Randomized Approach to Differential Inference in Multi-Subject Functional Connectivity

Manjari Narayan* and Genevera I. Allen*†‡
*Dept. of Electrical and Computer Engineering and
†Dept. of Statistics, Rice University, Houston, TX
‡Jan and Dan Duncan Neurological Research Institute, Houston, TX

***Abstract*—Inferring *functional connectivity*, or statistical dependencies between activity in different regions of the brain, is of great interest in the study of neurocognitive conditions. For example, studies [1]–[3] indicate that patterns in connectivity might yield potential biomarkers for conditions such as Alzheimer's and autism. We model functional connectivity using Markov Networks, which use conditional dependence to determine when brain regions are directly connected. In this paper, we show that standard large-scale two-sample testing that compares graphs from distinct populations using subject level estimates of functional connectivity, fails to detect differences in functional connections. We propose a novel procedure to conduct two-sample inference via resampling and randomized edge selection to detect differential connections, with substantial improvement in statistical power and error control.**

***Keywords*-functional connectivity; graphical models; differential edges; undirected Markov Networks; multiple testing; resampling;**

I. Introduction

Functional connectivity, or statistical dependencies between activity in different regions of the brain, has been used for biomarker discovery [1]–[3] for conditions such as Alzheimer's and schizophrenia. Graphical models, which specify brain regions as nodes and connections as edges, have become increasingly popular tools to represent functional connectivity [4]–[6]. Given two populations of subject graphs, where subjects within a group have a relatively homogenous graph structure but vary in structure across groups, our objective is to specifically test for differences in edge structure.

While there are many methods to infer functional connectivity, Markov Networks and specifically Gaussian graphical models have been shown by Smith, et al. [4] to reliably identify direct relationships between brain regions. Markov Networks form edges based on conditional dependencies provided by partial correlations, as opposed to correlations. Current methods for structural inference over multi-subject graphs often employ two-sample tests on binarized and unthresholded correlation matrices. Shrinkage or thresholding is necessary in high dimensions, where correlation matrices become degenerate. However, a problem arises when subject graphs have been subjected to model selection including thresholding, namely the failure to account for the uncertainty in estimated subject graphs. While we account for model uncertainty in Markov Networks in this paper, the problems with standard inference as used in [1], [2] are common to all high dimensional networks, regardless of the graphical method employed. In recent work, Meinshausen, et al. [7], [8] perform model selection and one-sample inference using subsampling and randomized lasso procedures. In contrast, our goal is to perform two-sample inference on multi-subject graphs using randomized edge selection

To the best of our knowledge, this is the first paper to address the novel challenges of large-scale two-sample inference for multi-subject graphical models. Our contributions in this paper are twofold: (1) We demonstrate that standard large-scale two-sample inference fails in high dimensions due to increased bias and variability in multi-subject graphs (2) We propose a novel inferential procedure based on resampling and randomized variable selection that demonstrates major improvement in statistical power and error control over standard large-scale two-sample inference.

II. Problem Framework

An undirected graph or Markov network consists of a set of vertices $V = \{1, \dots, p\}$ and a set of edges $E \subset V \times V$. The vertices correspond to $X = (X_1, \dots X_p) \sim P$, where P is the probability distribution of X. We will assume that P is that of a Gaussian graphical model. Vertices V correspond to regions in the brain and edges E are determined by conditional dependencies between the brain regions. In fMRI studies, groups of similar voxels are aggregated based on anatomical information or prior imaging studies to create brain regions of interest. These regions are then employed as vertices or nodes in a functional connectivity study. We infer graphical models for each subject using whitened time series activity in p brain regions. While many graph estimation procedures are available, we use the *graphical lasso* [9] to estimate $\Theta = \Sigma^{-1}$, a $p \times p$ matrix of edge weights. Such procedures employ sparse penalties to find the zeros in Θ. The optimal sparsity level can be determined using model selection procedures such as StARS [10] and stability selection [8], among others.

Suppose $A^1 = \mathbb{I}\left(\Theta^1\right)$ and $A^2 = \mathbb{I}\left(\Theta^2\right)$, where $\mathbb{I}(\cdot)$ is the indicator function, represent the true adjacency matrices of subjects in population groups $g \in \{1, 2\}$. Our objective is to recover the set of true *differential* edges $\mathcal{D}$, i.e. the edges present exclusively in one group, given estimates of $\{\hat{A}^j\}_{j=1}^{n_g}$ for each subject j in group g, with n_g subjects per group.This is distinct from testing the edge strength or correlation value, which is beyond the scope of this paper. We accomplish this by testing the following hypothesis for each edge i.

978-0-7695-5061-9/13 $26.00 © 2013 IEEE
DOI 10.1109/PRNI.2013.29

$$\mathcal{H}_{0,i}: \quad A_i^1 = A_i^2, \quad \forall\, i = \left\{1, 2, \ldots, \frac{p(p-1)}{2}\right\}$$

Since we have a multiple testing problem, we use the *Benjamini Yekutieli procedure* [11] to control false discovery rate, or the expected proportion of false positives.

Challenges for standard two-sample inference

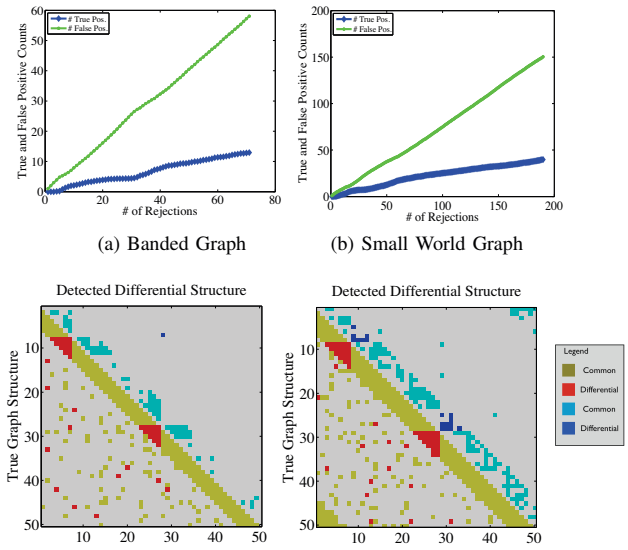
To test for differential edges, the standard approach is to first estimate the functional connectivity per subject j, $\{\hat{A}^j\}_{j=1}^{n_g}$ and then conduct standard large-scale two-sample testing on each edge. Since the estimated edges are binary, a two-sample score statistic $T_i = \frac{\hat{p}_{1,i} - \hat{p}_{2,i}}{\sqrt{\frac{\hat{p}_{1,i}(1-\hat{p}_{1,i})}{n_1} + \frac{\hat{p}_{2,i}(1-\hat{p}_{2,i})}{n_2}}}$ can be used to detect whether the edge proportions $p_{g,i} = \sum_j^{n_g} \hat{A}_i^j$ are higher in one group over the other. However, functional connectivity studies are unlike other common two-sample studies in genomics and neuroimaging. For example, in both differential gene expression studies for microarray data [12] and two-sample mixed-effects tests for voxel level analyses for fMRI data [13], inference is conducted directly on the raw data prior to any kind of model estimation. In contrast, functional connectivity studies usually perform model selection to determine the optimal number of edges in the graph in each subject, prior to conducting two-sample tests. This introduces challenges for maintaining statistical power with error control.

We illustrate this problem using a simulation study, explained later in section IV. Here, we first perform graph estimation and model selection for each subject and apply two-sample tests for each edge to detect differences. Fig 1 shows that such a method finds more false positives relative to true positives with increasing rejections of the null hypotheses. We also provide corresponding adjacency matrices of the graph populations in Fig 1 that indicate both common edge structures and differential edge structures that we wish to detect. Standard large-scale two-sample testing often rejects common edge structures as differential and fails to detect true differences, with false discovery proportions as large as 80%.

The poor performance of standard two-sample procedures is a consequence of testing for differences using model selected subject level graphs, $\hat{\Theta}^j$. When employing standard two-sample test statistics, we only account for within group variance, without accounting for subject variability due to uncertainty [14] in the model selection stage. Consequently, edges that are either false positives or false negatives at the subject level, easily translate to false differential edges at the population level.

III. Randomized Approach to Differential Edge Testing

To account for the uncertainty in subject level models, we simultaneously combine model selection with two-sample inference using resampling procedures. Bootstrapping procedures can estimate the uncertainty or variance in a statistic of interest via repeated sampling of the observations with replacement. Subsampling procedures are similar except they



(a) Banded Graph (b) Small World Graph

(c) Banded Graph, 78% FDP (d) Small World Graph, 83% FDP

Fig. 1. Receiver operating characteristics (a),(b) and edge recovery adjacency matrices (c),(d) of standard two-sample inference for detecting differential edges

sample the data without replacement. However, resampling observations alone does not account for systematic bias in model selection. We combine resampling and randomization to compute uncertainty in model selection. We then utilize the variability of edges across models to create a novel inferential procedure to detect differential edges.

A. Edge Selection Frequencies

The edge selection frequency (ESF) is the frequency with which an edge in a graphical model is selected after repeated subsampling or bootstraps and is popularly used in model selection procedures [8]. When an edge is truly present, it will often be selected by the model under random perturbations of the data. Therefore a high ESF is a strong indicator that an edge is truly present. We also introduce a randomization method to randomly sample from models in the vicinity of the selected model. Aggregating the ESF after additional randomization reduces the probability that null-edges will be selected. Moreover, if we treat edge selection as a Bernoulli random variable, the selection variance also provides a measure of edge disagreement or instability for an edge. For each bootstrap sample $k \subset \{1, \ldots, B\}$, we generate a $p \times p$ matrix of regularization parameters that randomly penalizes each edge with $\lambda^* \pm c$, for some model selected λ_j^* for each subject. Thus, we repeatedly estimate the graph using Λ^j generated as follows,

$$\Lambda_k^j = \lambda_j^* + cW$$

where $\Pr\{W_{m,n} = \pm 1\} = \frac{1}{2}$, and $c \in [.25, .5]$ is small fraction of $\lambda_{\max}$. We then aggregate resampled estimates of the edge selection $\Psi_k^j = \mathbb{1}\left\{\hat{\Theta}_k^j \neq 0\right\}$ to obtain the edge selection frequency. We subsequently threshold away any edges that fail to have at least one subject with $\frac{1}{B}\sum_k \Psi_{ik}^j > .5$.

B. Two Level Test Statistic

Given edge selection frequencies for every edge in every subject for two groups, our goal is to construct a two-sample test statistic that appropriately measures differences in the mean ESF of each group. We obtain the test statistic using an improved variance component estimate [13], one that accounts for model uncertainty within subjects, as well as between subject variability within each population.

Subject Level: Suppose each edge possesses a true ESF based on group membership that is unobserved. We approximate the true ESF with estimates based on randomized edge selection over a range of models in each subject. This provides us with a plug-in estimate of the ESF per subject given by μ^j as well as variance in edge selection given by $\mu^j(1-\mu^j)$. Subsequently, the group mean of the variance in edge selection provides an estimate of the uncertainty in model selection for that edge.

Population Level: The standard score statistic evaluates group edge proportions based on model selected subject graphs within each group. However, the within group subject variance in this statistic underestimates the true variance as it does not account for variability due to model selection. In our approach, we instead compare the group means of ESF based proportions in each subject $\hat{p}_g = \frac{1}{n_g}\sum_{j=1}^{n_g}\mu^j$. In addition to the population variance of the edge proportions within each group, the test statistic T_i in Eq. (1) includes sampling variance in the estimated proportions due to model uncertainty. Thus the test statistic for the difference in proportions, $\hat{p}_1 - \hat{p}_2$, includes both inter-subject variability as well as model uncertainty. However, this test statistic does not follow a known distribution. We therefore use a permutation procedure to infer the null distribution. The full procedure for obtaining the test-statistic T_i is described in the algorithm below.

IV. Simulations

We simulate multi-subject multivariate data according to a Gaussian graphical model with various graph structures. Suppose (A^1, A^2) represent the true adjacency structure in distinct populations. We construct differential edges A_{d_g} such that they are mutually exclusive, $A_{d_1} \cap A_{d_2} = 0$. Therefore $A^g + A_{d_g}$ forms the support for Θ^g. We simulate multivariate noise with covariance structure given by $\Sigma^1 = (\Theta^1)^{-1}, \Sigma^2 = (\Theta^2)^{-1}$ for each subject j. We simulate graphs with parameters $n = 400, p = 50, n_{g_1} = 20, n_{g_2} = 20$ as *small world*[1], *banded*[2], and *hub*[3] graphs.

To illustrate the source of false positives, we add differential edges to the above structures in two varieties. In the first case, we add differential edges that clump together like hubs to small world and banded structures. In the second case, we add differential edges, randomly distributed across nodes, to a common hub-like structure. Regardless of where hubs occur in graphs, edges on hub nodes are easily confounded with true edges. Our simulations enable us to study this obstacle to correctly identifying differential edges.

[1]where any node has a short path to any other node.

[2]where each node is only connected to a few other nodes.

[3]where some groups of nodes are highly connected to each other.

Randomized Approach To Differential Inference

1) **Resample observations and estimate graphs by randomizing edge selection**
2) **Compute test statistic for each edge i and for each subject j**

$$\mu_i^j = \frac{1}{B}\sum_{k=1}^{B}\Psi_k^j(i) \quad \text{and} \quad \sigma_i^j = \sqrt{\frac{1}{B}\left(\mu_i^j(1-\mu_i^j)\right)}$$

We obtain the following test statistic given that n_g are the number of subject samples in each population g.

$$T_i = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\sum_j^{n_1}(\hat{p}_1-\mu^j)^2 + n_1\bar{\sigma}_1^2}{n_1} + \frac{\sum_j^{n_2}(\hat{p}_2-\mu^j)^2 + n_2\bar{\sigma}_2^2}{n_2}}} \tag{1}$$

$$\hat{p}_g = \frac{1}{n_g}\sum_{j\in g}\mu_i^j \quad \text{and} \quad \bar{\sigma}_g^2 = \frac{1}{n_g}\sum_{j\in g}(\sigma_i^j)^2$$

3) **Permute subject labels Q times and compute test statistic $\tilde{T}^*$ to estimate null distribution. The p-values are given by**

$$\tilde{p}_i = \frac{1}{Qn_{\tilde{p}}}\sum_{l=1}^{n_{\tilde{p}}}\sum_{q=1}^{Q}\mathbb{I}\left(|\tilde{T}_l^{*q}| \geq |T_i|\right)$$

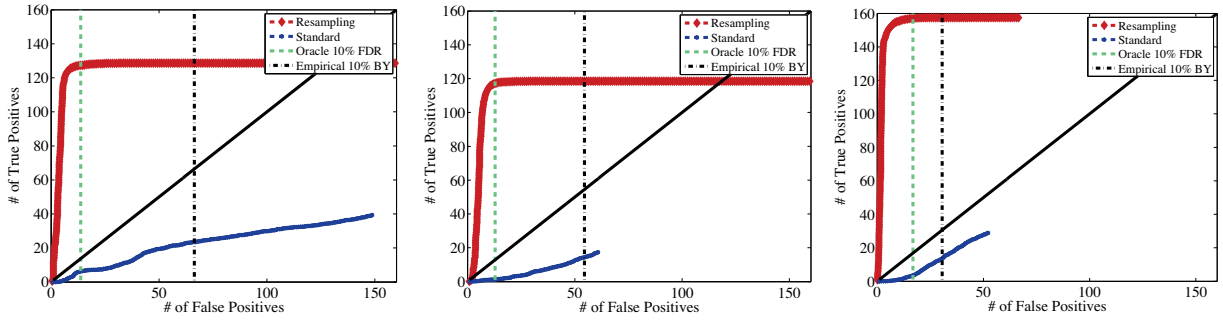
4) **Benjamini Yekutieli correction for multiple testing under dependence**

A. ROC Comparisons

We compare the operating characteristics of our resampling procedure and standard large-scale two-sample inference in Fig. 2. The figures show the true positive count against the false positive counts averaged over 50 random trials. For each graph type, i.e. hub, banded, and small world our procedure identifies differential edges with true positive rates (TPR) ranging from 92-97.8% and false discovery rates (FDR) from 10.3-21.5%. In contrast, the standard approach has almost 4 times lower statistical power with TPR between 18.5-26% and FDR from 64.6-78% for various graph types.

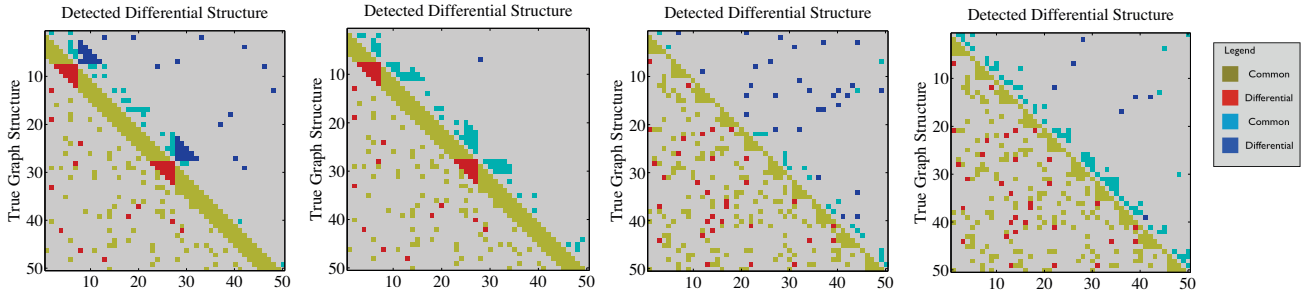
B. Location of False Positives

In Fig 3, we illustrate the benefits of our resampling procedure for the banded graph and hub graph via their adjacency matrices. The lower triangle consists of both true common edges and differential edges, while the upper triangle only consists of estimated differences. When the differential edges belong to hub nodes, our approach finds a few false positives on the same nodes as the true differentials. In contrast, the standard approach rejects many edges that share nodes with the differential edges. Our approach performs much better when the differential edges occur more randomly. Again, the standard approach often misidentifies common edges on hub nodes as differential. This illustrates that edge uncertainty from



(a) Small World Graph (b) Banded Graph (c) Hub Graph

Fig. 2. Operating Characteristics of two-sample Inference using the Resampling with Randomization Approach. Resampling with randomization (blue) outperforms standard two-sample inference. $p = 50, n = 400, n_{g_1} = 20, n_{g_2} = 20$



(a) Resampling, (0.97, 0.22) (b) Standard, (.15, .78) (c) Resampling, (0.88, 0.14) (d) Standard, (0.19, 0.69)

Fig. 3. Locations of Detected Differential Edges in Banded (a)-(b) and Hub (c)-(d) Graphs with true positive rate and false discovery proportion metrics (tpr, fdp). The lower triangular contains true graph structure including true differential and common structures, while the upper triangular displays estimated true and false positive differential edges.

randomized edge selection can eliminate edges that confound differential inference.

V. Discussion

Our paper provides insights into the challenges of edge inference over multi-subject graphical models when using any model selection procedure. We have proposed a novel inferential procedure for neuroimaging studies where the objective is to identify functional connections that are distinct cognitive markers. Incorporating model uncertainty in multi-subject models into a test-statistic that differentiates graph structure at the population level can offer dramatic improvements in power and error control. In future work, we expect to demonstrate results on heterogenous subjects in real and simulated data, extend this procedure beyond unweighted edges as well as address more complex differences in functional connectivity beyond edge inference.

Acknowledgment

This work was supported by the NSF grant DMS-1209017 and in part by the Data Analysis and Visualization Cyberinfrastructure funded by NSF grant OCI-0959097.

References

[1] T. B. Jones, P. A. Bandettini, L. Kenworthy, L. K. Case, S. C. Milleville, A. Martin, and R. M. Birn, "Sources of group differences in functional connectivity: An investigation applied to autism spectrum disorder," *NeuroImage*, vol. 49, no. 1, pp. 401 – 414, 2010.

[2] M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore, "Functional connectivity and brain networks in schizophrenia," *J. Neurosci.*, vol. 30, no. 28, pp. 9477–9487, 2010.

[3] E. Bullmore, "Functional network endophenotypes of psychotic disorders." *Biol. Psychiatry*, vol. 71, no. 10, p. 844, 2012.

[4] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fmri," *NeuroImage*, vol. 54, no. 2, pp. 875–891, 1 2011.

[5] S. Ryali, T. Chen, K. Supekar, and V. Menon, "Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty," *NeuroImage*, vol. 59, no. 4, pp. 3852–3861, 2 2012.

[6] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, and E. Reiman, "Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation," *NeuroImage*, vol. 50, no. 3, pp. 935–949, 2010.

[7] N. Meinshausen, L. Meier, and P. Bühlmann, "P-values for high-dimensional regression," *J. Amer. Stat. Assoc.*, vol. 104, no. 488, pp. 1671–1681, 2009.

[8] N. Meinshausen and P. Buhlmann, "Stability selection," *J. Roy. Statist. Soc. Ser. B*, vol. 72, no. 4, pp. 417–473, 2010.

[9] J. Friedman, T. Hastie, and R. J. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, July 2008.

[10] H. Liu, K. Roeder, and L. Wasserman, "Stability approach to regularization selection (stars) for high dimensional graphical models," in *Advances in Neural Information Processing Systems*, 2010.

[11] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[12] S. Dudoit and M. Van Der Laan, *Multiple testing procedures with applications to genomics*. Springer, 2007.

[13] J. A. Mumford and T. Nichols, "Modeling and inference of multisubject fmri data," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 25, no. 2, pp. 42–51, 2006.

[14] S. Buckland, K. Burnham, and N. Augustin, "Model selection: An integral part of inference," *Biometrics*, vol. 53, no. 2, pp. 603–618, 1997.