

Final Project: Graphical Models for ASD Connectivity

Zhenkun Fang - zf2352

2025-12-16

Goal and Motivation

The scientific goal of this project is to compare the brain connectivity networks of individuals with Autism Spectrum Disorder (ASD) versus Neurotypical (NT) controls. Specifically, I aim to test if the global network density (the proportion of active functional connections) differs significantly between these two groups.

Characterizing functional connectivity differences is essential for understanding the neurobiology of ASD. Graphical models allow us to move beyond simple correlations and estimate direct interactions (conditional dependence) between brain regions. If ASD brains show consistently higher or lower connectivity density, this could serve as a potential biomarker.

Data

The dataset comes from the ABIDE exchange. It contains fMRI timeseries for 27 subjects (14 ASD, 13 Controls) parcellated into 160 regions of interest (ROIs).

Exclusions/Preprocessing:

As instructed, we remove the 161st column from the data. We use the phenotypic_CMU.csv file to identify the diagnosis group (DX_GROUP: 1=ASD, 2=Control).

Methods

I will compare two methods to estimate the connectivity graphs.

1. Gaussian Graphical Model (GGM) via Graphical Lasso

This is the primary method. We assume the data follows a multivariate Gaussian distribution (or non-paranormal). The graph edges represent non-zero entries in the precision matrix (conditional independence).

- Algorithm: We use the Graphical Lasso (glasso) algorithm implemented in the huge package.
- Tuning: We use the Stability Approach to Regularization Selection (StARS) to choose the regularization parameter λ . StARS is robust for high-dimensional biological data as it selects the sparsest graph that is stable under subsampling (threshold set to 0.1).

2. Correlation Thresholding

As a baseline comparison, I will use a marginal independence graph.

- Algorithm: Compute the Pearson correlation matrix and threshold edges where $|r| > 0.3$.
- Interpretation: This does not control for other variables (marginal independence), unlike the GGM.

Statistical Test

For each subject, I will calculate the Network Density:

$$D = \frac{2E}{p(p - 1)}$$

where E is the number of edges and $p = 160$ is the number of nodes. I will then perform a Two-Sample t-test to compare the mean density between ASD and Control groups.

Results