Exploring the Impact of Autism Spectrum Disorder on the Structural Connectome

John Popeck
Computer Science
College of Computing and Informatics at Drexel University
Philadelphia, United States
jp3798@drexel.edu

***Abstract*—Recent studies have begun to examine whether patients with neurological disorders show distinctive network properties in their connectomes. If so, the findings have the potential to lead to increased understanding of the underlying nature of the disorders, as well as more accurate diagnoses. In this paper I explore the structural connectomes of individuals with autism spectrum disorder (ASD), and compare them to a control set of individuals without the disorder. I attempt to determine whether there is a group difference for any network measures between the ASD set and the control set, and whether a classifier can be trained to identify whether a connectome is from an individual with ASD. I also investigate whether any network measure shows a correlation with the level of severity of ASD.**

I. Introduction

Network neuroscience involves the study of connectomes, which are matrix representations of brain networks encompassing the pairwise connections between the various regions of the brain. Overall the field attempts to understand how connectomes relate to the brain's capacity and function. In recent years a specific area of focus has been how diseases and disorders impact the connectome. Since the connectome represents the network of communication patterns between different brain elements, it provides a powerful tool to study the nature of neurological disorders which impact the way the brain operates [2].

Autism spectrum disorder (ASD) is one such neurological disorder which has received this attention. ASD is a common disorder that affects close to one percent of the population, making it an obvious and important candidate for study [3]. Studies analyzing graph theory measures of the connectomes from subjects with ASD have shown that ASD connectomes exhibit some distinctive network properties. A 2015 study of structural connectomes found that subjects with ASD had decreased global efficiency, increased characteristic path length, and decreased strength of connections [3]. A 2021 study of functional connectomes found a higher mean modular variability in ASD subjects, suggesting more overall instability in global brain dynamics [4].

In addition to network analysis, researchers have also used connectomes to investigate finding classifiers and biomarkers for ASD. A biomarker for ASD would be an objective empirical measure which can be shown to clearly predict whether an individual has the condition or not. A classifier is a machine learning model that takes input data, and uses features from the data (such as biomarkers) to predict which group the data belongs to. Research on ASD classifiers does not always involve connectomes. For example, Tunc et al. [1] used evaluation scores like MESL (Mullen Scales of Early Learning) and VABS (Vineland Adaptive Behavior Scales) as input data to a machine learning technique called a Support Vector Machine (SVM) to establish an ASD classifier. The classifier achieved high performance, with an accuracy of 84% [1]. Several studies searching for classifiers using functional connectomes have also been performed, with good results. One in particular showed that a functional connectome collected from a subject at the age of six months could be used to predict autism status at the age of 24 months [5]. These studies have used a variety of machine learning methods including SVMs, linear regression, random forests, and neural networks [5].

In this paper I will attempt to add to the previous work done on both network analysis and on finding a classifier for ASD. I will use a dataset with structural connectomes from both ASD subjects and healthy controls, compute graph theory measures on the connectomes, and attempt to determine if there is a significant statistical difference between any measure for the ASD set and the control set. I will then input the data into a Support Vector Machine and use it to generate a classification boundary which can be used to predict whether an individual has ASD or not. Finally, I will also analyze whether any of the computed measures correlate to the severity of ASD diagnosed in each subject.

II. Materials and Methods

A. Materials

My analysis will be performed on the CHARM dataset, provided by Penn Medicine. The CHARM dataset contains structural connectomes from 450 subjects. The data was collected from eleven different sites, and the subjects are a mix of patients with ASD and patients without ASD (healthy controls).

Even though 450 subjects were included in the study, only 313 subjects produced connectomes which passed a quality assessment review. I will only include the connectomes from these 313 in my analysis. Of these 313 subjects, 163 are diagnosed with ASD, while 150 are not diagnosed with ASD.

Each subject has only one brain image, but connectomes were generated using two different atlases, with the end result

that there are two connectomes for each subject. The two atlases used are the Desikan atlas and the Schaefer atlas. The Desikan atlas has 86 regions, and thus connectomes created using this atlas have 86 nodes in their network. The adjacency matrix from these connectomes is size 86 x 86 and has non-zero values in every position where there is a connection between regions. The edges are weighted, with greater values indicating a stronger connection between the two regions. The Schaefer atlas has 220 regions, and connectomes created using this atlas have 220 nodes. Edges are also weighted in these connectomes, and so adjacency matrices are similar as to what was described for the Desikan atlas connectomes, except they have size 220 x 220.

Each subject also has their sex, age, diagnosis (ASD or non-ASD), Autism Diagnostic Observation Schedule Calibrated Severity Score (ADOS_CSS), Social Communication Questionnaire (SCQ) score, and IQ recorded. I will use the diagnosis to distinguish between subjects with ASD and subjects without ASD. I will use the ADOS_CSS and SCQ score to interpret ASD severity. The ADOS_CSS score ranges from 1 to 10 with a higher score representing a more severe level of ASD [1]. The SCQ score can range from 0 to 39 with a higher score representing more social impairment.

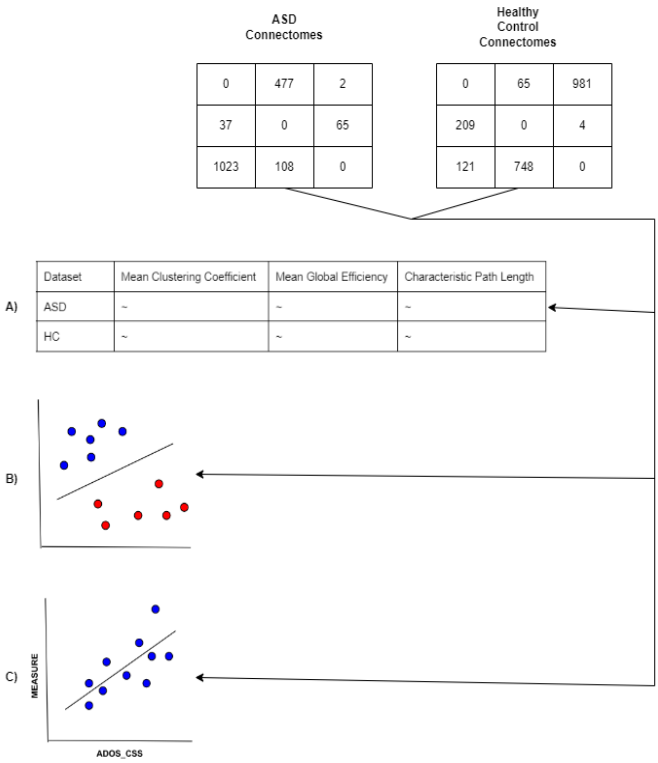

Fig. 1. Structural connectomes from both ASD subjects and Healthy Control subjects will be used A) to calculate and analyze graph theory measures B) as input into a Support Vector Machine, to determine a classifier for ASD C) to check for a correlation between graph theory measures and autism severity scores like ADOS_CSS

B. Methods

To calculate graph theory measures, I used code from the Brain Connectivity Toolbox (BCT) [6]. I computed several key node level and network level measures we discussed in class that are supported by the BCT. Next, I attempted to determine whether there was a statistically significant group difference between the ASD set and healthy control set for several network measures. For this analysis, I chose to focus on measures which were identified in previous research as distinctive in connectomes from ASD subjects [3], as well as measures that highlight certain characteristics about the overall brain networks. This includes measures for degree, density, strength, clustering, centrality, path length, and assortativity.

To attempt to generate a classifier for ASD, I created an SVM with a linear kernel using the python sci-kit learn library [10]. The input features to the SVM were selected from the calculated graph theory measures. Selected measures included the same network level measures used in my group difference analysis, as well as node level measures that corresponded to the network level measures which showed significant difference between the two populations. This resulted in 359 features for my SVM based on the Desikan atlas connectomes, and 895 features for my SVM based on the Schaefer atlas connectomes. Since four node level measures were selected, and the Schaefer atlas has 134 more nodes than the Desikan atlas, this accounts for the difference in number of features between the two. Additionally, I filtered these features through a feature selection process called Sequential Forward Floating Selection (SFFS) [9], in order to pre-select only the features which are most relevant to classification. This process can help reduce over-fitting in the resulting classifier, as it significantly reduces the dimensionality of the input data. SFFS uses an iterative algorithm to evaluate which individual features or combinations of features will contribute most to the success of the classifier, stopping when a pre-determined number of desired output features (k) are identified. I experimented with three different k values (10, 20, 30) for an SVM based on each atlas. These selected features are then used to train the SVM. For training, I used 70% of the available dataset (219 connectomes). The remaining 30% (94 connectomes) were reserved for a testing set. I evaluated the accuracy of each trained classifier SVM on the testing set.

To test for a relationship between specific graph theory measures and autism severity, I calculated Pearson's correlation between computed graph theory measures and the ADOS_CSS and SCQ scores of the subjects. I used the same 18 network level graph theory measures for this comparison as I used for my analysis of group difference.

III. Results

A. Group Difference

Results of group difference analysis on selected network measures from connectomes constructed using the Desikan atlas are shown in Table I , and from connectomes constructed using the Schaefer atlas in Table II. As the set of data in

Desikan Atlas Group Difference		
Measure	Effect Size	P Value
degree average	-0.29	0.04696
density	-0.29	0.04696
clustering coefficient	0.21	0.109766
characteristic path length	0.03	0.824202
global efficiency	0.13	0.354301
betweenness centrality	0.1	0.474474
eigenvector centrality	0.22	0.100847
global strength	0.05	0.709829
assortativity	-0.16	0.234072
degree between module	-0.27	0.04696
degree inter hemisphere	-0.31	0.04696
degree intra hemisphere	-0.08	0.596056
degree within module	-0.49	0.000339
global modularity	0.21	0.109766
global modularity negative	0.24	0.072305
strength inter hemisphere	-0.27	0.04696
strength intra hemisphere	-0.27	0.04696
strength self connections	0.04	0.737407

TABLE I
GROUP DIFFERENCE CALCULATION ON NETWORK MEASURES USING THE DESIKAN ATLAS

Schaefer Atlas Group Difference		
Measure	Effect Size	P Value
degree average	-0.17	0.309874
density	-0.17	0.309874
clustering coefficient	0.19	0.309874
characteristic path length	-0.14	0.34979
global efficiency	0.1	0.49463
betweenness centrality	0.32	0.083648
eigenvector centrality	-0.08	0.599656
global strength	0.05	0.694127
assortativity	0.1	0.49463
degree between module	-0.16	0.309874
degree inter hemisphere	-0.21	0.309874
degree intra hemisphere	0.05	0.694127
degree within module	-0.21	0.309874
global modularity	0.15	0.325631
global modularity negative	0.11	0.49463
strength inter hemisphere	-0.17	0.309874
strength intra hemisphere	0.17	0.309874
strength self connections	0.04	0.694289

TABLE II
GROUP DIFFERENCE CALCULATION ON NETWORK MEASURES USING THE SCHAEFER ATLAS

each group comes from different subjects, and the data is parametric, the independent group t-test was used to calculate a P value for each measure. The P value represents whether there was a statistically significant difference between the groups for the measure. Calculated P values were corrected using the Benjamini-Hochberg procedure, in order to appropriately adjust for chance. Cohen's D was used to calculate the effect size for each measure, which represents the magnitude of the difference between the two groups.

The P value should be less than 0.05 in order for a result to be considered significant. Most of my results do not meet this requirement. However, the group difference calculations involving average degree, degree within and between modules, degree within and between hemispheres, and strength within and between hemispheres have a P value less than 0.05 for connectomes constructed using the Desikan atlas. The most significant difference appears in the degree within module measure. It has a relatively strong effect size of -0.49, indicating that Desikan connectomes from the ASD group typically have less connections within each brain module than the healthy control group. The other significant differences in the Desikan connectomes indicate that nodes in the ASD group have less degree on average, less connections across brain modules, less connections within their own hemisphere, less strongly weighted connections within their own hemisphere, and more strongly weighted connections which span hemispheres.

Interestingly, the calculations on connectomes constructed using the Schaefer atlas do not corroborate these findings. No significant group differences (P value less than 0.05) were found at all in the measures computed on the Schaefer connectomes.

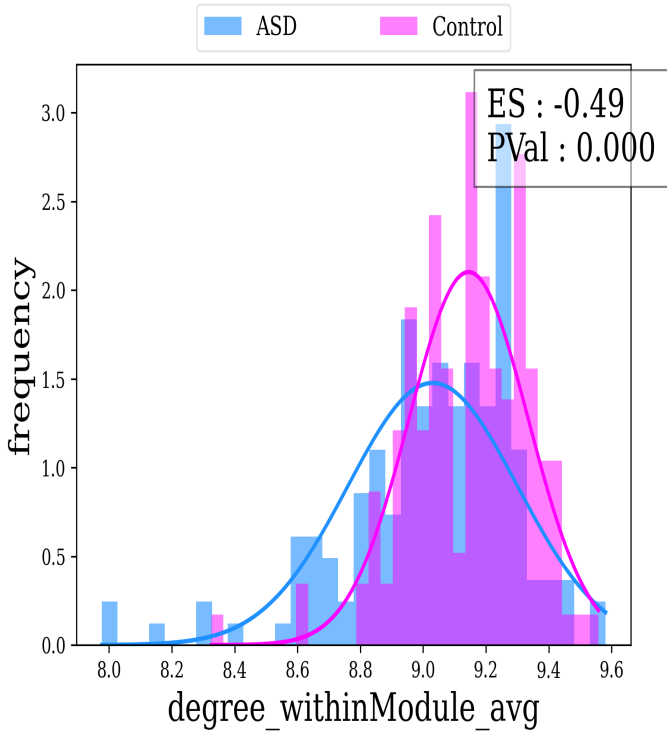

Fig. 2. A histogram showing the distribution of the degree within module measure across the Desikan connectomes for both the ASD and healthy control groups. The groups are statistically different, with the ASD group exhibiting nodes with substantially lower degrees within their own brain modules.

B. *Classifier*

The accuracy results for each evaluated SVM are found in Table III and Table IV. The SVMs trained on features from the Desikan connectomes, using 20 and 30 selected features respectively, produced modest results. Their accuracies on evaluating the testing set were slightly above 60%, showing that they were 10% to 15% better than random chance at identifying a connectome as having an ASD diagnosis or

not. The other evaluated SVM classifiers were not effective, with accuracy scores hovering around 50%. This is not too surprising for the SVMs trained on features from the Schaefer connectomes, as earlier statistical analysis did not show any group difference between the ASD and healthy control sets for these features. The SVM trained on 10 features from the Desikan connectomes perhaps just did not have enough input information.

Desikan Atlas SVM Results		
k	Selected Features	Accuracy
10	1 D, 8 DWM, 1 SIH	52.13%
20	2 D, 16 DWM, 2 SIH	63.83%
30	2 D, 22 DWM, 3 DIH, 3 SIH	61.70%

TABLE III
SVM CLASSIFIER ACCURACY RESULTS USING THE DESIKAN ATLAS. THE TABLE SHOWS THE NUMBER OF SELECTED FEATURES (K), AND WHICH KINDS OF FEATURES WERE SELECTED. ALL SELECTED FEATURES WERE AT THE NODE LEVEL. D = DEGREE, DWM = DEGREE WITHIN MODULE, DIH = DEGREE INTER HEMISPHERE, SIH = STRENGTH INTER HEMISPHERE

Schaefer Atlas SVM Results		
k	Selected Features	Accuracy
10	1 D, 5 DWM, 4 DIH	52.13%
20	7 D, 6 DWM, 5 DIH, 2 SIH	50.00%
30	6 D, 13 DWM, 6 DIH, 5 SIH	47.87%

TABLE IV
SVM CLASSIFIER ACCURACY RESULTS USING THE SCHAEFER ATLAS. THE TABLE SHOWS THE NUMBER OF SELECTED FEATURES (K), AND WHICH KINDS OF FEATURES WERE SELECTED. ALL SELECTED FEATURES WERE AT THE NODE LEVEL. D = DEGREE, DWM = DEGREE WITHIN MODULE, DIH = DEGREE INTER HEMISPHERE, SIH = STRENGTH INTER HEMISPHERE

Interestingly, the features selection process (SFFS) for every SVM selected all node level measures, ignoring the network level measures. Table III and Table IV also display which kinds of node level measures were selected. The Desikan SVMs in particular seemed to focus on degree within module scores for particular nodes, which makes sense considering that the network level degree within module scores showed the most significant difference between the populations for these connectomes. The modules of the particular nodes selected here may show which brain modules are structured most differently in the connectomes of these ASD subjects. I plan to investigate this point further in my final report.

C. Correlation with Severity

Pearson's correlation produces a P value, which represents whether a result is statistically significant, and an R value which represents the magnitude of the correlation. A P value should be less than 0.05 to be considered significant. R values can range from -1 to 1, with 1 representing a perfect linear relationship, -1 representing a perfect inverse relationship, and 0 representing no correlation at all. A few sample P and R values from my experiments are displayed in Table V and Table VI. Like with my group difference results, the P values were corrected to account for chance using the Benjamini-Hochberg procedure.

The results show no significant correlation at all between any computed graph theory measure and either the ADOS_CSS score or SCQ score. This holds for both the Desikan and Schaefer connectomes. For this dataset at least, it appears we cannot conclude any relationship between our analyzed graph theory measures and autism severity.

Desikan Atlas Correlation with Severity Scores			
Measure	Severity Score	R Value	P Value
degree average	ADOS_CSS	-0.07	0.578745
degree within module	ADOS_CSS	-0.11	0.578745
characteristic path length	ADOS_CSS	0.06	0.578745
degree average	SCQ	0.02	0.961343
degree within module	SCQ	0.01	0.961343
characteristic path length	SCQ	-0.09	0.961343

TABLE V
PEARSON'S CORRELATION BETWEEN NETWORK MEASURES AND AUTISM SEVERITY SCORES USING THE DESIKAN ATLAS

Schaefer Atlas Correlation with Severity Scores			
Measure	Severity Score	R Value	P Value
degree average	ADOS_CSS	-0.06	0.660012
degree within module	ADOS_CSS	-0.05	0.660012
characteristic path length	ADOS_CSS	-0.08	0.660012
degree average	SCQ	0.03	0.990824
degree within module	SCQ	0.01	0.990824
characteristic path length	SCQ	0	0.990824

TABLE VI
PEARSON'S CORRELATION BETWEEN NETWORK MEASURES AND AUTISM SEVERITY SCORES USING THE SCHAEFER ATLAS

REFERENCES

[1] Tunç, B., Pandey, J., St. John, T., Meera, S.S., Maldarelli, J.E., Zwaigenbaum, L., Hazlett, H.C., Dager, S.R., Botteron, K.N., Girault, J.B. and McKinstry, R.C., 2021. Diagnostic shifts in autism spectrum disorder can be linked to the fuzzy nature of the diagnostic boundary: a data-driven approach. Journal of Child Psychology and Psychiatry, 62(10), pp.1236-1245.

[2] van den Heuvel, M.P. and Sporns, O., 2019. A cross-disorder connectome landscape of brain dysconnectivity. Nature reviews neuroscience, 20(7), pp.435-446.

[3] Roine, U., Roine, T., Salmi, J., Nieminen-von Wendt, T., Tani, P., Leppämäki, S., Rintahaka, P., Caeyenberghs, K., Leemans, A. and Sams, M., 2015. Abnormal wiring of the connectome in adults with high-functioning autism spectrum disorder. Molecular Autism, 6(1), pp.1-11.

[4] Xie, Y., Xu, Z., Xia, M., Liu, J., Shou, X., Cui, Z., Liao, X. and He, Y., 2022. Alterations in connectome dynamics in autism spectrum disorder: A harmonized mega-and meta-analysis study using the Autism Brain Imaging Data Exchange Dataset. Biological Psychiatry, 91(11), pp.945-955.

[5] Horien, C., Floris, D.L., Greene, A.S., Noble, S., Rolison, M., Tejavibulya, L., O'Connor, D., McPartland, J.C., Scheinost, D., Chawarska, K. and Lake, E.M., 2022. Functional connectome-based predictive modelling in autism. Biological Psychiatry.

[6] Rubinov, M. and Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), pp.1059-1069.

[7] Kazeminejad, A. and Sotero, R.C., 2019. Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification. Frontiers in neuroscience, 12, p.1018.

[8] Liu, Y., Xu, L., Li, J., Yu, J. and Yu, X., 2020. Attentional connectivity-based prediction of autism using heterogeneous rs-fMRI data from CC200 atlas. Experimental Neurobiology, 29(1), p.27.

[9] Pudil, P., Novovičová, J. and Kittler, J., 1994. Floating search methods in feature selection. Pattern recognition letters, 15(11), pp.1119-1125.

[10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, pp.2825-2830.