Structure of and Inquiries about Healthy Connectivity Matrix Data

Nooreen S Dabbish and Kaijun Wang

Feb 1, 2016

Outline

Basics

How it is organized. What types; what measurements are selected?

Connectivity Matrices

- symmetric/undirected n x n matrix
- n x 3 list of Montreal Neurological Institute (MNI) coodinate spaces for each node
- n x 1 lists of node names and abbreviations
- [data types] fMRI, Diffusion Tensor Images (DTI), Diffusion Spectrum Images (DSI), structural MRI, EEG, MEG

http://umcd.humancommectomeproject.org

Description in (Brown 2016)

- central repository for connectivity matrices
- click-of-the-mouse analyses
- (as of 1/17/16) 2254 brain networks (CMs), 21 studies
- all ages fetus to 89 yo.
- healthy, ADHD, autism, OCD, APOE-4 carrier status (risk for AD)

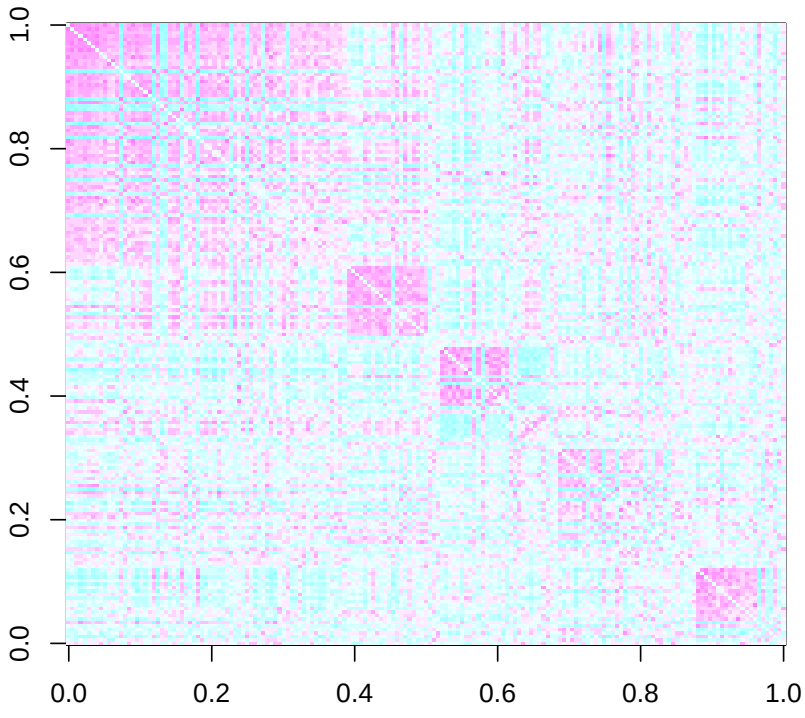
Bejing Zang

The dataset to be used: R-fMRI BOLD data from 148 subjects (74 female and 74 male; matched by age (21 years old) recruited as part of larger studies conducted in Beijing China. See the MATERIALS AND METHODOLOGY, Resting State Data of `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3271304/pdf/TONIJ-6-1.pdf`

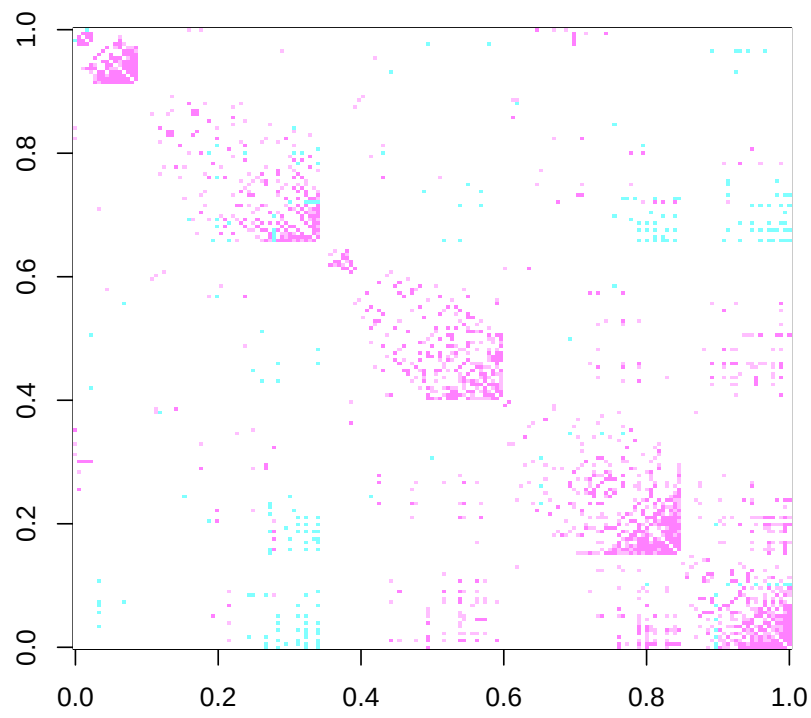
Overview

number of samples/variables/..histogram plot of some measurement to give a "feel" for the data. List of interesting modeling questions and why they are scientifically interesting?

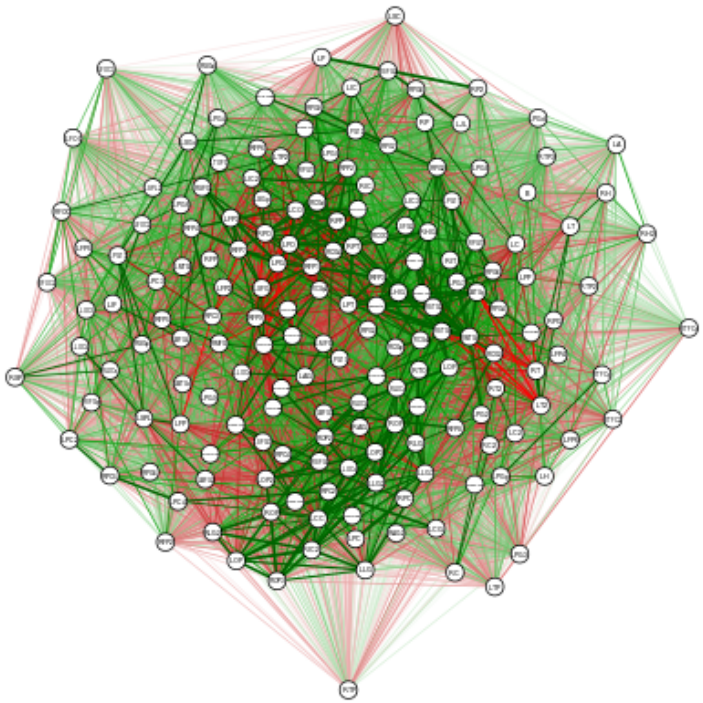
Before Thresholding



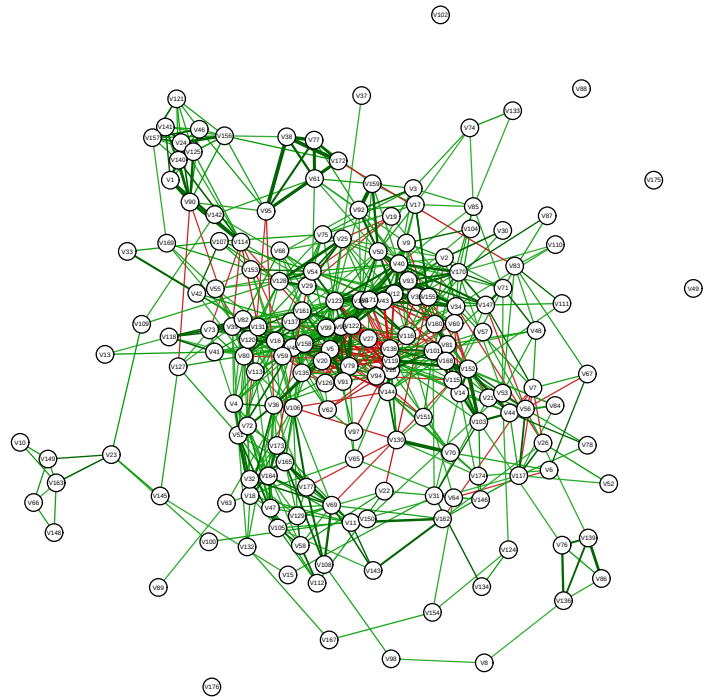
After thresholding (with 0.5)

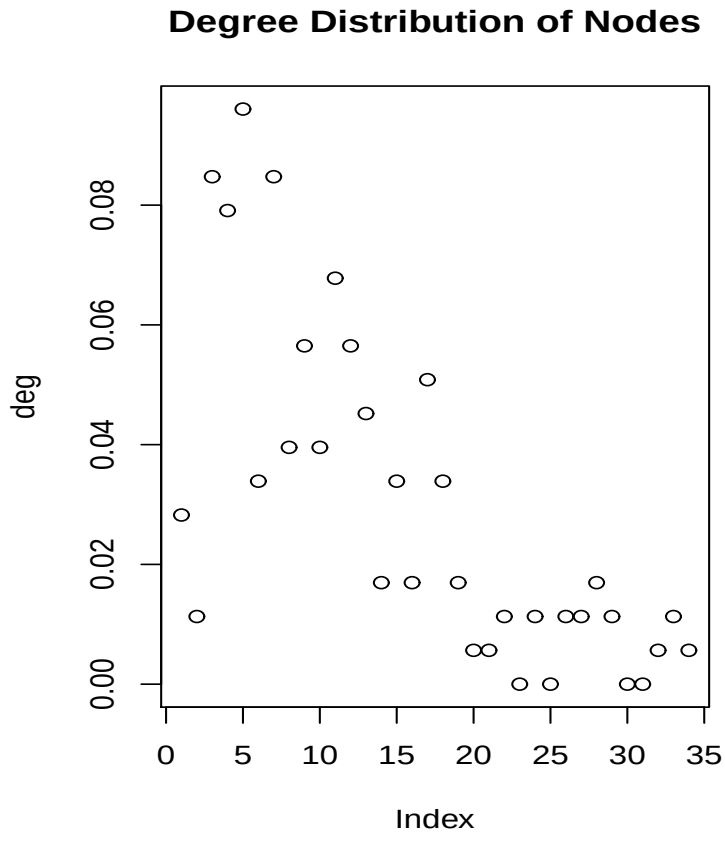


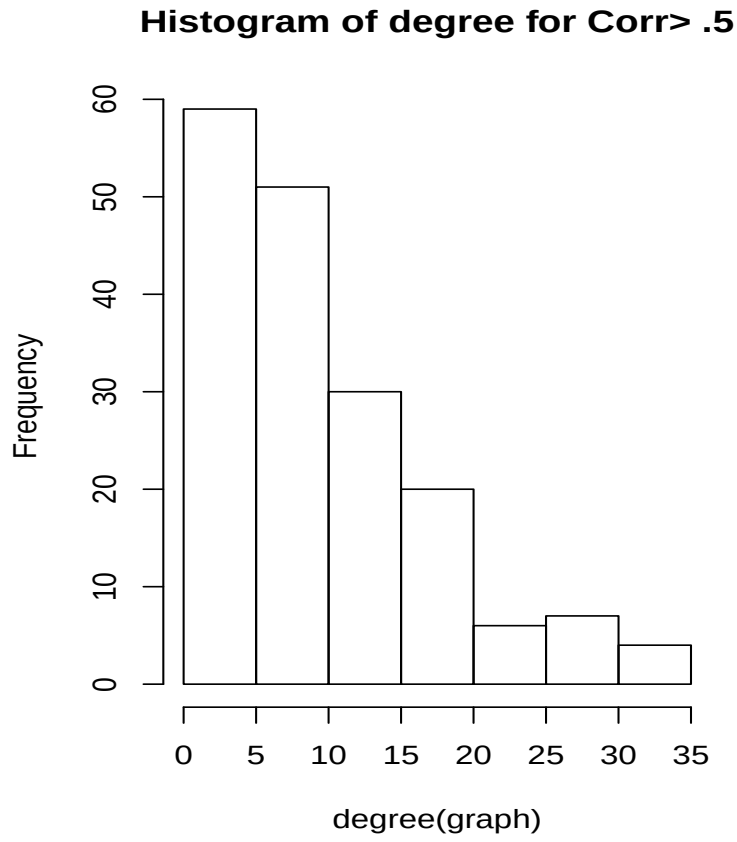
Before Thresholding



After Thresholding




Degree Distribution of Nodes
deg
0.00
0.02
0.04
0.06
0.08
0
5
10
15
20
25
30
35
Index


Histogram of degree for Corr> .5
Frequency
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35
degree(graph)

Literatures: On fMRI Data

- Network Centrality in the Human Functional Connectome Xi-Nian Zuo, et al. (2011)
- Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space Xi-Nian Zuo, et al. (2013)
- Fully exploratory network independent component analysis of the 1000 functional connectomes database Klaudius Kalcher, et al.(2013)

Literatures: Classification Problems

- Combining Graph and Machine Learning Methods to Analyze Differences in Functional Connectivity Across Sex (2012) Casanova R., et al.
  Response: Gender
  Random forest and Lasso
- The Kernel Two-Sample Test for Brain Networks, Emanuele Olivetti et al. (2015)
  Response: Gender
  Classification based method vs Kernel Two-sample

Literatures: Classification Problems

- Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data Damien A. Fair, et al. (2013)
  Response: ADHD
- The Autism Brain Imaging Data Exchange: Towards Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism Di Martino, et al.(2014)
  Response: Autism Spectrum Disorders
  Intrinsic Functional Connectivity Analyses between Autism and Control Group

Possible Topics

- Classification according to the connectivity matrices, given exogenous data.
- Signal Path Clustering, normal pattern in brain activities
- Evaluate different classification methods on this data (e.g. random forest and lasso in [Cassanova 2012])

Graph Kernel

Graph pairs $G_i, G_j \rightarrow$ Kernel $k_{ij} = \kappa(G_i, G_j) \rightarrow$ Classifier

KTST

Functional connectomes, 148 graphs, 74 from group A and 74 from group B
Kernel matrix: 148-by-148, of pairwise kernels

graphs	A_1	A_2	$\cdots$	A_m	B_1	B_2	$\cdots$	B_n
A_1	$k_{A1,A1}$	$k_{A1,A2}$	$\cdots$		$k_{A1,B1}$	$\cdots$		
A_2	$k_{A2,A1}$							
$\vdots$	$\vdots$		$\ddots$					
A_m	$k_{Am,A1}$	$\cdots$						
B_1	$k_{B1,A1}$				$k_{B1,B1}$	$k_{B1,B2}$	$\cdots$	
B_2		$\ddots$			$k_{B2,B1}$			
$\vdots$	$\vdots$				$\vdots$		$\ddots$	
B_n	$k_{Bn,A1}$							

KTST

maximum mean discrepancy (MMD):

$$MMD^2 = E[k(x_A, x'_A)] + E[k(x_B, x'_B)] - 2E[k(x_A, x'_B)]$$

$$\widehat{MMD}^2 = \frac{1}{m(m-1)} \sum_{i \neq j} k(x_i^A, x_j^A) + \frac{1}{n(n-1)} \sum_{i \neq j} k(x_i^B, x_j^B) - 2\frac{1}{nm}$$

Test hypothesis: $H_0 : MMD = 0$

KTST

- MMD changes with different matrix permutations
- in practice we calculate different MMDs according to various permutations, to see its distribution
- Total calculation complexity: $M \cdot N^2$, where, M is the number of permutations, and N is the number of graphs.

Factor Analysis

- input factor number k, covariance matrix A
- find k factors $F_1, \cdots F_k$ according to A
- find projection from factors to the original variables in A

$$X_1 = \lambda_{11} F_1 + \lambda_{12} F_2 + \cdots + \lambda_{1k} F_k$$
$$\cdots$$
$$X_1 = \lambda_{n1} F_1 + \lambda_{n2} F_2 + \cdots + \lambda_{nk} F_k$$

Factor Analysis

- The coefficients λ_{ij} are called loadings, group X_is according to their highest loading into k groups
- permute the original matrix

Example:
if we have two groups with $\{X_1, X_2, X_4\}$, $\{X_3, X_5\}$ the original matrix will exchange the third and fourth row and column.