Integrative brain network analysis Using Gaussian Graphical Models

Suprateek Kundu

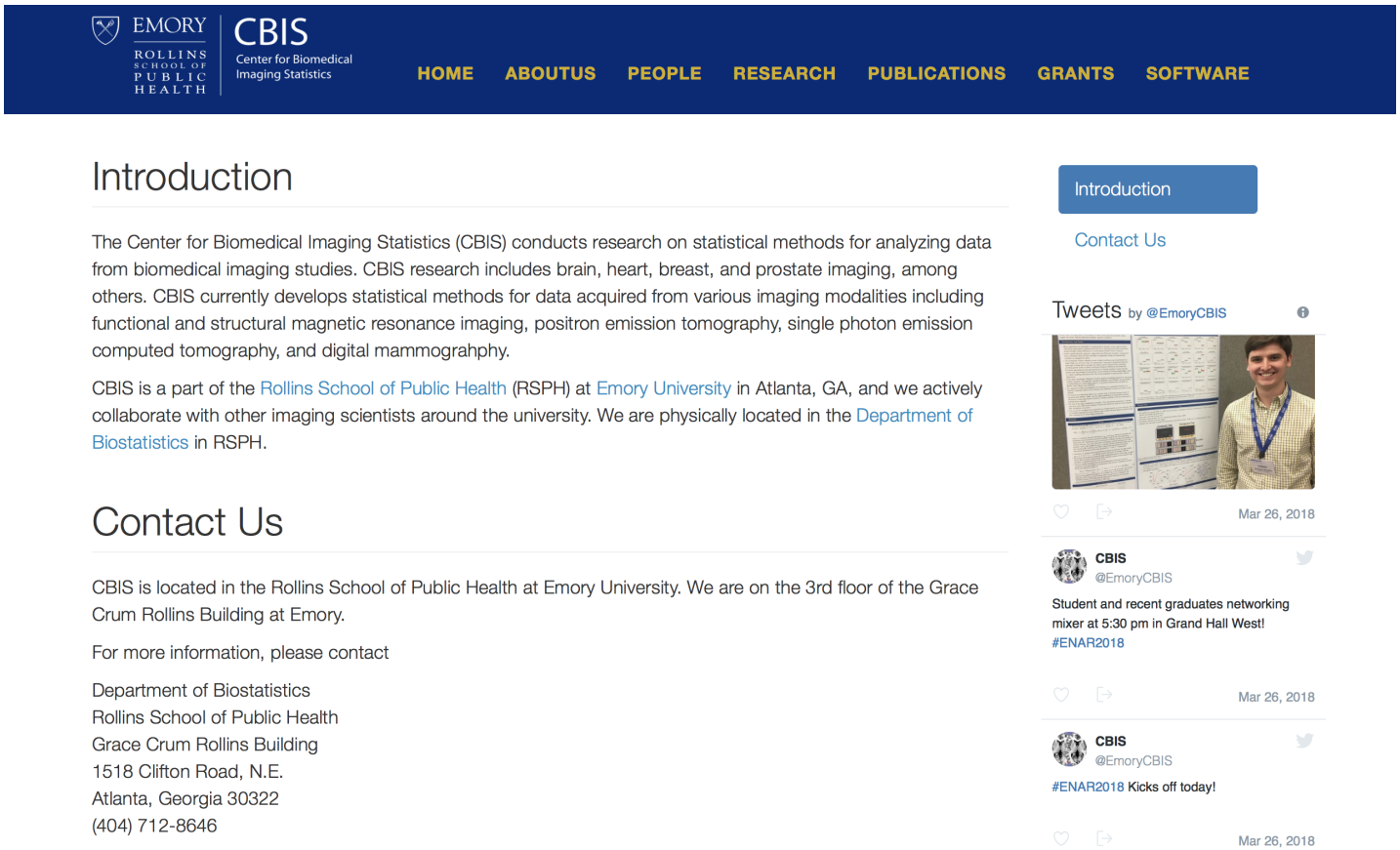
Department of Biostatistics and Bioinformatics
Emory University

October 17, 2019

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Funding Sources

- My research was supported by National Institutes of Mental Health grant number R01MH120299 and Georgia CTSA grant ULITR002378
- Center for Biomedical Imaging Statistics, Emory University



1 Introduction

2 Bayesian Joint Network Learning
- Background
- Model
- Simulation
- Stroop Task Data
- Summary

3 Anatomically Guided Brain Functional Connectivity
- Simulation
- Philadelphia Neurodevelopment Cohort

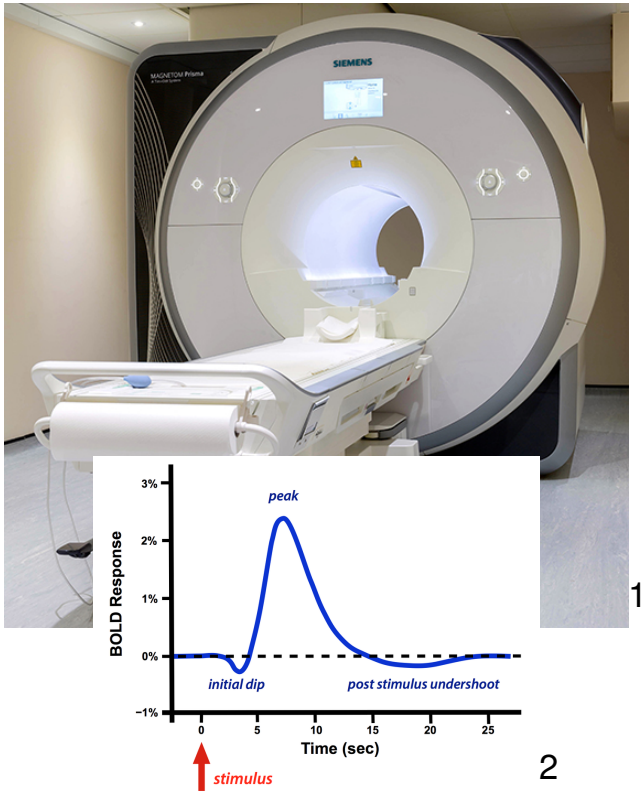
Papers

This talk is based on the following papers:

- Lukemire*, J., **Kundu**, S., Pagnoni, G., and Guo, Y. (2019+). Bayesian joint modeling of multiple brain functional networks. *JASA A&CS*, Accepted conditional on reproducibility review. https://arxiv.org/abs/1708.02123.
- Higgins*, I., **Kundu**, S. and Guo, Y., 2018, Integrative Bayesian analysis of brain functional networks incorporating anatomical knowledge, *NeuroImage*, Volume 181, Pages 263-278

* indicates student advisee

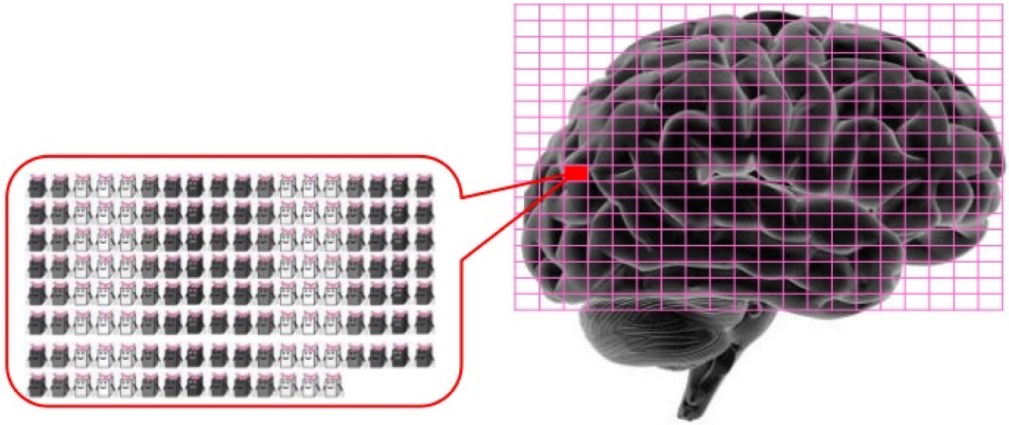
fMRI Background



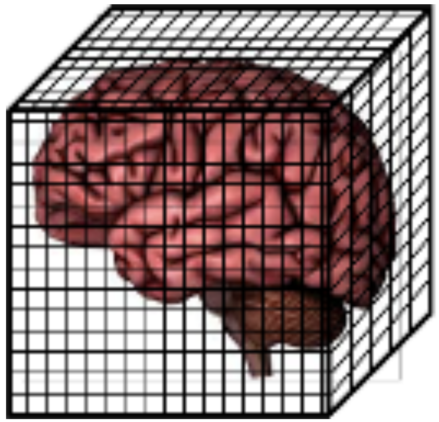
1 2

Blood Oxygen Level Dependent Signal

Preprocessing →



4D Data (width, height, depth, time series)



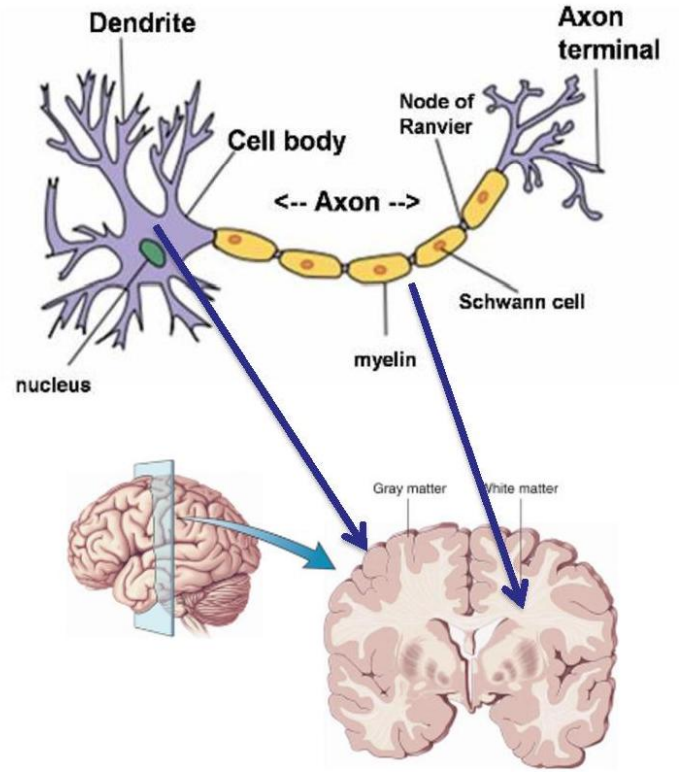
1 https://www.ndcn.ox.ac.uk/divisions/fmrib/what-is-fmri/introduction-to-fmri
2 http://mriquestions.com/deoxy-oxyhb-activity.html

Brain Tissues

Two main types of brain tissue:

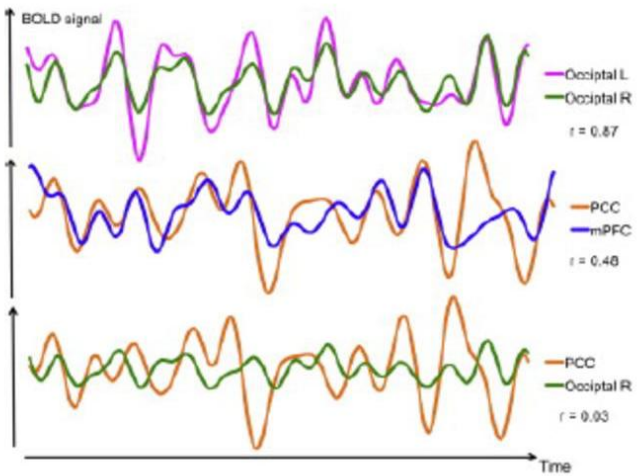
- **Gray Matter**
  - Makes up the surface of the cortex
  - Composed of neuron cell bodies, dendrites

- **White Matter**
  - Lies beneath the cortex
  - Composed of myelinated axons (fiber tracts) that connect nerve cells
  - Enables communication between grey matters by carrying electrical impulses.



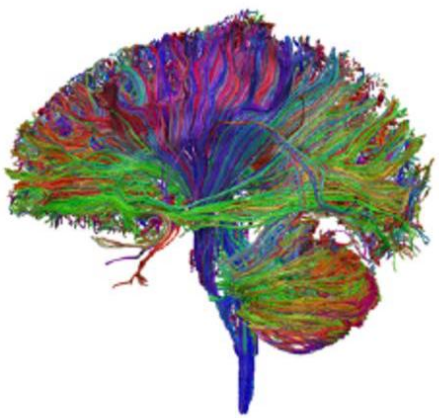
Types of Connectivity

Functional Connectivity: temporal coherence of activity between brain areas



(fMRI data)

Structural Connectivity: anatomical connections (white matter tracts) between brain areas



(DTI data)

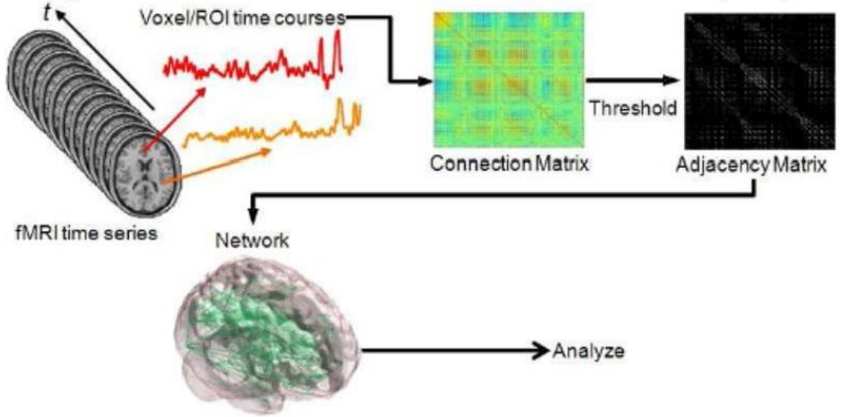
Functional Connectivity

There are many different approaches to studying functional connectivity, each with their own advantages and drawbacks.

- Seed voxel/ROI analyses - pick out a region of the brain and measure the correlation of its time course with the time courses of all other regions
- Graph Theoretical Approach - View the brain as a set of nodes and edges - estimate the edge set.
- Based on Gaussian graphical models (GGM) which will focus for this talk

Functional Connectivity

Schematic for generating brain networks from fMRI time series data. (Simpson et al., 2013)



- Brain Network Representation: $V \times V$ matrix (V: number of nodes).
  - Σ: covariance matrix for marginal connectivity
  - Ω: precision matrix for direct connectivity
- Steps in brain network analysis:
  - Defining nodes (brain parcellations)
  - Network Estimation
  - Thresholding
  - Network comparisons between groups (e.g. normal vs. depressed)

Functional Connectivity - Challenges

Current challenges in brain network analysis:

- Often unsatisfactory reliability and reproducibility due to low signal to noise ratio in fMRI
- Network comparisons between groups may not be straightforward
- No obvious way to fuse data across multiple fMRI experiments or across two imaging modalities
- Others ...

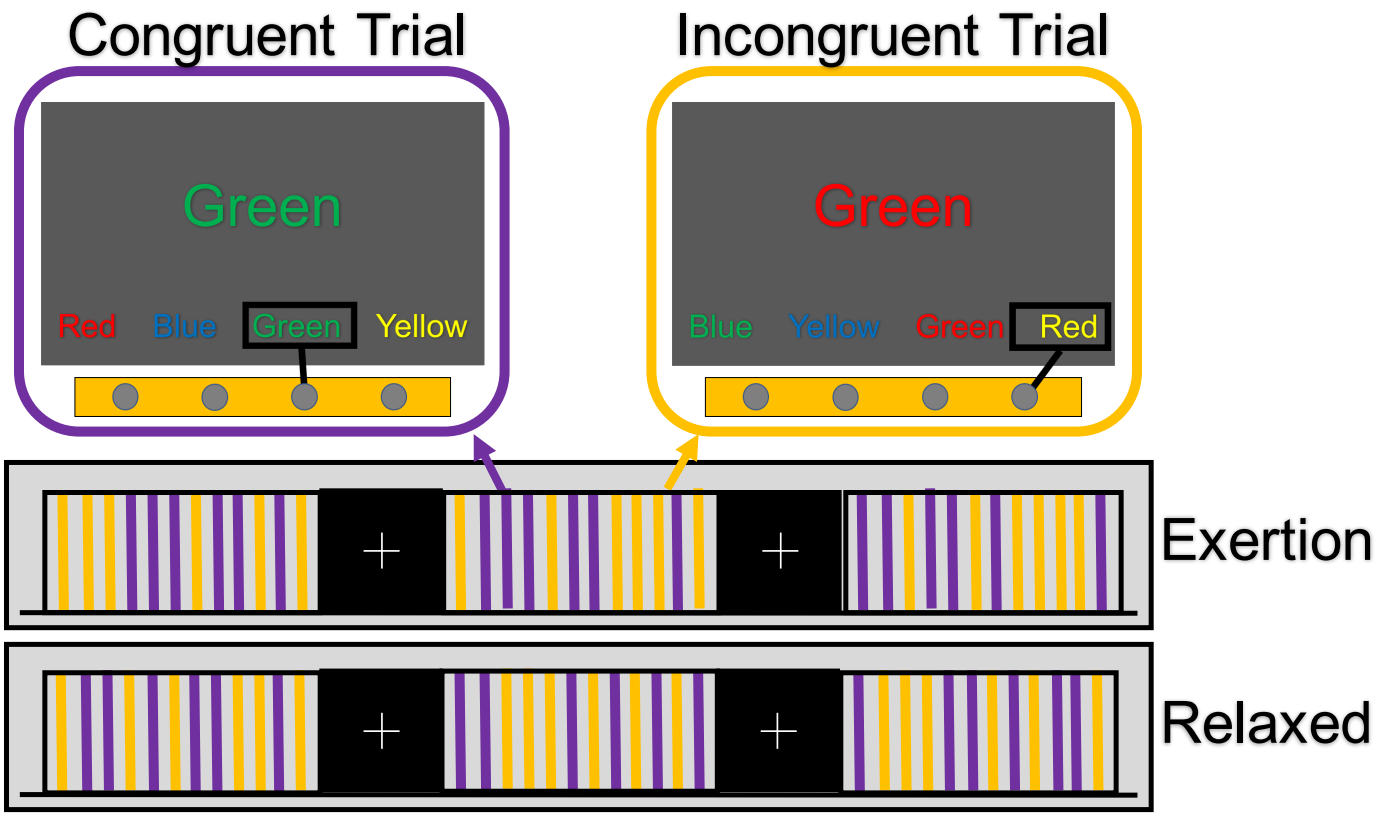
Functional Connectivity - Objectives

- We will integrate data from multiple imaging modalities and experiments in a manner that improves network estimation accuracy and reproducibility
- In the first part, I will describe an approach for joint estimation of multiple group level functional networks corresponding to different experimental conditions experienced by each individual in a group
- In the second part of the talk I will describe an approach for fusing fMRI and DTI data, with a goal to guide the estimation of functional connectivity using structural connectivity information

Bayesian Joint Network Learning

A Bayesian approach for joint learning of multiple related networks corresponding to different experimental conditions, longitudinal visits and so on

Motivating Stroop Task Application



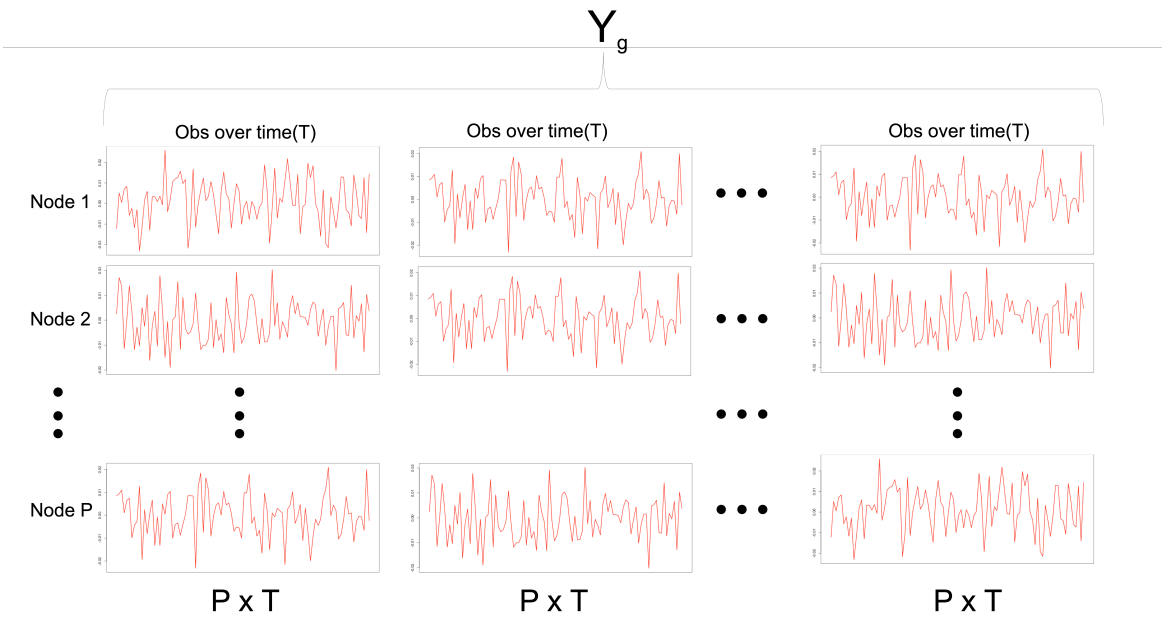
Stroop Task - Research Questions

- How does the brain network differ between (a) exertion vs relaxed task performance; and (b) task performance and passive fixation (resting state)?
- Data available from same subject on multiple experimental conditions
- Hence, the networks for fixation, EXR and RLX tasks share some similarities, and also differences
- Important to develop an approach that can borrow information across experimental conditions in order to compute these networks

Data Format

Each subject's prewhitened $P \times T$ data are concatenated along the time dimension for each group:

- P: number of nodes
- T: number of observations per subject
- N_g: number of subjects in group g



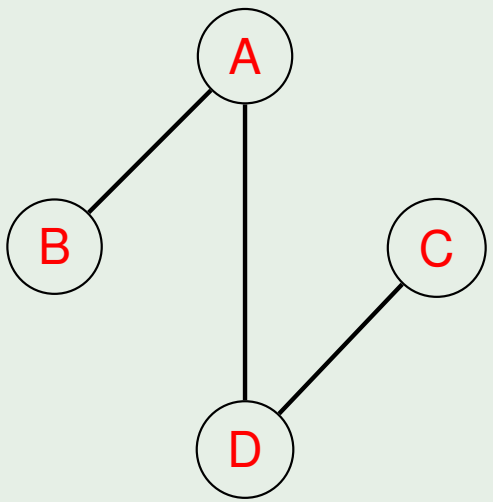
Total data matrix size for each group: $P \times TN_g$

Network Estimation - Gaussian Graphical Model

- Model the observed data across the p nodes for group g at time t as $\mathbf{Y}_{g,t} \sim N_p(\mathbf{0}, \mathbf{\Omega}_g^{-1})$ for $t = 1, \ldots, TN_g$.
- Estimate a sparse precision matrix - elements that are 0 correspond to absent edges, elements that are non-zero correspond to edges. Larger $|\omega_{kl}|$ are stronger connections.

Graph Example

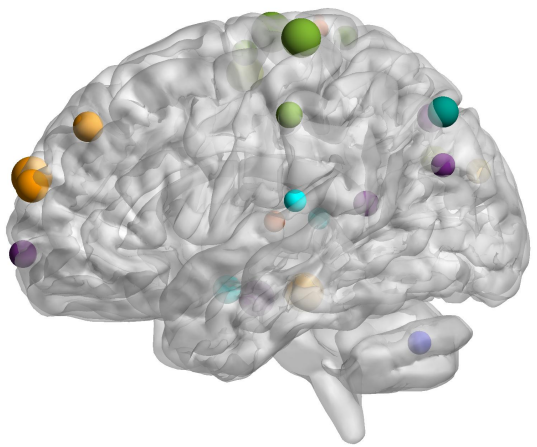
$$\Omega = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 0.2 & 0 & -0.1 \\ 0.2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ -0.1 & 0 & 0.5 & 1 \end{bmatrix} \end{array} \longrightarrow$$



Graphical Models

View the brain as a graph - set of regions (nodes) and the connections between them (edges). Edges represent functional connections.

- Use anatomically defined ROIs as nodes & average across voxels within each ROI.
- We use the Power atlas having 264 regions (Glasser et al., 2016).
- Normality assumption reasonable for modeling fMRI data - GGMs
- Edges can be determined based on the patterns of zeros in sparse precision matrix



Penalized Network Estimation

- Key is getting sparse estimates of the precision matrix - one possibility is penalized approaches
- Graphical lasso (GL) approach is popular (Friedman et al., 2008).

$$\boldsymbol{\Omega} = \arg\min_{\boldsymbol{\Omega}} \left\{ tr(\mathbf{S}\boldsymbol{\Omega}) - \log\det(\boldsymbol{\Omega}) + \lambda \sum_{k \neq l} |\omega_{kl}| \right\},$$

where $\mathbf{S} :=$ sample covariance, λ is a penalty term (larger λ implies more sparse networks), and ω_{kl} is the (k, l)th element of $\boldsymbol{\Omega}$ (FC strength).

- Not equipped to jointly compute multiple networks

Joint Estimation Approaches

Penalized Approaches Extend idea of graphical lasso by penalizing the *differences* between the two estimated graphs (Guo et al., 2011; Danaher et al., 2014). Example of fused lasso penalty from JGL

$$P(\mathbf{\Omega}) = \lambda_1 \sum_{k \neq l} \sum_{g=1}^{G} |\omega_{g,kl}| + \underbrace{\lambda_2 \sum_{k<l} \sum_{g \neq g'} |\omega_{g,kl} - \omega_{g',kl}|}_{\text{Edge differences across groups}}$$

Bayesian Approaches Typically operates by jointly learning probabilities for each edge across multiple networks. Limited development and not scalable to high dimensions (Peterson et. al, 2015). Some recent work on jointly estimating multiple temporally dependent brain networks (Qiu et al., 2016; Lin et al., 2017) not applicable here

Joint Estimation - Motivation & Goals

- Similar edge structure does not mean similar edge strengths - smoothing over edge strength in JGL may fail to capture this.
- Penalized approaches, although often scalable, only provide point estimates - can have a direct adverse influence when comparing networks
- Bayesian approaches characterize uncertainty but often not scalable
- Our goal: Develop scalable Bayesian model to pool information across groups to estimate the edges without forcing similarity in the edge strengths.
- Use edge probabilities to compare networks
- We introduce a flexible model for edge probabilities that pool information across experiments while also allowing network specific differences

Bayesian Joint Network Learning (BJNL) - Overview

- Assume G experimental conditions, with $T_1, \ldots, T_G$ scans respectively
- The pre-whitened fMRI measurements for g-th experimental condition are modeled as $\mathbf{Y}_t(g) \sim N_p(\mathbf{0}, \mathbf{\Omega}_g^{-1})$, $t = 1, \ldots, NT_g, g = 1, \ldots, G$, where

$$\pi(\mathbf{\Omega}_g) = C_g^{-1} \prod_{k=1}^{p} E(\omega_{g,kk}; \frac{\alpha}{2}) \Big\{ \prod_{k<l} w_{g,kl} N(\omega_{g,kl}; 0, \tau_{g,kl}^{-1}) + (1 - w_{g,kl}) DE(\omega_{g,kl}; \lambda_0) \Big\}$$

- Here $\mathbf{\Omega}_g \in M_p^+$, space of all $p \times p$ symmetric pd matrices
- $\pi(\mathbf{\Omega}_g)$ looks like the Bayesian GL prior but has a spike & slab flavor to it
- We denote $\pi(\mathbf{\Omega}_g)$ as the *spike and slab graphical lasso* prior

Group-Specific Edge Strengths

- The spike and slab graphical lasso prior induces sparsity in Ω
- For edge k, l in group g this prior is akin to

$$\pi(\omega_{g,kl}|w_{g,kl}, \tau^{-1}_{g,kl}, \lambda_0) \propto w_{g,kl} \underbrace{N(\omega_{g,kl}; 0, \tau^{-1}_{g,kl})}_{slab} + (1 - w_{g,kl}) \underbrace{DE(\omega_{g,kl}; \lambda_0)}_{spike}$$

- The edge weights $w_{g,kl}$ are the probability of a connection between nodes k and l in group g
- Larger λ_0 leads to sharper spike
- We pool information across groups to model the edge probabilities

Pooling Information across networks

- We use $h(w_{g,kl}) = \beta_{0,kl} + \sum_{g'=1}^{G} \beta_{g,kl} I_{\{g=g'\}}$, $h(\cdot) :=$ link function
- Uses information across all groups to estimate the shared effect $\beta_{0,kl}$.
- Non-zero differential effects $\beta_{g,kl}$ define edge differences across networks
- How do we model the shared and differential effects?
- p nodes imply $\frac{p(p-1)}{2} X(G+1)$ coefficients - curse of dimensionality
- Cluster the edge probabilities using a Dirichlet Process Mixture prior:

$$w_{g,kl} = h(\beta_{0,kl}, \beta_{g,kl}),\ \beta_{0,kl} \sim f_0,\ \beta_{g,kl} \sim f_g$$

$$f_0 \sim DP(MP_0),\ f_g \sim DP(MP_0),\ P_0 \equiv N(0, \sigma_\eta^2) \text{ is base measure}$$

Modeling of edge weights: comments

- The model is purposely overcomplete, i.e. G + 1 parameters in the weights model when G parameters would suffice.
- Allows us to pool information in a systematic manner, and ensures computational efficiency and interpretability Does not pose any problems w.r.t. identifiability of functionals of interest
- Motivated by existing literature on modeling binary or ordered categorical responses using mixture distributions (Kottas et al., 2005; Jara et al., 2007; Gill and Casella, 2009; Canale and Dunson, 2011).
- DP on the shared and differential effects reduces the sensitivity to the link function and enables straightforward posterior computation.

Posterior Computation

- Posterior computation is carried out via Markov Chain Monte Carlo.
- All updates are Gibbs
- Standard tools such as slice sampler are used for updating DP related parameters
- The precision matrix is updated one column at a time using closed form posteriors

Simulation

Goal: compare performance to other sparse precision estimation techniques.

- Generate data from two groups.
- Number of nodes (p): 40, 100.
- 3 network types : Erdos-Renyi (random graphs), Watts-Strogatz (small-world networks), scale-free networks.
- 3 levels of similarity across graphs (0.25, 0.50, 0.75).

60 subjects, each with 200 time points (typical fMRI study size), 50 replicates per combination of settings.

Simulation Setup - Methods Comparison

Joint Estimation Techniques

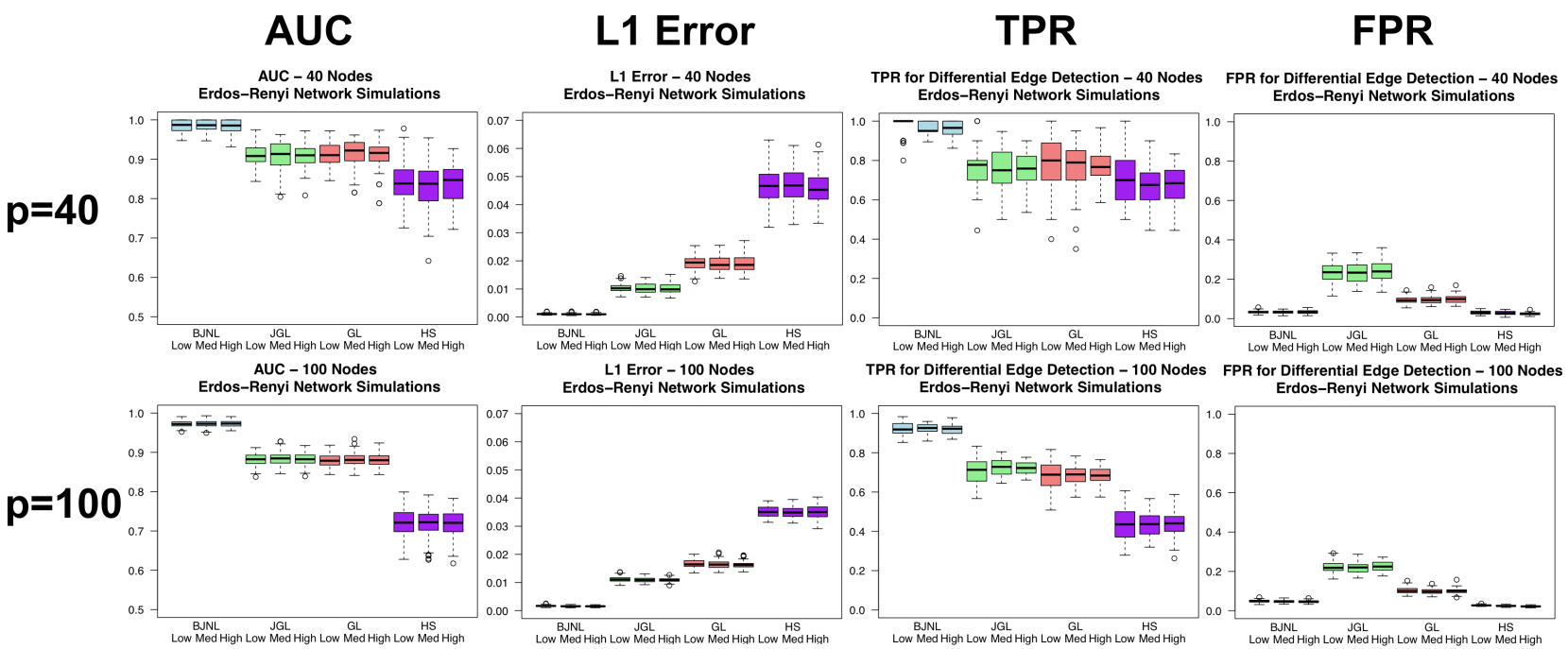
- Bayesian Joint Network Learning (BJNL) - pool information across groups to estimate the edge probabilities
- Joint Graphical Lasso (JGL) - smooth over edge strengths to pool information (Danaher et al., 2014). Penalty function: $P(\mathbf{\Omega}) = \lambda_1 \sum_{j \neq k} \sum_{g=1}^{G} |\omega_{g,jk}| + \lambda_2 \sum_{j \neq k} \sum_{g \neq g'} |\omega_{g,jk} - \omega_{g',jk}|$

Separate Estimation Techniques

- Graphical Lasso (GL) - Estimate each group separately, L1 penalty on the off diagonal elements (Friedman et al., 2008). Penalty function: $P(\mathbf{\Omega}_g) = \lambda_1 \sum_{j \neq k} |\omega_{g,jk}|$
- Horseshoe graph estimator (HS) - Bayesian approach. Shrinkage of each off-diagonal element (Carvalho et al., 2010; Li et al., 2019).

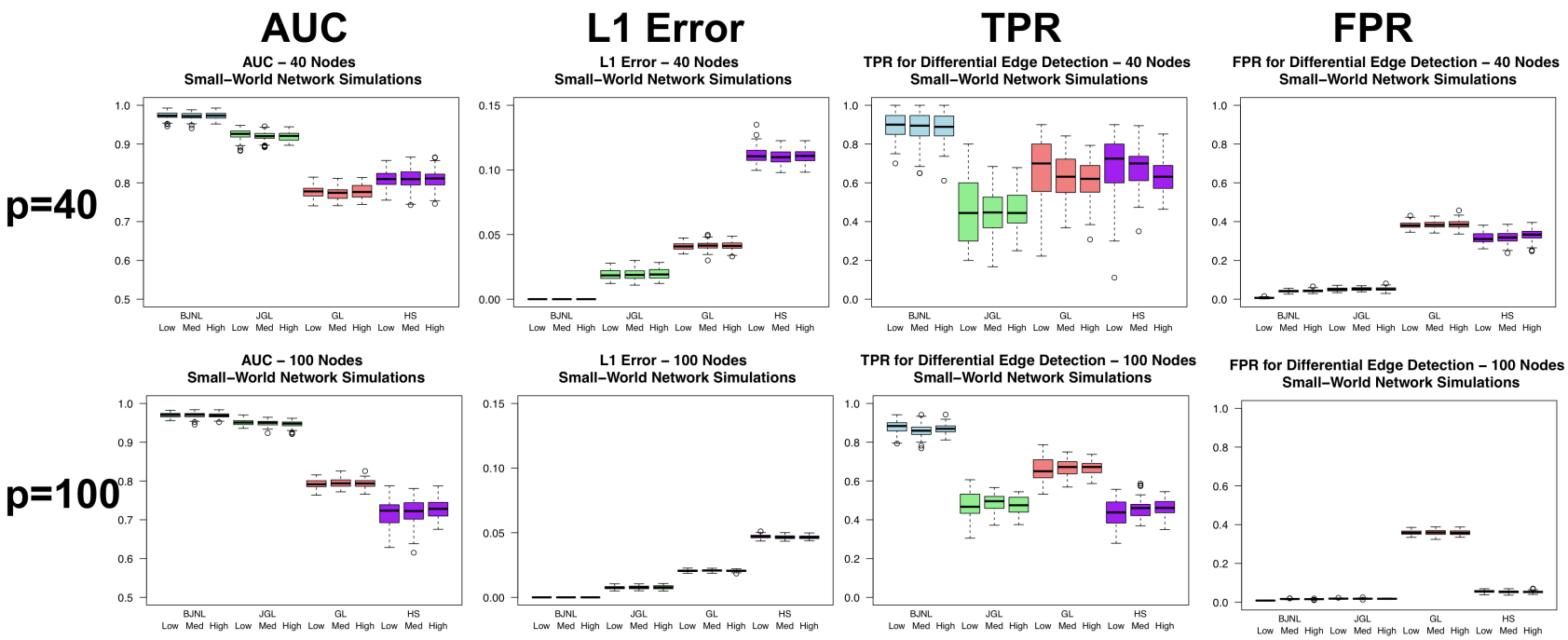
JGL has two tuning parameters, GL has one. We find the optimal values for them using a grid search over 30 values. Computationally demanding

Simulation Results - Erdos-Renyi Random Graphs



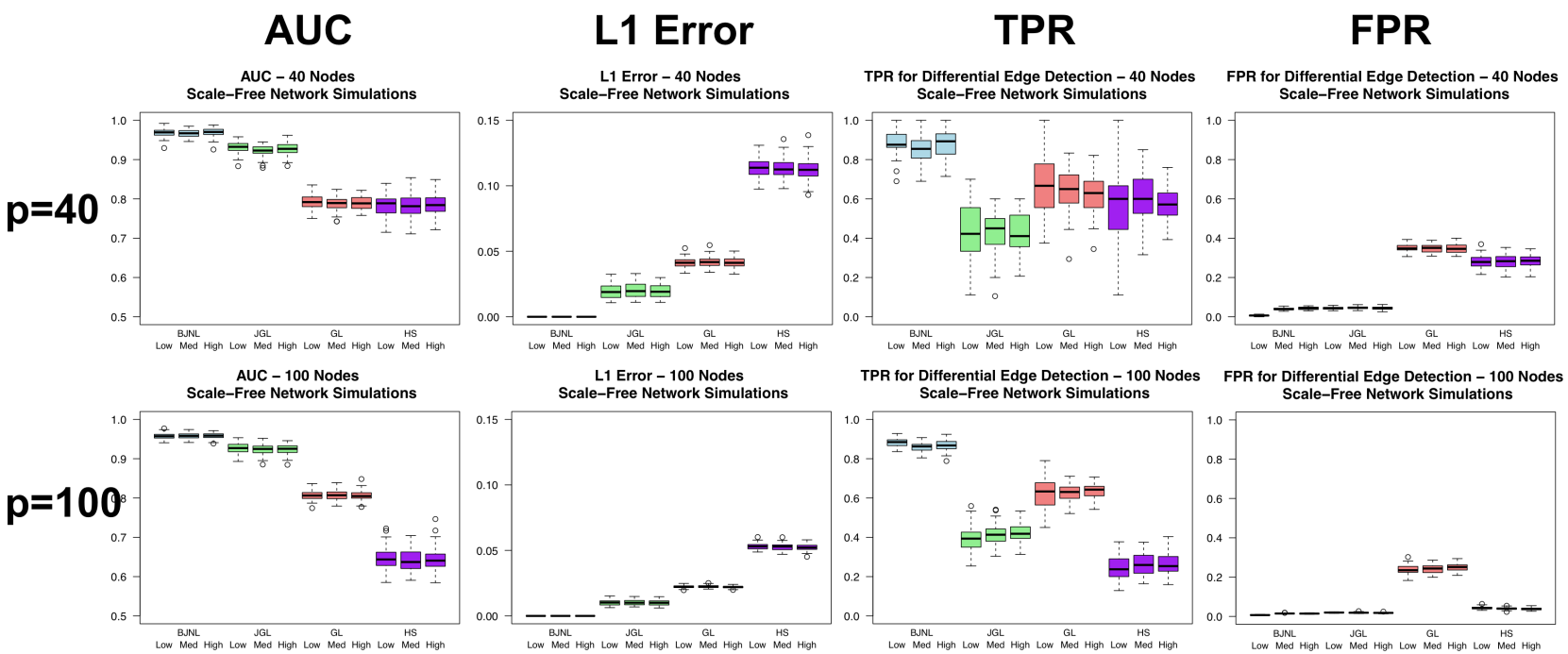
Boxes, left to right: BJNL (blue), JGL (green), GL (red), Horseshoe (HS)

Simulation Results - Small-World Random Graphs



Boxes, left to right: BJNL (blue), JGL (green), GL (red), Horseshoe (HS)

Simulation Results - Scale-Free Random Graphs



Boxes, left to right: BJNL (blue), JGL (green), GL (red), Horseshoe (HS)

Simulation Summary

- BJNL outperforms the other methods in both edge detection, differential edge detection, and L1 error.
- JGL is too conservative - misses many of the differential edges. We believe this is due to smoothing over the edge strengths, which makes it more difficult to identify edges with differential edges.

Stroop Task - fMRI Data

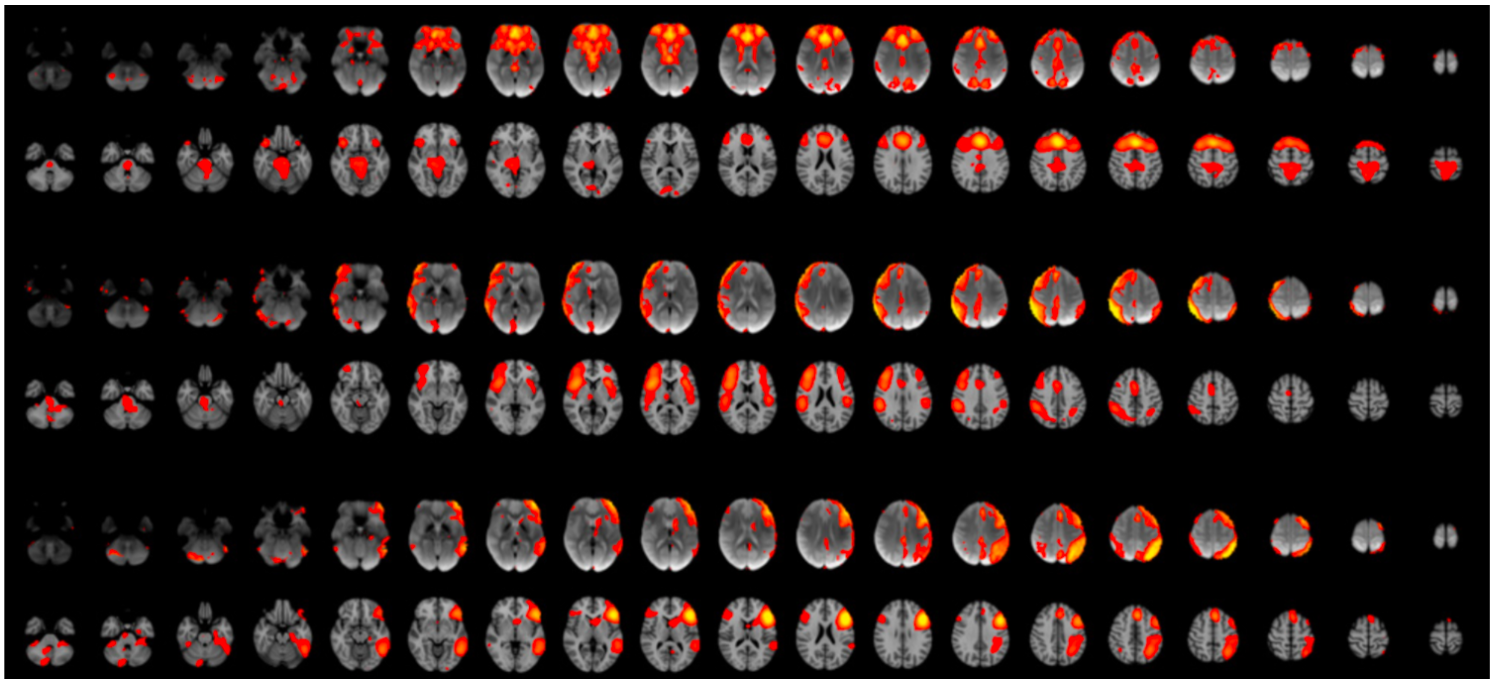
- 45 Healthy, right handed subjects
- Average age 21.9
- Stroop task and resting-state data collected. Groups are:
  - Task performance and resting state
  - Exertion and Relaxed task performance

  where the data are collected from the same subjects. Our interest is in comparing across conditions.
- Data in AAL 90 node system (Tzourio-Mazoyer et al., 2002).

Stroop Task - Edge Differences

Edges with significantly different strengths, organized by brain functional module.

Important Brain Networks



[4]

EC - cognition, action inhibition, emotion, pain

FPL - perception, cognition/language

FPR - cognition/language

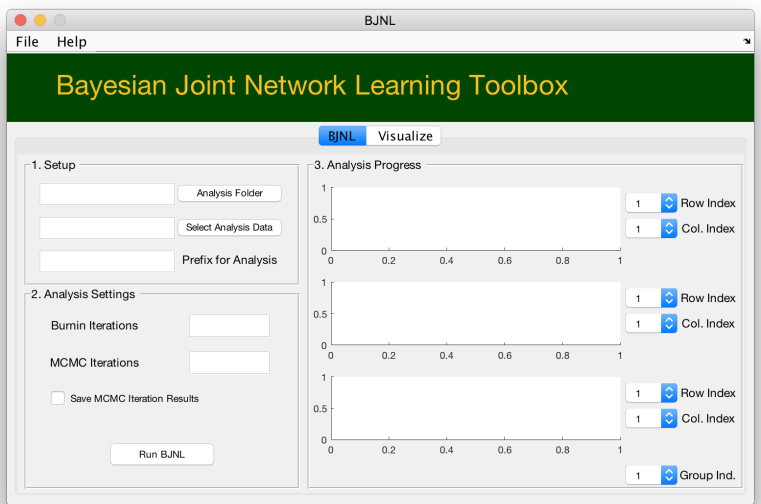
[5] Image from Smith et al. (2009)

Analysis Results under JGL and GL

- In contrast for **EXR vs RLX**, permutation p-values did not yield any significant edges for the GL, and only 2 significant edges for JGL.
- For **TASK vs REST**, JGL identified no differential edges and 624 common edges (versus 763 differential edges and 1211 common edges under BJNL).
- GL was able to identify 136 edges with differential strengths (27 of which overlap with those identified by the BJNL), and 661 common edges.
- Absence of differential edges between **EXR vs RLX** under GL, and between **TASK vs REST** under JGL, seems unrealistic

Summary

- Bayesian approach to joint estimation of multiple brain networks.
- Pools information to model edge probabilities instead of edge strengths.
- Demonstrated through simulations that BJNL works better for two-group simulations when the edge strengths are not constrained to be the same.
- Applied BJNL to stroop task data.
- Available as a Matlab toolbox and as a Julia package.



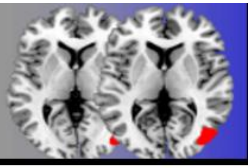
Anatomically Guided Brain Functional Connectivity

Using brain structural connectivity to adaptively guide the estimation of functional connectivity

FC-SC Relationship

FC and SC

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH
CBIS Center for Biomedical Imaging Statistics



Complex association between anatomical and functional connectivity

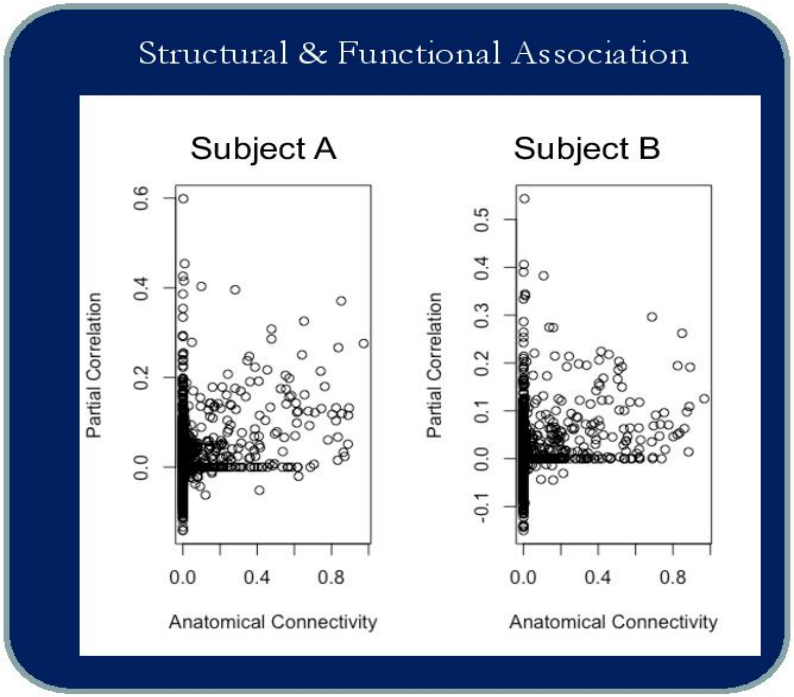
- Honey et. Al (2007,2009)

strong Structural $\Rightarrow$ strong Functional

$\not\Leftarrow$

- Messè (2013)

~15% of variation in FC explained by SC



Motivation

- Strong evidence for the role of white matter fiber tracts in regulating FC
- Appealing to incorporate information about brain structural connectivity when estimating functional connectivity
- Clearly desirable since it is expected to produce more accurate and reproducible estimates of the network,
- However, several considerations need to be taken into account, such as the complexity of the structure-function relationship, heterogeneity in FC for a given SC, strength since FC is only partially dependent on SC

Existing Challenges

- Limited advances in FC approaches which are guided by underlying anatomical knowledge.
- The majority of approaches have considered group level analysis that ignores substantial variability in cortical anatomy and function
- Only handful of recent approaches for estimating FC guided by SC for single subject data (Ng et al. (2012) and Pineda et al. (2014)).
- Unfortunately these methods may not adequately capture the complex underlying structure-function relationships and does not account for heterogeneity in FC for a given SC strength

Possible Solutions

- One could potentially use the earlier spike and slab graphical lasso framework to incorporate supplemental SC knowledge when estimating FC
- However the spike and slab lasso is best applied to group level networks
- We would like to design a shrinkage based approach that specifies a flexible FC-SC relationship and that works well for single subject data
- Single subject analysis important to avoid difficulties arising from the variability in brain anatomy over different ages, disease groups and so on

Proposed Model

Method: *si-GGM*

Structurally-Informed Gaussian Graphical Model (si-GGM)
(Higgins, Kundu and Guo, *NeuroImage*, 2018)

$$\mathbf{y}_t \sim N_p(0, \mathbf{\Omega}^{-1}).$$

$$\pi(\mathbf{\Omega}|\boldsymbol{\lambda}) = C_{\lambda,\nu}^{-1} \prod_{k=1}^{p} \mathrm{Exp}\left(\omega_{kk}; \frac{\nu}{2}\right) \prod_{j<k} \mathrm{DE}(\omega_{jk}; \nu\lambda_{jk}) \mathrm{I}\left(\mathbf{\Omega} \in M_p^+\right),$$

$$\pi(\boldsymbol{\lambda}|\boldsymbol{\mu}, \eta) = C_{\lambda,\nu} \prod_{j<k} \mathrm{LN}\left(\mu_{jk} - \eta p_{jk}, \sigma_\lambda^2\right),$$

"non-structural" variations in FC not explained by SC measure p_{jk}

Variations in FC related to SC p_{jk}

Parameter Interpretations

λ_{jk}: **structurally-informed** sparsity parameter for off-diagonal of the precision matrix

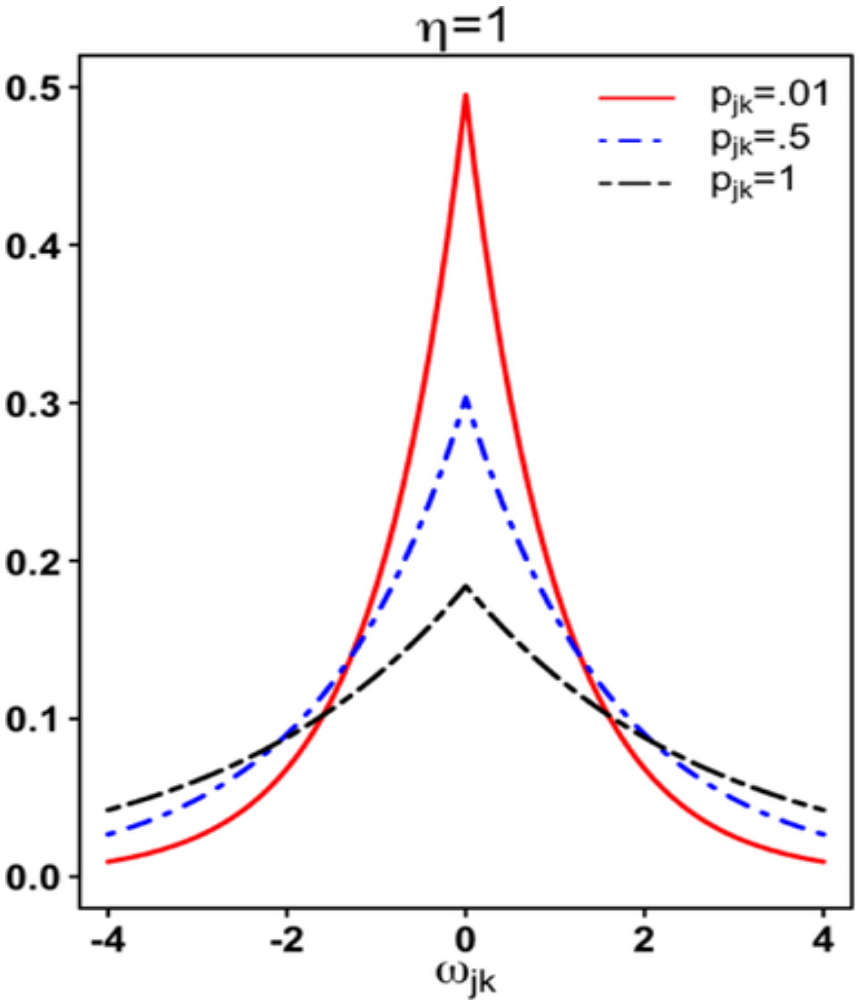
ν : overall network sparsity

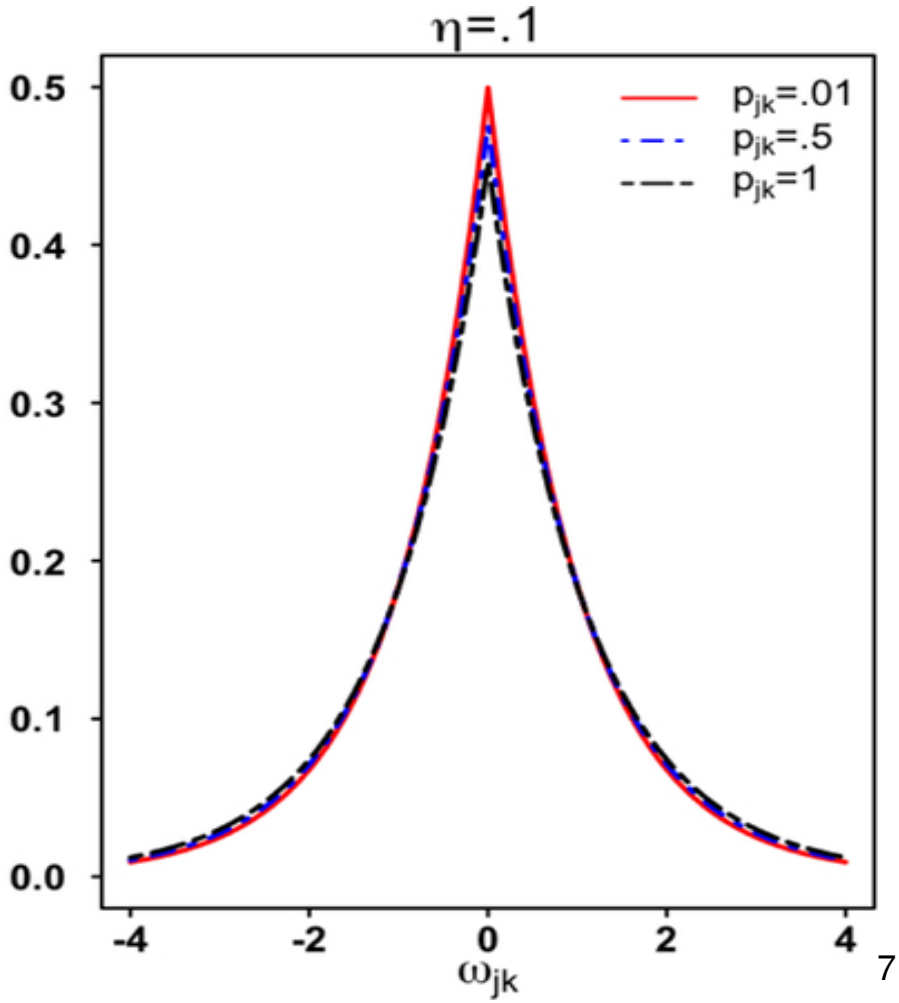
μ_{ij}: "non-structural" sources of variations regulating functional connection(FC)

η: effect of structural on functional

p_{jk} : probability of structural connection (SC)

Shrinkage Property

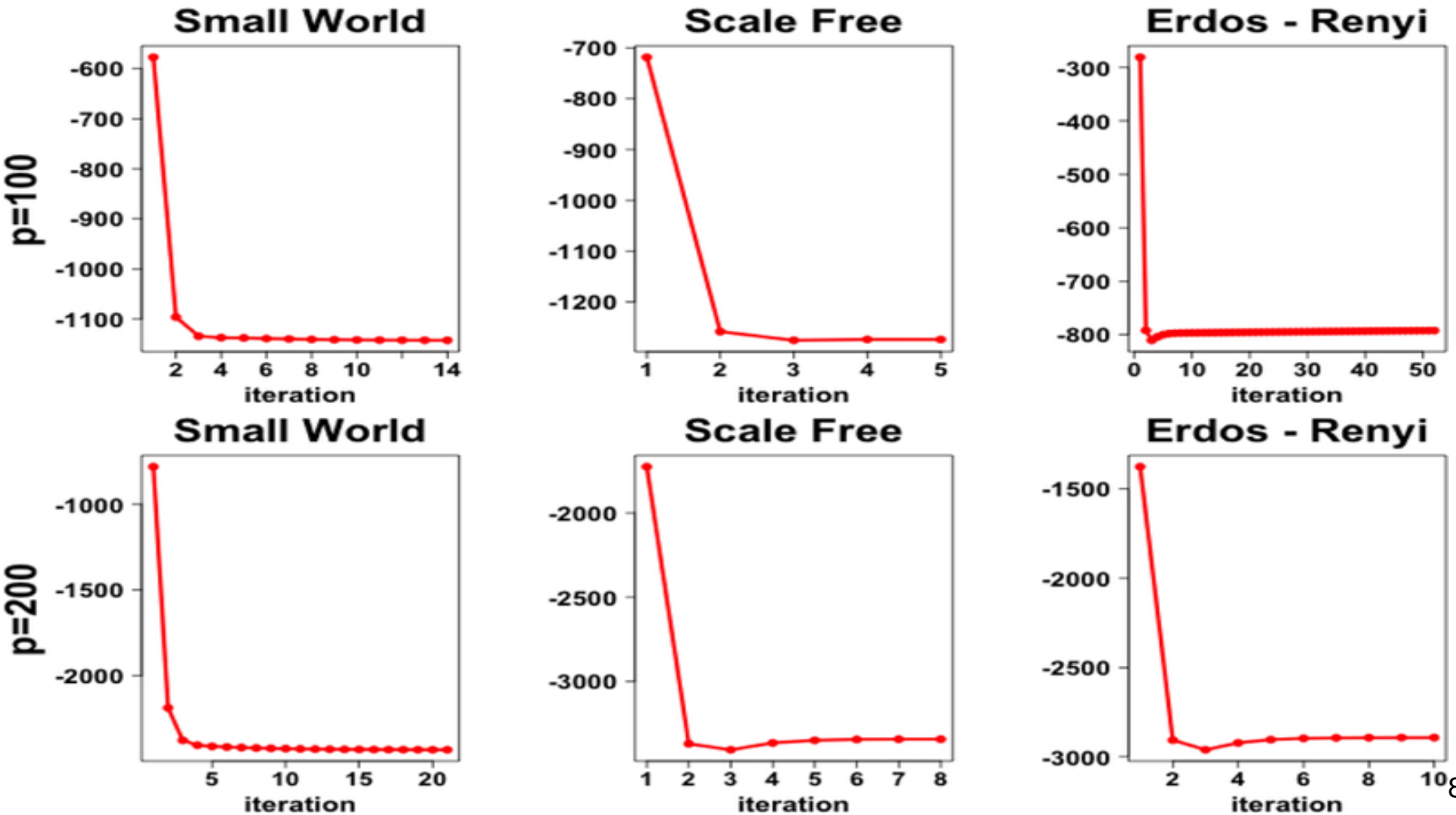




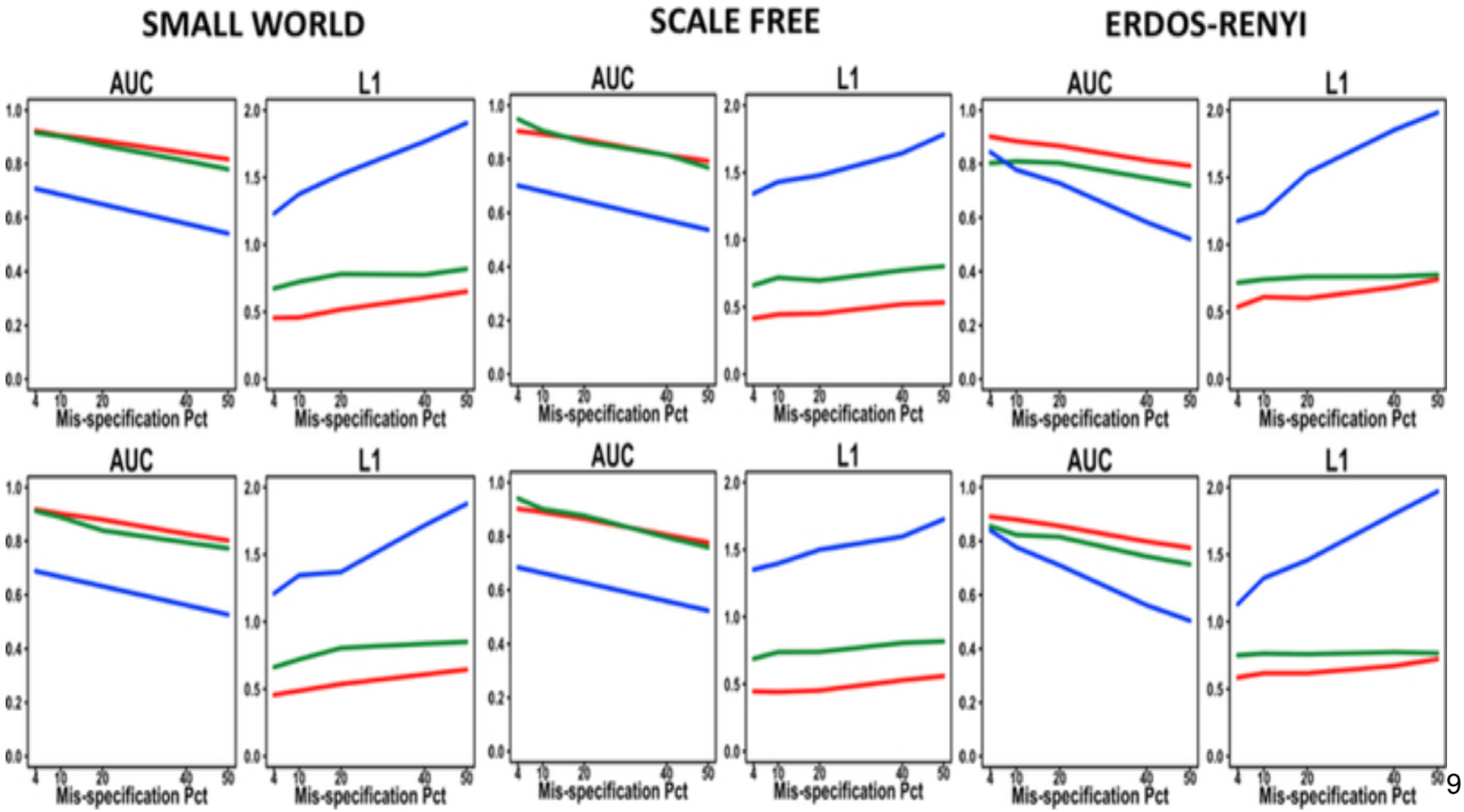
7

Computation

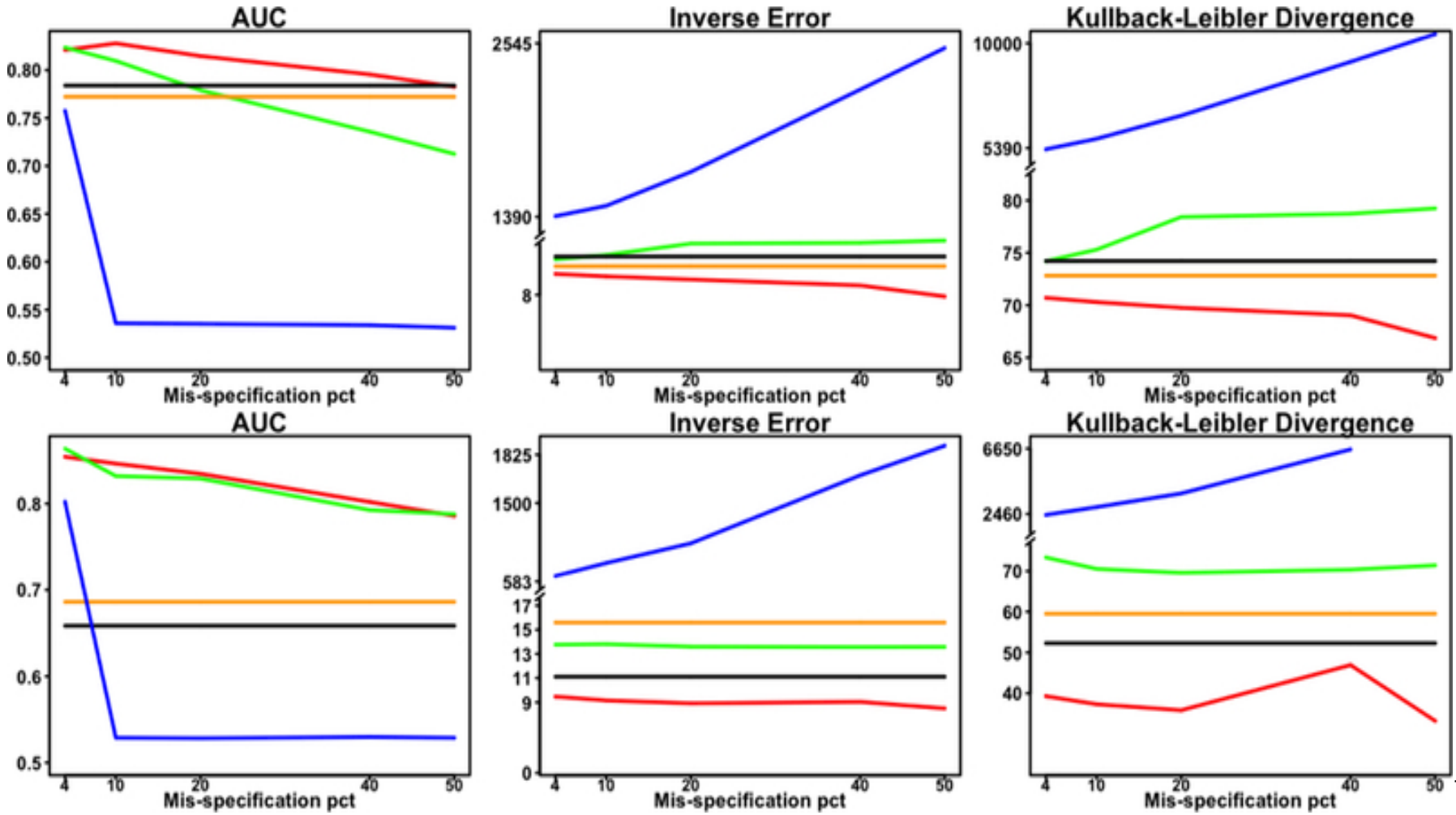
We use a coordinate descent algorithm to compute the MAP estimates.



Simulation Results: Data generated from GGM



Simulation Results: Data generated from ICA based models



10

Philadelphia Neurodevelopment Cohort

Study comprised fMRI & DTI data from children & adolescents aged between 8-21 years. All subjects are right-handed, physically, and mentally healthy

Network based on 120 brain volumes and Power atlas (264 ROIs)

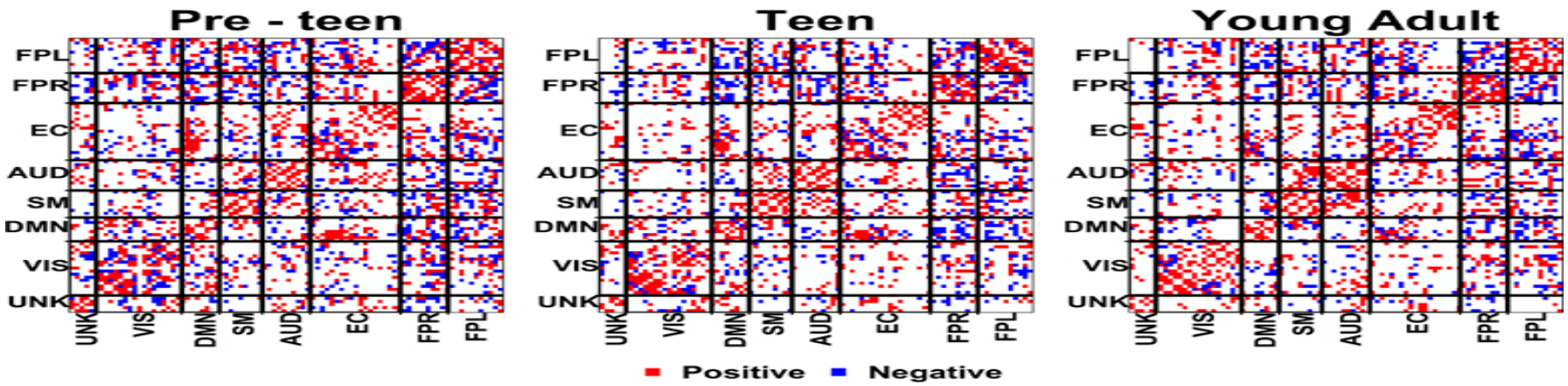
Goal is to assess gender based network differences

We also assess our methodâs ability to reliably estimate functional networks in terms of the intraclass correlation coefficient or ICC for seven network metrics

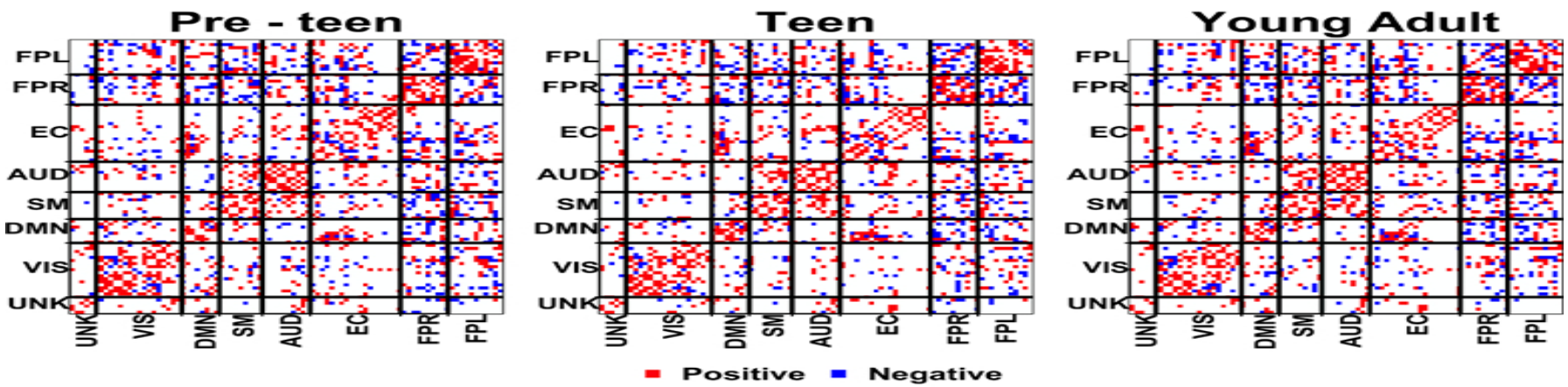
We split each subjects' resting state fMRI time series into two equally sized scanning sessions (60 scans each) to compute network reproducibility

Anatomical & cortical differences in the brain prohibit group level analysis

Gender Based Network Summaries



(A) Female

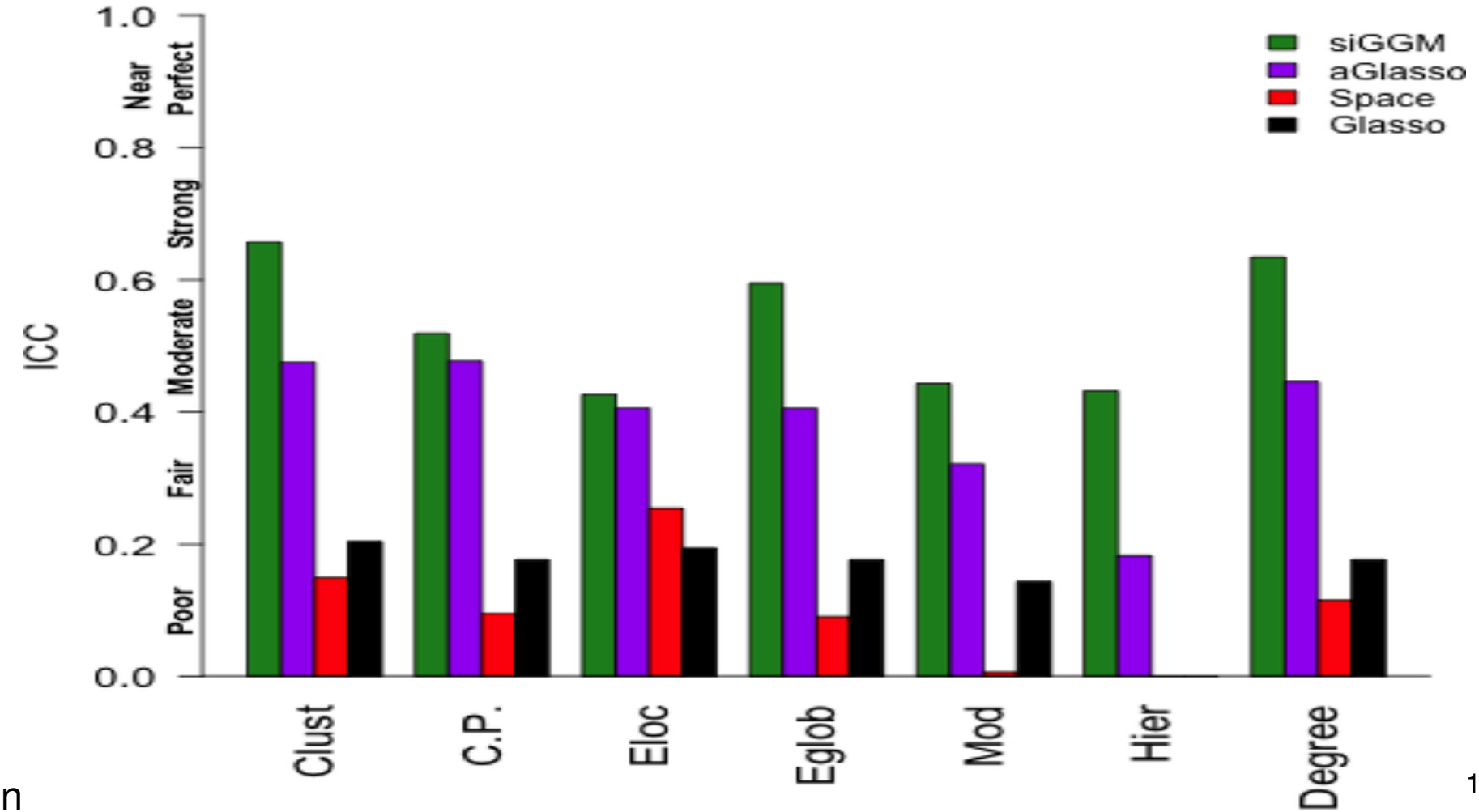


(B) Male

in

11

Network Reproducibility



in

12

Summary

- Incorporating SC information improved accuracy and reproducibility of the estimated network
- The proposed method is scalable to high dimensional networks
- Questions?

The Statistical Methods in Imaging Conference 2020

May 18-21, Atlanta, Georgia

Search...

Home | Program | Travel and Accommodation | Venue | Sponsors

SMI 2020

The Statistical Methods in Imaging (SMI) conference is the annual meeting of the American Statistical Association (ASA) Statistics in Imaging Section. For information about SMI 2020 at Emory University, please contact Dr. Ying Guo (yguo2 at emory dot edu).

References I

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15.

References II

Li, Y., Craig, B. A., and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, (just-accepted):1–23.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.