5-1-2024

Detection of Brain Communities in the General Children Population

Britny Farahdel

Recommended Citation

Farahdel, Britny, "Detection of Brain Communities in the General Children Population." Thesis, Georgia State University, 2024.

doi: https://doi.org/10.57709/36974806

Detection of Brain Communities in the General Children Population

by

Britny Farahdel

Under the Direction of Jingyu Liu, Ph.D.

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2024

ABSTRACT

The fingerprint is known to be unique in every individual, and there is evidence that such individuality exists with the brain. Neuroimaging studies that research brain fingerprint patterns typically consider relationships between individuals and their brain patterns. However, there remains a question as to how such fingerprint patterns can be grouped among the general population. In this study, we implemented clustering-based methods to evaluate whether such subgrouping exists among individuals and evaluated the relationships between these clusters and individuals' developmental, cognitive, demographical, psychological status in the Adolescent Brain and Cognitive Development study cohort. Multiplex community detection and K-means clustering revealed the existence of clusters in our cohort, as well as significant group differences between these clusters in these datasets, indicative of heterogeneous subgrouping of brain fingerprint patterns in the general population.

INDEX WORDS: Community Detection, Brain Fingerprinting, Brain Imaging

Detection of Brain Communities in the General Children Population

by

Britny Farahdel

Committee Chair: Jingyu Liu

Committee: Vince Calhoun

Armin Iraji

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2024

DEDICATION

I want to thank my family and friends for all of their love and support. I appreciate their acts of kindness and encouragement they gave me throughout my academic journey.

ACKNOWLEDGMENTS

I want to extend my gratitude to my advisor Dr. Jingyu Liu for all of her support throughout my academic journey. I'm grateful that I had the opportunity to work with her, as she has always encouraged me to strive for excellence. Her mentorship and guidance has been invaluable to my personal and professional growth. I would also like to thank my committee members Dr. Vince Calhoun and Dr. Armin Iraji for their expertise and contributions to my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Evidence of "fingerprint" patterns of the brain have been found in previous studies with brain structure, function, and signals. Brain anatomical features, "brain prints", can be differentiated between individuals for cortical thickness, area, and volume Valizadeh et al. (2018). Functional brain connectivity profiles have been used for identification of individuals and prediction of cognitive behavior Finn et al. (2015); Cai et al. (2021, 2020). Significant group differences have been found between localized regions ("fingerprints") of brain white matter connectivity across monozygotic twins, dizygotic twins, non-twin siblings, and other genetically unrelated subjects Yeh et al. (2016). These studies all indicate that individuals can be characterized based on brain "fingerprint" patterns. However, are there individuals that share some commonalities? In other words, is there evidence of subgrouping of individuals? Under the hypothesis that a natural subgrouping might exist in the general population of children in terms of brain structure and function, which could partially underlie the heterogeneity in the brain responses to stimuli observed in fMRI data Fahle & Spang (2003), we performed a pilot attempt to answer this question in this study.

The aim of this study is to investigate structural and functional neuroimaging features with clustering methods to identify potential brain communities embedded in the general population of children. We applied community detection which identifies dense groups of nodes referred to as communities in graph networks Traag et al. (2019). Multiplex community detection Magnani et al. (2021) identifies communities using multiple layers of graph

networks on the same set of nodes. Another clustering method, K-means clustering, was also applied for verification of communities, which involves the selection of k points as clustering centers (centroids) iteratively Zhu et al. (2021).

CHAPTER 2
METHODS

2.0.1 Cohort

The Adolescent Brain and Cognitive Development (ABCD) study, an ongoing longitudinal study following a cohort of over 10 thousand children recruited at ages 9-10 Barch et al. (2018) from 21 sites across the USA, has collected rich datasets ranging from genetics, environmental, physical and mental health, to cognition and brain MRI imaging. A total of 1,685 brain MRI-imaging derived features, including measures for white matter, structure, stop signal task functional MRI (fMRI) activation, Resting-State fMRI network correlations, Emotional N-Back task fMRI activation, white matter fractional anisotropy and mean diffusivity, were provided by the ABCD study for a total of 7,371 subjects at baseline. Previous studies have shown evidence of scanner effects for some of these features Hagler Jr et al. (2019). To disentangle scanner effects in our analysis, we separated data based on the manufacturer of the scanner. For the Siemens scanner, data from a total of 4,962 subjects are further split into the discovery (3,473 subjects) and replication (1,489 subjects) sets. Data from both GE and Philips scanners with 2,409 subjects form the 2^{nd} replication set with different scanners. Finally, the 2^{nd} replication set combined with the 1^{st} Siemens replication set forms the 3^{rd} replication set for a large sample of 3,898 subjects. Each scanner set was harmonized with ComBat Fortin et al. (2018) to remove the variance contributed by collection site differences.

Cognitive assessments from NIH Toolbox Cognitive battery are used for analyses of group

differences between different communities, including scores from the Picture Vocabulary Task, Oral Reading Recognition Task, Flanker Task, Pattern Comparison Processing Speed Test, and Picture Sequence Memory Test Thompson et al. (2019). Diagnostic assessment from Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime version (K-SADS-PL) are used for association analyses with depression, bipolar, psychosis, anxiety, obsessive-compulsive disorder, eating disorder, attention deficit hyperactivity disorder, oppositional defiant disorder, and autism Nishiyama et al. (2020). In addition, the pubertal developmental scale Barch et al. (2018), the general psychopathology factor (p-factor) score estimating individuals' tendency to develop psychiatric disorders Farahdel et al. (2021), early life stress score Thapaliya et al. (2021) summarizing physical abuse/trauma/sexual abuse, household substance abuse, household mental illness, criminal in household, parent separation/divorce, emotional neglect, are also tested for association with clusters in this study.

2.0.2 Community Detection and K-means Clustering

A pairwise Pearson correlation matrix was first constructed by computing the Pearson correlations for the 1,685 brain-imaging features between subjects. This correlation matrix was used to construct two adjacency matrices for multiplex community detection: one for the positive correlation layer and one for the negative correlation layer, each layer with equal weights Traag & Bruggeman (2009). To construct the adjacency matrix, the correlations were assigned a value of 1 if the absolute values were greater than 0.1 (approximately one standard deviation from the mean) or assigned 0, otherwise. Communities are determined based on the edges connecting the nodes by maximizing modularity computed as the differ-

ence between the actual and expected number of edges in a community Traag et al. (2019). The modularity ranges from 0 to 1 with a typical range of 0.3 to 0.7 Newman & Girvan (2004).

Community detection is a method for identifying dense groups of nodes referred to as communities in graph networks Traag et al. (2019). Multiplex community detection involves a compact network model consisting of multiple layers of graphs which are formed from graphs identified on the same set of nodes of interactions between the same type of entities Magnani et al. (2021). Communities can be determined based on the weights of the links, or edges, that connect the nodes, which can be positive or negative Traag & Bruggeman (2009). Communities can be detected by optimizing modularity. Modularity is obtained based on the maximizing the difference between the actual and expected number of edges in a community Traag et al. (2019), and ranges from 0 to 1 (strong community structure) with a typical range of 0.3 to 0.7 Newman & Girvan (2004).

Multiplex community detection with the Leiden algorithm using the leidenalg python package Traag et al. (2019) and NetworkX Hagberg et al. (2008) for plotting was performed on the discovery and replication sets with the positive and negative layers. To test the robustness of the results from the discovery set, we implemented community detection for 1000 runs with different initiations and obtained the average intersection between the obtained communities. Furthermore, we also obtained the average intersection percentage between clusters from community detection and clusters from K-means clustering over 1000 runs. K-means clustering selected the same number of clusters as community detection using Pearson

correlation as a distance metric.

Additionally, for replication purposes, K-means clustering using the means from the communities of the discovery set as the centroids Pedregosa et al. (2011), (termed replication K-means clustering) was performed on all 3 replication sets. The multiplex modularity of these clusters was computed using the labels from the K-means results. We further compared these multiplex modularity values with those derived with K-means clustering using random centroids for 1000 runs.

2.0.3 Interpretation of Communities

To understand the impact of communities, we tested differences between communities in terms of pubertal developmental status, sex, early life stress, general p factor score, as well as K-SADS diagnostic status and cognitive assessments. Specifically, ANOVA and Tukey's multiple comparisons ($\alpha = 0.05$) were performed for variables of pubertal developmental status, early life stress, general p factor scores, and cognitive assessments. A chi-square test of independence was applied to sex and K-SADS diagnostic data. We implemented these analyses on the discovery set and all replication sets (replication K-means clustering results used).

The general psychopathology dimension (also called the p factor) is derived from a broad range of symptomatic measures, along with two or three other specific dimensions as latent factors in a bifactor model verified by confirmatory factor analysis Caspi et al. (2014). Confirmatory factor analysis was used to verify a proposed or hypothesized model presented as factor structures. As suggested in previous studies Caspi et al. (2014); Bornovalova et al.

(2020), we hypothesized a bifactor model which included three latent factors: general p factor model, internalizing and externalizing factors within broad spectral behavioral measures Farahdel et al. (2021). In our model, the three latent factors (general p, internalizing, and externalizing factors) were uncorrelated and had unique variances, and included the ABCD Child Behavior Checklist (CBCL) subscale scores as the inputs. The general p factor scores were extracted for our subjects using this model.

We explored brain differences between the communities using the original brain-imaging features and the principal components of imaging features for both the discovery and replication sets. In the discovery set, pairwise t-tests were first performed between each community for each of the 1,685 features. PCA Pedregosa et al. (2011) was applied to the features to reduce the dimensionality to 100 principal components (PCs) ($\geq$ 65% variance explained). Pairwise t-tests were then implemented between communities for each PC. To decide the contributing features for each PC, PCA weights were standardized using the Z-score method and the contributing features for each PC were determined as features that had absolute z-scores above two standard deviations. PCA weights of the discovery set were used to transform the replication datasets to PCs. ANOVA and Tukey's multiple comparisons were implemented onto the PCs of replication sets.

T-Distributed Stochastic Neighbor Embedding (t-SNE) Pedregosa et al. (2011) was applied to the discovery and replication sets to reduce the dimensionality of the top five PCs to obtain a three-dimensional visual representation.

CHAPTER 3
RESULTS

3.0.1 Community Detection and K-means Clustering

Community detection on the discovery set resulted in 4 communities with the multiplex modularity score of 0.495. Each community is labeled with a specific color code [Fig. 3.1], and the percentages of subjects in each community were 29.00% for blue, 25.34% for green, 24.42% for red, and 21.25% for yellow. The intersection percentage of each community from 1000 random initiation runs were approximately 80% for the red, green, and blue communities, while both three or four communities were detected. About half of the runs resulted in detection of the yellow community with an average percent intersection of approximately 60%. The intersection percentage between the communities from community detection and communities from K-means clustering with four initial cluster centers out of 1000 runs was approximately 70% for each community. Community detection on the Siemens replication set resulted in three communities and the percentages of subjects in each community were 38.15% blue, 35.66% green, and 26.19% red, with a multiplex modularity of 0.526.

Replication K-means clustering resulted in the following percentages for each community in the Siemens replication set: 29.08% blue, 26.86% green, 26.66% red, and 17.39% yellow [Fig. 3.1] with a multiplex modularity of 0.503. The multiplex modularity values of K-means clustering with random centroids was less than 0.503 for all 1000 runs. For the 2^{nd} replication set, the percentages of each community were 28.97% blue, 22.75% green, 25.45% red, and 22.83% yellow with a multiplex modularity of 0.427. For the 3^{rd} replication set the

percentages of each community were 29.09% blue, 24.24% green, 25.91% red, and 20.75% yellow with a multiplex modularity of 0.467.

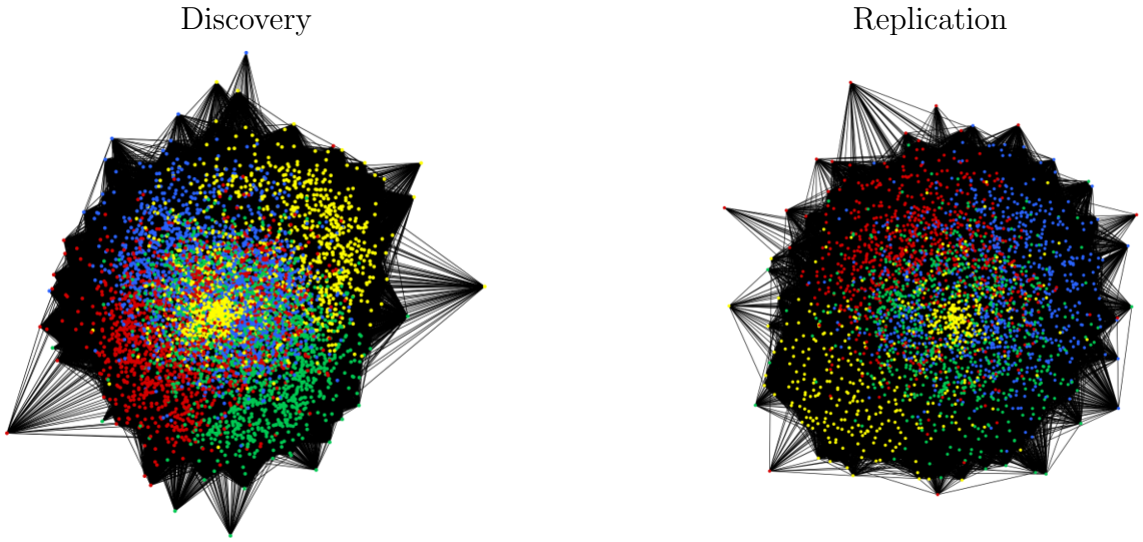

Figure 3.1 Communities in the Discovery and K-means Replication set

3.0.2 Interpretation of Communities

The results of the ANOVA for the general p factor scores were not significant for the discovery and the 2^{nd} replication sets. Figure 3.2 shows results from the 1^{st} and 3^{rd} replication sets, where the red community was found to be significantly greater than the blue community.

Across communities, there were significant sex differences in all discovery and replication sets [Tab. 3.1]. For the pairwise comparisons, except the blue and yellow communities in the 1^{st} replication set, all other pairs showed different sex distributions. The green community had significantly more males, the red community had significantly more females, while the blue community had almost equal females and males.

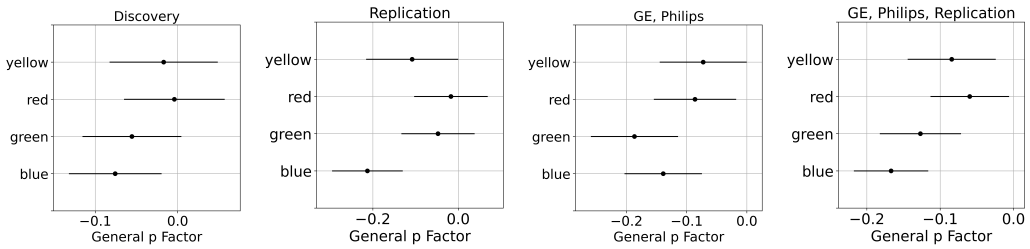

Figure 3.2 Multiple Comparisons Interval Plots (95% CI) of the General p Factor

Table 3.1 Percentage of males in each community

	Disc.	Rep.	2^{nd} rep. (GE, Philips)	3^{rd} rep. (GE, Philips, Rep.)
Blue	49.45	^{ns}50.34	49.71	49.91
Green	65.00	66.49	64.04	65.10
Red	31.74	31.47	27.90	29.41
Yellow	54.87	^{ns}57.56	57.64	57.59

ns means No Significance between indicated clusters (all other differences are significant).
Disc. is Discovery and Rep. is replication.

For the male pubertal developmental status, the ANOVA tests were significant for discovery set and the 2^{nd} and 3^{rd} replication sets. Multiple comparisons showed significant differences between the blue and green, and red and green communities of the Discovery set [Fig. 3.3]. For the female pubertal developmental status, ANOVA tests were only significant in the discovery set, where the blue community showed significantly higher values than the green community [Fig. 3.3].

In the replication, 2^{nd} and 3^{rd} replication sets, the red community had consistently lower scores than all other communities for the Flanker Task, and the blue community was found to

have higher scores than all other communities with the Pattern Comparison Speed Processing Test [Fig. 3.4].

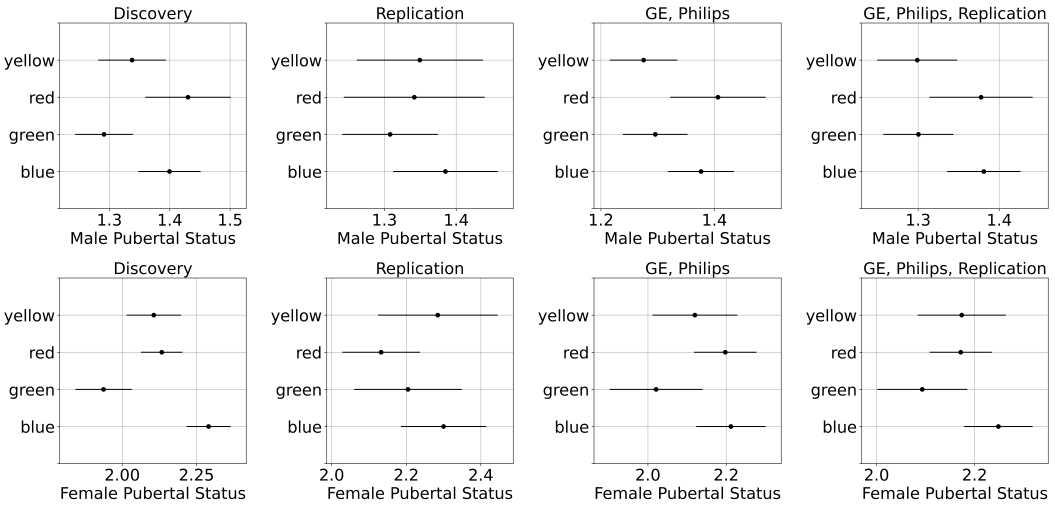

Figure 3.3 Multiple Comparisons Interval Plots (95% CI) of Male and Female Pubertal Developmental Status

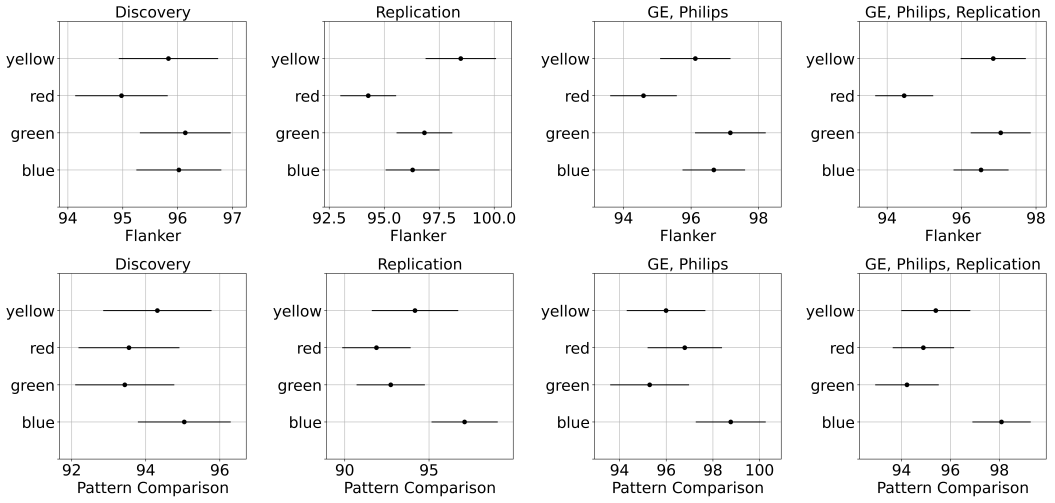

Figure 3.4 Multiple Comparisons Interval Plots (95% CI) of the Flanker Task and Pattern Comparison Processing Speed Test

For Picture Vocabulary, Oral Reading, and Picture Sequence Memory tests, there were significant differences across all communities except for the Picture Sequence Memory assessment scores in the 2^{nd} and 3^{rd} replication sets. As shown in Figure 3.5, the red community has consistently lower scores than the other communities in all three assessments and all datasets except the Picture Sequence Memory assessment scores in the 2^{nd} replication set.

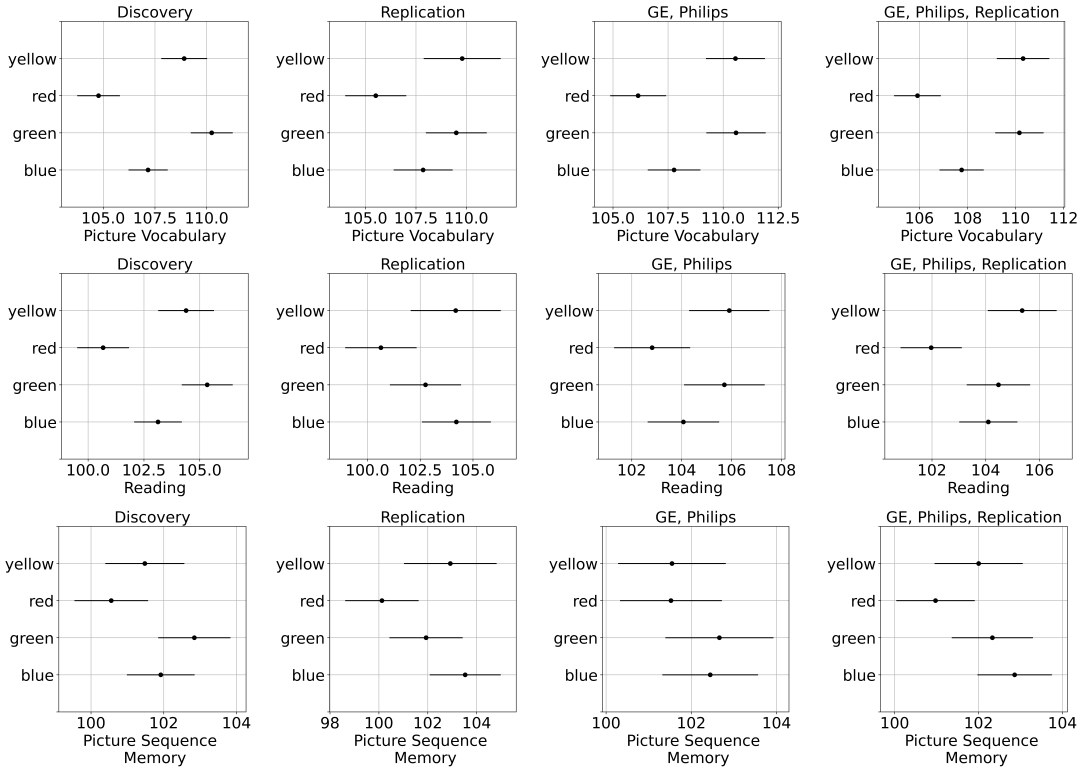

Figure 3.5 Multiple Comparisons Interval Plots (95% CI) of the Picture Vocabulary, Oral Reading, and Picture Sequence Memory Tests

The ANOVA tests showed significant differences between communities for age and early life stress scores for all datasets. Multiple comparisons were performed and plotted in Figure 3.6. The chi-square test for the K-SADS-PL assessments reported no significant differences.

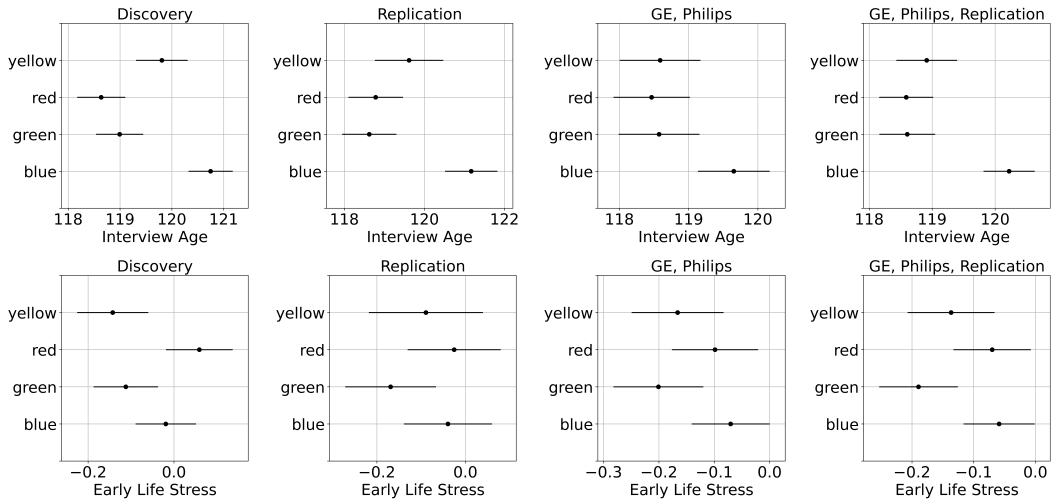

Figure 3.6 Multiple Comparisons Interval Plots (95% CI) of Age, Early Life Stress

For visual presentation purposes, t-SNE was applied to the top five PCs and visualized in three dimensions of embedded space shown in Figure 3.7. There is some overlap between clusters.

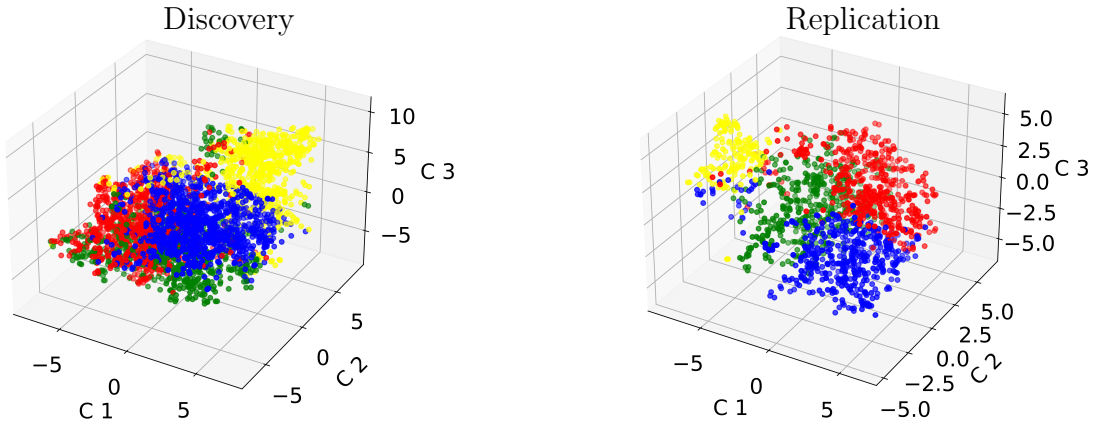

Figure 3.7 Three dimensions of the t-SNE Embedded Space

For the brain features associated with each community, pairwise t-tests on the original feature space showed that greater than 30% of features had significant differences between communities in the discovery set for each pair ($\alpha = 0.05$, Bonferroni correction = $0.05/1685 =$ 2.97E-5). Given the large amount of significant differences with the original features, further analysis was done on the principal components. Pairwise t-tests on the 100 PCs revealed that 6 to 10 PCs had significant differences between each community pair with $\alpha = 0.05/100$ $= 0.0005$ in the discovery set. The comparison of the first five PCs across communities of the discovery set are shown in Figure 3.8. For all significant differences, the relative order of communities remained unchanged between the discovery and replication sets within each of the corresponding five PCs.

For example, the yellow community had the highest mean in PC 1 of the discovery set as well as in PC 1 for all replication sets, and the green community had the lowest mean in PC 2 for the discovery and each of the replication sets. There were significant differences in all PCs except between the blue and green communities of PC 1 in the 1^{st} replication set, and between the blue and yellow communities in PC 5 for the discovery and all of the replication sets. The top feature variables in these principal components were derived from brain white matter, EN-back, SST, cortical area, and cortical thickness [Tab. 3.2].

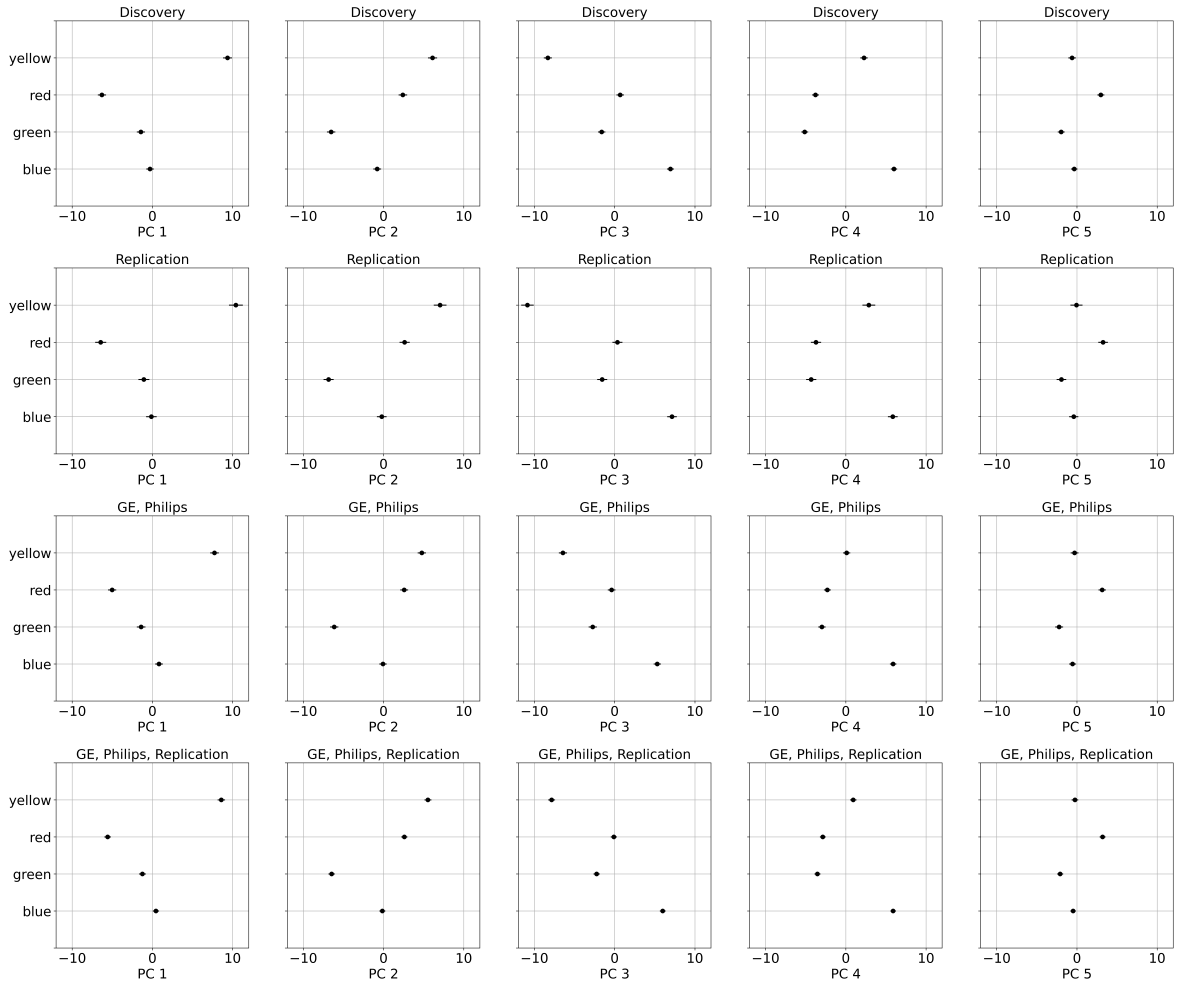

Figure 3.8 Multiple Comparisons Interval Plots (95% CI) of the Scores from each PC

Table 3.2 Contributing features for the top five PCs

PC	ABCD Imaging Feature Type	Top Feature Variables
1	EN-back (0-back, 2-back, place, emotion) APARC ROI	Posterior cingulate, precuneus, insula
	White matter (fractional anisotropy) Destrieux ROI	Superior frontal gyrus, superior temporal sulcus, middle frontal gyrus
2	SST (Correct Stop vs. Correct Go contrast) Destrieux ROI	Supramarginal gyrus, opercular part of the inferior frontal gyrus, superior frontal sulcus
	SST (Incorrect Stop vs. Correct Go contrast) Destrieux ROI	Inferior temporal gyrus, anterior transverse collateral sulcus, parahippocampal gyrus
	White matter (fractional anisotropy) Destrieux ROI	Precuneus, superior parietal lobule, cuneus
3	White matter (fractional anisotropy) Destrieux ROI	Superior parietal lobule, intraparietal sulcus and transverse parietal sulci, superior frontal gyrus
	EN-back (0-back, 2-back, place, emotion) APARC ROI	Posterior cingulate, caudal anterior cingulate, precuneus
	White matter (mean diffusivity) Destrieux ROI	Central sulcus, cuneus, middle occipital sulcus and lunatus sulcus
4	White matter (fractional anisotropy) Destrieux ROI	Orbital gyri, middle frontal sulcus, middle temporal gyrus
	Cortical Thickness Destrieux ROI	Superior temporal sulcus, superior parietal lobule, superior frontal gyrus
5	EN-Back (0-back, place) APARC ROI	Paracentral, caudal anterior cingulate, lateral occipital
	EN-Back (2-back vs. 0-back, 2-back) APARC ROI, 2-back vs. 0-back ASEG ROI	Right caudate, caudal anterior cingulate, posterior cingulate
	Cortical Area Destrieux ROI	Superior frontal gyrus, anterior part of the cingulate gyrus and sulcus, orbital gyri

Blue-colored rows represent features with positive weights. Red-colored rows represent features with negative weights.

CHAPTER 4
DISCUSSION

In a population of a healthy cohort of children ages 9-10, where we would typically assume no significant heterogeneity in brain structure and function, we found evidence of heterogeneous subgrouping based on the children's brain imaging derived features, indicating that brain "fingerprint" patterns can be grouped between individuals. In the discovery set, four communities were detected from community detection with graph networks formed from pairwise Pearson correlations between 1,685 brain imaging features among subjects. The multiplex modularity was around 0.5 for all graph networks labeled based on community detection or K-means clustering, indicating a moderate cohesion between the clusters of each network. Further investigation revealed that there were large amounts of significant differences between the communities from the t-tests of the 100 principal components as well as the original feature space of the discovery set. Additionally, in the top five principal components, there were significant differences between all clusters except for the blue and yellow community in the fifth principal component. Visualization of the t-SNE three-dimensional embedded space illustrated that the clusters are visually distinct with some overlap.

Most of the contributing features in the top five PCs are from brain white matter, especially in the PCs with the most negative contributions. Brain white matter is known to be highly unique between individuals and has been used as a local connectome fingerprint Yeh et al. (2016). Brain imaging features from EN-back task were also commonly associated with top PCs. EN-back fMRI is mainly used to measure working memory and emotional

processing. Previous research has shown evidence of functional connectome fingerprinting in predicting cognitive behavior Finn et al. (2015). These results are in line with significant differences found in cognitive assessments between the communities, specifically in picture vocabulary and reading test scores for all datasets.

Further investigation into our findings reveals some patterns that are specific to communities. For example, the red community was found to have significantly lower scores in the cognitive assessments and higher scores in early life stress than other communities (some differences were not significant). Moreover, the red community also had significantly higher values than most communities for male pubertal developmental status, indicating higher levels of maturity. We would expect higher early life stress with lower scores on the cognitive assessments based on previous research involving the association of cognitive decline with increased psychopathology Krugers & Joëls (2014). The red community also was found to have higher general p factor scores than the blue community for the replication and 2^{nd} replication sets. Higher early life stress has been found to be associated with increased risk of developing psychopathology Krugers & Joëls (2014). The red community also had significantly higher scores than other communities in the 5^{th} PC, indicating the key involvement of brain structure and functional features of the cingulate. The blue community had the highest age, Pattern Comparison Processing Speed Test scores, and modest pubertal developmental status. It also had the highest scores in the 3^{rd} and 4^{th} PCs, which are predominantly associated with brain white matter features and cortical thickness. Cortical thinning and white matter maturation are associated with brain development Tamnes et al. (2010). The green

community was found to have relatively higher cognitive assessment scores, lower early life stress scores, younger ages, and lower pubertal developmental status (though some of the differences may not be significant). The green community also has the lowest scores in the 2^{nd} PC. The yellow community that had the highest scores in 1^{st} and 2^{nd} PCs was found to have lower pubertal developmental status, higher cognitive assessment scores, and lower early life stress scores (some differences were not significant). Overall, in all the differences that were significant, there were no differences in the order of the clusters between each scanner dataset, indicating that there was no evidence of scanner effects.

CHAPTER 5
CONCLUSION

We found evidence of heterogeneous subgrouping based on children's brain features, indicating that brain fingerprint patterns can be grouped between individuals. Future studies should consider further investigation into this subgrouping, as well as ways to account for this heterogeneity.

REFERENCES

Barch, D. M. et al. 2018, Developmental cognitive neuroscience, 32, 55

Bornovalova, M., Choate, A. M., & Wiernik, B. M. 2020, Biological Psychiatry, 88, 18

Cai, B. et al. 2021, Human Brain Mapping, 42, 2691

——. 2020, arXiv preprint arXiv:2006.09928

Caspi, A. et al. 2014, Clin Psychol Sci, 2, 119

Fahle, M., & Spang, K. 2003, International Review of Sociology, 13, 507

Farahdel, B., Thapaliya, B., Suresh, P., Ray, B., Calhoun, V. D., & Liu, J. 2021, in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE), 3540–3547

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. 2015, Nature Neuroscience, 18, 1664

Fortin, J.-P. et al. 2018, NeuroImage, 167, 104

Hagberg, A., Swart, P., & S Chult, D. 2008, Exploring network structure, dynamics, and function using NetworkX, Report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States)

Hagler Jr, D. J. et al. 2019, Neuroimage, 202, 116091

Krugers, H. J., & Joëls, M. 2014, Behavioral neurobiology of stress-related disorders, 81

Magnani, M., Hanteer, O., Interdonato, R., Rossi, L., & Tagarelli, A. 2021, ACM Computing Surveys (CSUR), 54, 1

Newman, M. E., & Girvan, M. 2004, Physical review E, 69, 026113

Nishiyama, T. et al. 2020, Comprehensive Psychiatry, 96, 152148

Pedregosa, F. et al. 2011, the Journal of machine Learning research, 12, 2825

Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. 2010, Cerebral cortex, 20, 534

Thapaliya, B., Calhoun, V. D., & Liu, J. 2021, 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2330

Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B. J., Nixon, S. J., & Luciana, M. 2019, Developmental cognitive neuroscience, 36, 100606

Traag, V. A., & Bruggeman, J. 2009, Physical Review E, 80, 036115

Traag, V. A., Waltman, L., & Van Eck, N. J. 2019, Scientific reports, 9, 5233

Valizadeh, S. A., Liem, F., Mérillat, S., Hänggi, J., & Jäncke, L. 2018, Scientific reports, 8, 5611

Yeh, F.-C., Vettel, J. M., Singh, A., Poczos, B., Grafton, S. T., Erickson, K. I., Tseng, W.-Y. I., & Verstynen, T. D. 2016, PLOS Computational Biology, 12, e1005203

Zhu, A., Hua, Z., Shi, Y., Tang, Y., & Miao, L. 2021, Entropy, 23, 1550