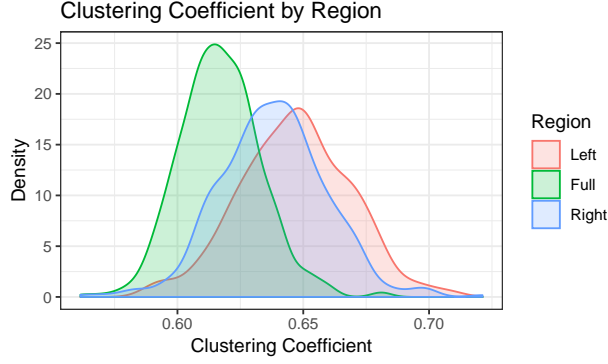
By analyzing our first hypothesis in this way, we are implicitly assuming that each statistic collected, and therefore each network graph, is independent. However, we know that each student was imaged twice which immediately violates this assumption. Moreover, it may be the case that brain scans are not be independent across relevant phenotype subgroups such as Age and Sex. This leads to our second research question; is there a difference in the structure of the brain, as measured by Barrat's clustering coefficient, between phenotype subgroups? We analyze this research question in a similar way as we did as our first hypothesis, this time comparing distributions of the clustering coefficient for the left, right, and full brain between Age, Sex, and imaging session. If there is a difference between these subgroups, we expect to see a difference in the distribution of the clustering coefficients.

Finally, we look to investigate if their is a difference between structural centrality and physical centrality of the brain. For example, it may be the case that central vertices to the DTI network graph do not lie in

Figure 2: The observed distribution of Barrat's clustering coefficient by region. It appears that vertices in the left and right side form slightly more cohensive clusters than collections of vertices in the full network.

the center of the brain. Leveraging the fact that the DTI networks are embedded into $\mathbb{R}^3$, we can compare the physical center of the brain to the "center" of each observed network graph. Let $p_1, p_2, \ldots, p_{200} \in \mathbb{R}^{200}$ be the physical locations of the vertices in the CC200 parcellation. We consider the physical center, denote P_{PC}, as the mean of these points. To measure structural centrality, we propose the following metric based on the eigen-centrality of the graph. Recall the eigencentrality defines the centrality $\mathbf{x}_i$ of vertex i by the recursive definition $\mathbf{x}_i = t^{-1} \sum_{j \sim i} \mathbf{x}_j$ where t is the number of neighbors of i. In order to incorporate the weighted nature of the graph, let $\mathbf{x}$ be the eigenvector corresponding to the largest eigenvalue of the *weighted*-adjacency matrix $\mathbf{W}$. Then $\mathbf{x}_i$ measures the weighed eigen-centrality of vertex i. Next, define the *normalized* eigen-centrality $\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_1$ so that $1 = \sum_{i=1}^{N_v} \tilde{\mathbf{x}}_i$. From here we define the structural center P_{SC} as the weighted sum of the p_i with respect to $\tilde{\mathbf{x}}_i$. In this way, the weights encode the centrality of each position and the weighed sum gives a centrality measure based instead on the eigen-centrality of each physical location. Formally, we define two measures of centrality in this report

$$P_{PC} = \sum_{i=1}^{N_v} \frac{p_i}{N_v} \qquad\qquad P_{SC} = \sum_{i=1}^{N_v} \tilde{\mathbf{x}}_i p_i \qquad\qquad (3)$$

Notice that P_{PC} is fixed with respect to the parcellation. As we have 227 individuals and consequently 454 brain images, P_{SC} will vary with respect to the DTI image. We will therefore collect $\{\hat{P}_{SC}^{(i)}\}_{i=1}^{454}$ for each brain image. Having this cloud of points in $\mathbb{R}^3$, we will then construct multivariate t-distribution confidence ellipses and compare P_{PC} to these intervals.

3   Results & Conclusion

We first investigate the results pertinent to testing the first hypothesis. Plotted in Figure 2 is the distribution of the observed clustering coefficients by each region of the brain. In green is the full brain clustering coefficients while the red and blue are the clustering coefficients from the left and right brain. It appears that the distribution of the left and right brain are shifted to the right when compared to the full brain. Indeed, the 95% empiracle interval of the full brain was $(0.588, 0.651)$ as compared to the left brain $(0.603, 0.690)$ and right brain right side $(0.597, 0.678)$ While these intervals largely overlap, they do appear to shifted slightly as evident by Figure 2 as well as the change in the center of these intervals. We also note that it appears the variances of these distributions is higher among the left and right brains than the full brain. This observation is also supported by increased width of left and right brain intervals given above. Therefore, we believe this serves as exploratory evidence to suggest that vertices may tend to cluster slightly more cohesively within

anatomical regions of the brain than across the full brain.
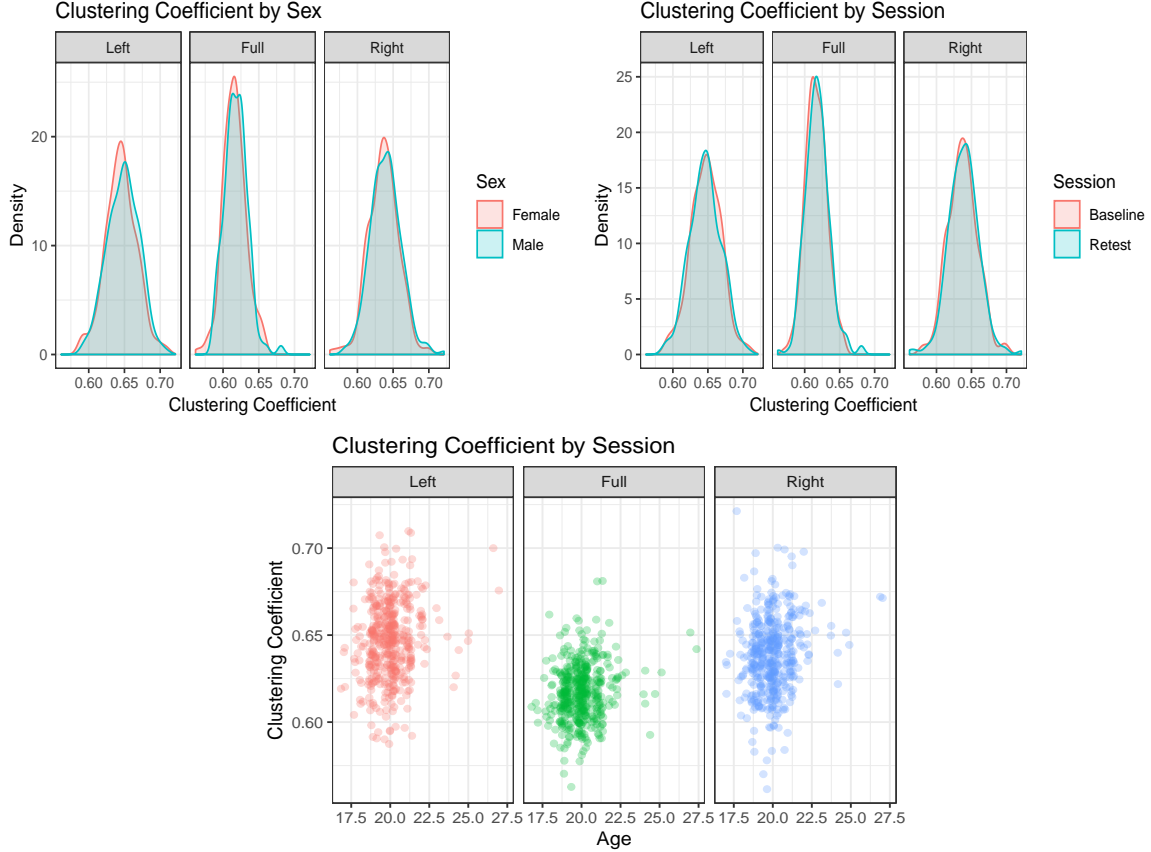

Figure 3: *Top left*: The distribution of clustering coefficient for the left brain (left), full brain (center), and right brain (right) by Sex (Male & Female). *Top right*: The distribution of clustering coefficient by session (Baseline & Retest). *Bottom*: A jittered scatter plot of clustering coefficient by age. In all three plots it appears that the clustering coefficient is independent of each independent variable.

Next we turn to analyzing our second hypothesis and hope to understand if there is a difference in the clustering of nodes within the left, right, or full brain between Sex, Age, or session number. Three plots illuminating this questions are provided in Figure 3. In the top left panel is a density plot of the clustering coefficient for the left, full, and right brain for both Male and Female participants. Immediately, it is clear that there is very little disagreement between the two distributions for any region of the brain. Indeed, it is almost difficult to distinguish the two distributions in the same plot. Similarly, in the top right panel, the clustering coefficient for each region of the brain is compared across imaging session one and two. Again, it appears that there is little to no discrepancies between these distributions. Lastly, we plot the observed clustering coefficients for each region of the brain against the age of the participant. As no clear pattern emerges in this plot, these two variables appear to be uncorrelated. These results suggest that the propensity of vertices in the network to cluster together - in any region of the brain - does not differ between imaging session, Sex, or Age.

Finally, we look to investigate if there a difference between physical centrality as measured by P_{PC} as compared to the estimated structural centers $\hat{P}_{SC}$. Consider Figure 4 for the results of this analysis. Plotted in blue are the physical locations of the vertices $\{p_i\}_{i=1}^{200}$. In order to visualize these points in $\mathbb{R}^2$, we plot

their projections onto the xy-plane (left), the xz-plane (center), and the yz-plane (right). The blue X marks the center of these points, P_{PC}. The red points are the appropriate projections of each $\{\hat{P}_{SC}\}_{i=1}^{200}$ and the red dashed ellipses are the estimated 95% multivariate t-confidence intervals visualized in each plane.
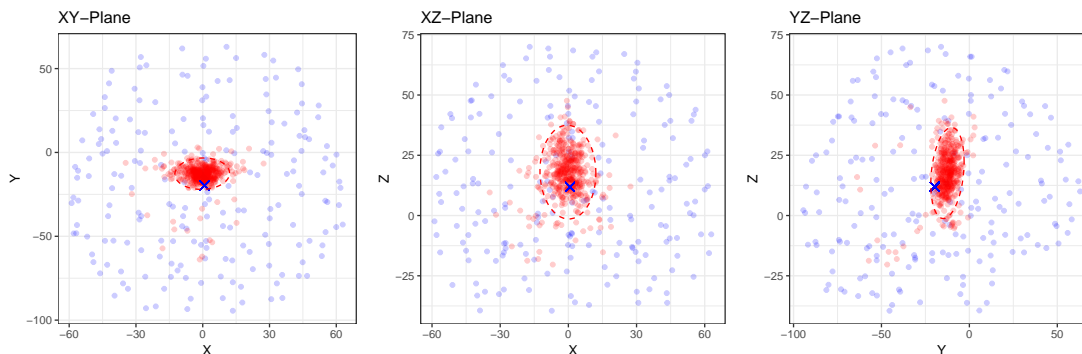

Figure 4: Plot of the vertex locations (blue points), the parcellation center (blue X), and estimated centers (red points). The dashed red lines are estimated 95% multivariate t confidence intervals. These points were projected into the XY-plane (left), XZ-plane (center), and the YZ-plane (right) for visualization.

From the figure, we first note that there is very little variability along the y axis (back to front), followed by variability on the x axis (ear to ear), while the most variability existed among the z axis (top to bottom). Therefore, there was more consistent agreement of centrality from the $\hat{P}_{SC}^{(i)}$ along the y-axis than any other. Also from the figure, each confidence ellipse contains the central P_{PC} point. This suggests that the central of the CC200 parcellation, P_{PC}, coincides with our estimate of structural centrality P_{SC}.

These figures and statistics help elicit insight into our stated research objectives. Figure 2 suggests that there may be a slight tendency of vertices to cluster more within anatomical regions than across anatomical regions of the brain. Moreover, Figure 3 suggests this phenomena is independent of the participants Sex, Age, or which session the image was taken. Finally, Figure 4 suggests that the physical center of the brain coincides with the center of the structural organization of the brain. Results like these can aid in our understanding of the structure of the brain and how that structure relates to its function.

References

[1]   Craddock et al. "A whole brain fMRI atlas generated via spatially constrained spectral clustering". In: *Human brain mapping* 33.8 (Aug. 2012), pp. 1914–1928.

[2]   Kiar et al. "A High-Throughput Pipeline Identifies Robust Connectomes But Troublesome Variability". In: *bioRxiv* (2018). DOI: 10.1101/188706. eprint: https://www.biorxiv.org/content/early/2018/04/24/188706.full.pdf.

[3]   Liu et al. "Longitudinal test-retest neuroimaging data from healthy young adults in southwest China". In: *Scientific Data* 4 (Feb. 2017).

[4]   Soares et al. "A hitchhiker's guide to diffusion tensor imaging". In: *Frontiers in neuroscience* 7 (Mar. 2013), pp. 31–31.

[5]   Zuo et al. "An open science resource for establishing reliability and reproducibility in functional connectomics". In: *Scientific Data* 1 (2014).

[6]   A. Barrat et al. "The architecture of complex weighted networks". In: *Proceedings of the National Academy of Sciences* 101.11 (2004), pp. 3747–3752. ISSN: 0027-8424. DOI: 10.1073/pnas.0400087101. eprint: https://www.pnas.org/content/101/11/3747.full.pdf. URL: https://www.pnas.org/content/101/11/3747.

[7]   M. C. Corballis. "Left brain, right brain: facts and fantasies". In: *PLoS Biol.* 12.1 (Jan. 2014), e1001767.