# Topological Structure Learning and Classification in Individuals with Autism Spectrum Disorders and Typical Controls - An fMRI analysis

**Qihang Wu***
Department of Biostatistics
Mailman School of Public Health
Columbia University
New York, NY 10032, USA

## Abstract

Neuroimaging technique is a useful tool nowadays. One popular technique is functional magnetic resonance imaging (fMRI) analysis. The objective of this project is to apply two network modeling algorithms: the Peter and Clark (PC) and the Max-Min Hill Climbing (MMHC) to understand the topological structures of the functional networks of individuals with Autism Spectrum Disorders (ASD). Based on the global metrics of graphs, we distinguish ASD patients from the typical controls (TC). The result shows the logistic regression model-based MMHC actually outperforms PC with respect to the prediction abilities.

## 1 Introduction

Nowadays, neuroimaging has become a powerful tool for studying the structure and function of the human brain in a non-invasive way. The researchers can then interpret findings and understand the relevant neural activities or brain structural changes. Among multiple types of brain imaging techniques, functional magnetic resonance imaging is a skill to detect regional and time-varying changes in brain metabolism especially related to human cognition. A common way to perform the fMRI analysis is to restrict the analysis to specific spatial regions of interest (ROIs). From each ROI, the time series of blood oxygen level-dependent (BOLD) volume of each cube (i.e., voxel) are extracted and recorded to represent the neuronal activity within the ROI. ASD is one of the most serious neuropsychiatric disorders worldwide, affecting one's daily interactions and thereby impairing communications with others. According to CDC, ASD occurring before the age of 3 will also have a negative and devastating influence throughout one's life.

Despite some challenges embedded in the field of fMRI analysis such as the noises and some unknown psychological signals and the high dimensionality of the data in contrast to the sample size, the probabilistic graphical modeling, fortunately, sheds light on these issues.

Based on the fMRI time series data, this project takes advantage of the graphical modeling to investigate the topological difference of the functional connectivity network (FCN) between people with ASD vs. health controls, that is, to determine the connection between any pairs of ROIs. Although directionality is important for causal inference, it will not be a focus here. According to the structural learning results, we try to distinguish the ASD cases from the healthy controls using logistic regression, a generalized linear model.

*Email: qw2331@cumc.columbia.edu

Preprint. Under review.

## 2 Data

The Autism Brain Data Exchange (ABIDE) initiative is to aggregate both functional and structural brain imaging data from 17 international sites to help the scientific community better understand the mechanism behind ASD. In this project, we use the fMRI data during the resting state (rs-fMRI) collected from Carnegie Mellon University (CMU).

In the data, 27 subjects aged 19-40 years underwent the MRI scanning with a SIEMENS scanner, 14 of them were previously diagnosed with ASD (those individuals were medically healthy without identifiable genetic etiology for their disorder) while the rest 13 were regarded as the typical controls. The main parameters involved are echo time (TE) = 30 ms, repetition time (TR) = 2000 ms, and 3.0mm$^3$ voxels. Data was downloaded from ABIDE website and preprocessed using the NeuroImaging Analysis Kit (NIAK) pipeline, which parcelled the data into Dosenbach's 160 functional ROIs[1]. Within each ROI, the average BOLD signals were calculated.

To describe the data in a more quantitative way, we have $\tilde{y}_{ktp} = \{\mathbf{y}_{k1}, \mathbf{y}_{k2}, ..., \mathbf{y}_{kp}\}$ to represent the $p$-th ROI for the $k$-th subject at the corresponding $t$-th timepoint (s.t., $t \in \{1, ..., T\}$), where $\mathbf{y}_{kp}$ is a random variable generating T samples of BOLD signals of $p$-th ROI for the subject $k$. The number of time points varies by different subjects since some samples are dropped through data preprocessing steps to remove potential artifacts or outliers. Meanwhile, the supplementary data contains some phenotypic information such as sex, age, estimated IQ, handedness, etc.

## 3 Methods

To better compare different structural learning algorithms and apply the findings further for class prediction, this project includes three parts: **Exploratory Data Analysis (EDA)**, **Individual-Level Analysis**, and **Group-Level Analysis**. The data preprocessing steps and related methods are described in this report as follows along with the parameters tuning strategies.

### 3.1 Data Preprocessing & EDA

No many additional data processing steps are required for the original fMRI data[2]. For further analysis, we have also incorporated the six identifiable adult rs-fMRI networks (i.e., cingulo-opercular, fronto-parietal, default, occipital, sensorimotor, and cerebellum) based on Dosenbach, N. U. F, et.al [1]. For the demographic table, the p values for continuous variables and categorical variables are calculated using Welch Two Sample t-test and Pearson's Chi-squared test, respectively.

### 3.2 Structural Learning Algorithms

We first evaluate Pearson's correlation and partial correlation between each pair of ROIs since they are the simplest measure of pairwise similarity between two ROIs [2]. For time series data, Pearson's correlation is defined as follows [3]:

$$\hat{\rho}_{i,j} = \frac{\sum_{t=1}^{T}(x_i^t - \bar{x}_i)(x_j^t - \bar{x}_j)}{\sqrt{\sum_{t=1}^{T}(x_i^t - \bar{x}_i)^2 \sum_{t=1}^{T}(x_j^t - \bar{x}_j)^2}} \tag{1}$$

where $\bar{x}_i$ and $\bar{x}_j$ are the sample means of the BOLD signals of $i$-th ROI and $j$-th ROI, respectively. It is also called *Full correlation* if the time series data are normalized to unit variance. Meanwhile, each element from the resulting matrices is normalized so that all elements are ranging from 0 to 1. The element with a higher score represents the higher probability or confidence of the existence of a connection. We also implement two structural learning algorithms: a constrained-based PC algorithm and a hybrid MMHC algorithm.

#### 3.2.1 Peter and Clark (PC) algorithm

As one the oldest algorithms for causal discovery, the rationale behind it is to learn the CPDAG (completed Partially Directed Acyclic Graph), an equivalence class of DAGs (Directed Acyclic

[2]As all 27 subjects' fMRI data include an extra ROI (column 161) instead of 160 ROIs, we remove those columns first and also rename all data files as .csv files.

Graphs) through a sequence of consistent conditional independence tests under no latent confounders assumption. By the DAG Markov properties, it searches for the conditional sets containing all adjacent vertices of any two interested vertices $X_i$ and $X_j$ to determine the independence of these two vertices, examine colliders via v-structure, and apply several orientation rules to orient the remaining undirected edges. Here we use Fisher's z-transformed of the partial correlation for testing on Gaussian distributed random variables[3]. Parameter tuning with respect to $\alpha$[4] is discussed in the next section.

#### 3.2.2 Max-Min Hill Climbing (MMHC) algorithm

While PC is different from the GES (Greedy Equivalence Search), a score-based learning algorithm, MMHC combines both ideas of constraint-based and score-based techniques in a more efficient way. The first step of this algorithm is to construct the "skeleton" of a Bayesian network based on a local discovery algorithm MMPC (Max-Min Parent and Children) with multiple conditional independence tests (a similar procedure to PC but requiring fewer test on average) and it then performs a greedy hill-climbing search to determine the orientations of edges. Hill climbing is commonly used in the field of artificial intelligence, although it is heuristic and cannot guarantee a global maximum. It is believed that this algorithm will outperform some state-of-the-art algorithms including PC and GES in empirical evaluations [4]. The parameter tuning for $\alpha$, a significance level for assessing the p values, is also described in the next section.

### 3.3 Parameter Tuning

The tuning of $\alpha$ for PC is always tricky. Although some investigators have proposed several strategies including choosing the lowest BIC score with the assumption of parametric models and performing non-parametric StARS (Stability Approach to Regularization Selection) algorithm, both have their own drawbacks such as strict assumptions and computational intensity [5]. As we have a large number of variables in the data, we intend to select smaller $\alpha$ so as to reduce the density of the estimated graph and benefit the computational efficiency. Therefore, we randomly select one subject and then implement the PC algorithm with a grid search of $\alpha$ from 0.0067 to 0.135 (-5 to -2 in the log scale). The optimal $\alpha$ chosen in this case (i.e., 0.0383) is the one where the number of edges is stable.

The procedure for choosing $\alpha$ in MMHC is similar to the PC, although we also try to make the number of edges closer to that in the PC within the same subject. To achieve that, a range of $\alpha$ from 0.0025 to 0.050 (i.e., log(-6), log(-3)) is used and final $\alpha = 0.0088$ is decided for further analysis.

### 3.4 Graph Measures & Metrics

For each algorithm mentioned above, the graph global and topological measures are calculated for each individual. We include a total of six metrics for the unweighted and undirected graphs here: **Eccentricity**, **Mean Distance**, **Assortativity**, **Transitivity**, **Modularity**, and **Density**. The number of metrics selected is limited as we have a relatively small sample size. These measurements are described below in simple terms.

- **Eccentricity:** The eccentricity $\epsilon$ of a graph $\mathcal{G}$ is computed by averaging eccentricity $\epsilon_i$ over all vertices in the graph where $\epsilon_i$ of a specific node $i$ is defined as the longest hop count[5] between this vertex and the others. This measures the distance-link property of one graph.
- **Mean Distance:** This metric evaluates the average shortest paths from one vertex to another.
- **Assortativity:** As a class of connection metrics, assortativity is used to compute the correlation among the vertices' degrees based on Pearson's correlation coefficient and describe the similarity of one vertex to another. This coefficient for an undirected graph is given as:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \tag{2}$$

  where $q_k$ is the distribution of the remaining degree and $e_j k$ represents the joint distribution of the remaining degrees of the two vertices $j$ and $k$.

[3]Conditional independence tests for discrete and binary variables are also available

[4]Threshold for controlling the sparsity of the output, a higher $\alpha$ will render denser graphs and vice versa.

[5]Hop count of a path $\pi$ is the number of links given a path between vertex $i$ and vertex $j$, where $i \neq j$.

- **Transitivity:** Also in the connection class, it reveals the existence of tightly connected clusters (or communities) within a graph. It is sometimes called the clustering coefficient. In this project, we use the "global" argument in the function, which is simply the ratio of the count of the connected triples in an undirected graph.
- **Modularity:** It measures the density of connections within a community. A higher module usually represents the dense intra-communities edges while sparse inter-communities edges. We define the communities based on the adult brain maturity networks from Dosenbach [1].
- **Density:** It is calculated by dividing the number of edges by the number of possible edges.[6]

### 3.5 Classification Prediction

As we have a limited sample size (i.e., 27 subjects in total), we decide not to use a complex model for class fitting and prediction. To fit this generalized linear model, we split[7] the original 27 subjects into 85% for the training set and the remaining 15% for prediction evaluation. For each observation (i.e., subject), the above graph metrics are calculated based on either PC or MMHC. Evaluation for the prediction of the test data is then based on accuracy, sensitivity, and specificity.

## 4 Results

The demographic information is shown below.

Table 1. Demographics for People with ASD vs. TC

| Characteristics | **ASD**, N = 14[a] | **TC**, N = 13 | p-value[b] |
|---|---|---|---|
| **Age, yrs** | 25.5 (21.2, 30.8) | 27.0 (21.0, 30.0) | 0.8 |
| **Sex** | | | >0.9 |
| Male | 11 (79%) | 10 (77%) | |
| Female | 3 (21%) | 3 (23%) | |
| **Handedness** | | | 0.6 |
| Ambi | 1 (7.1%) | 1 (7.7%) | |
| L | 1 (7.1%) | 0 (0%) | |
| R | 12 (86%) | 12 (92%) | |
| **Intelligence Quotient** | | | |
| Full-Scale (FIQ) | 117 (105, 123) | 110 (109, 124) | >0.9 |
| Verbal (VIQ) | 111 (99, 120) | 111 (108, 122) | 0.7 |
| Performance (PIQ) | 115 (108, 121) | 109 (108, 123) | 0.7 |

[a]Median (IQR); n (%)

[b]Welch Two Sample t-test; Pearson's Chi-squared test

**Table 1** shows there is no statistical significance between ASD vs. TC across all demographic characteristics including age, sex, handedness, and IQ scores, i.e., $P>0.05$, which means that the distribution of these two groups is relatively balanced and potential confounding problems are minimized in this case, at least for these covariates.

### 4.1 Individual-Level Analysis

Based on the selected tuning parameter $\alpha$ above for both PC and MMHC, we randomly select one sample from each group (ASD vs. TC). Pearson's correlation and partial correlation are calculated and compared. A higher score (i.e., those regions colored in red) represents higher confidence in connections. **Figure 1** illustrates that the red area in the typical control is larger than that in an ASD patient for both correlations, which may imply the denser connections for a TC subject and that the distinction between these two disease groups based on the network connection may be reasonable. This correlation is simple and intuitive. However, it does not consider possible confounders under each pair of connections, that is, the differentiation between the direct and indirect effects is difficult. Partial correlation is then used to overcome such a disadvantage by controlling all the other vertice,

[6]In our case, the possible number of edges in a graph is 12,720.

[7]The data partition is based on the distribution of outcomes, i.e., ASD vs. TC.

although it has its own shortcomings.

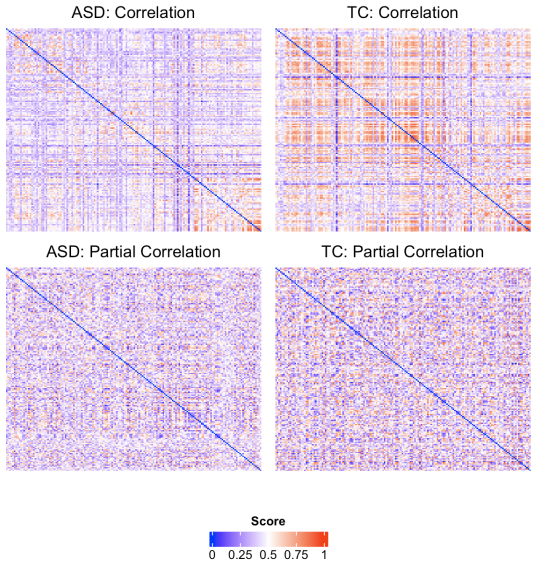


Figure 1: Correlation and Partial Correlation of ASD vs. TC

To simply compare the topological structures from both disease groups, we make the undirected graphs according to the PC algorithm. The resulting plot (**Figure 2**) is then color-labeled based on their corresponding networks as mentioned before. The layout is modified so that the vertices within the same network are closer than the others. The vertices in TC are more widely distributed compared to ASD, which may indicate better and well-separated communities (or networks) for those typical controls on average. This speculation is verified by selecting different individuals from both groups.

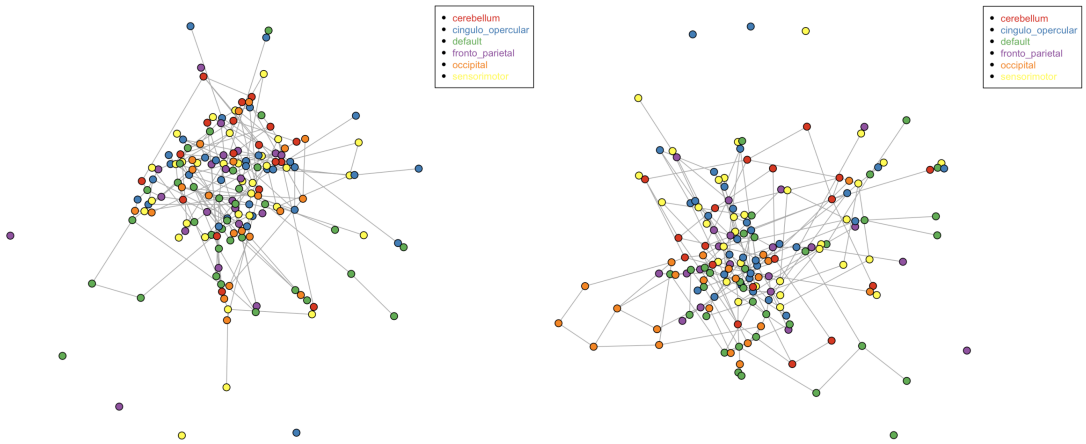


Figure 2: Topological Structure of ASD (L) vs. TC (R) from PC

### 4.2 Group-Level Analysis

For group-level analysis, we perform both PC and MMHC algorithms on all 27 subjects under the previously selected parameters and calculated those six global metrics for each one. The evaluation of the testing data demonstrates the possibility of utilizing the topological structures (i.e., global graph features) to classify ASD and TC. The result shows the prediction accuracy for PC is around 0.67 while for MMHC, it is close to one. Compared with the random guess of 50%, these two algorithms

are feasible but not the best for our dataset. We have also varied the ratio of sample size between training and testing, the result is consistent across ratios.

# 5 Discussion

## 5.1 Conclusions

The demographic table indicates the balanced distribution across all indicated covariateS (i.e., P$>$0.05.) For the single-subject level analysis, the heatmaps of both Pearson's correlation and partial correlation show the denser connections in the TC group. However, they have their own weaknesses. The community-based graphs not only illustrate the topological structure within each disease group but reveal the differences between them: graphs from the TC group are more well-separated than the ASD group. Both PC and MMHC provide relatively promising predictive abilities under the current dataset with an average prediction accuracy of about 0.83.

## 5.2 Limitations

Instead of considering the spatial ROIs as the vertices, independent component analysis (ICA) may be another suitable tool for determining the vertices of the network.

The parameter tuning for the sparsity $\alpha$ is always tough and more holistic considerations are needed for further analysis. Here we only choose the one based on a single randomly selected subject. Subjects from different disease groups may require different thresholds. Meanwhile, the criterion we used is to find the $\alpha$ so that the number of edges will be stable, however, the term "stable" is quite subjective, and such a criterion still requires examinations.

The figures of topological structures from both ASD and TC are not intuitive enough and the distribution of every vertex is not just in two dimensions as they represent the regions of interest in human brains.

There are several problems with class prediction in current data analysis. 1) extra data preprocessing may be needed before running analyses; 2) we could only utilize the simpler model for classification due to the small sample size; 3) the choice of the features for prediction may not well-describe a specific graph, although we have tried to cover as many different types of properties as possible; 4) the accuracy for prediction may not be reliable as the size of the testing set is small.

## References


[1] Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., Nelson, S. M., Wig, G. S., Vogel, A. C., Lessov-Schlaggar, C. N., Barnes, K. A., Dubis, J. W., Feczko, E., Coalson, R. S., Pruett, J. R., Barch, D. M., Petersen, S. E., Schlaggar, B. L. (2010). Prediction of Individual Brain Maturity Using fMRI. Science, 329(5997), 1358–1361. http://www.jstor.org/stable/41075813

[2] Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., Woolrich, M. W. (2011). Network modelling methods for FMRI. NeuroImage, 54(2), 875-891. https://doi.org/10.1016/j.neuroimage.2010.08.063

[3] Nie, L., Yang, X., Matthews, P.M. et al. Inferring functional connectivity in fMRI using minimum partial correlation. Int. J. Autom. Comput. 14, 371–385 (2017). https://doi.org/10.1007/s11633-017-1084-9

[4] Tsamardinos, I., Brown, L.E. Aliferis, C.F. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn 65, 31–78 (2006). https://doi.org/10.1007/s10994-006-6889-7

[5] Strobl, E. V. (2021). Automated hyperparameter selection for the PC algorithm. Pattern Recognition Letters, 151, 288-293. https://doi.org/10.1016/j.patrec.2021.09.009


## A Appendix

For codes in detail please click here.