Bayesian Joint Modeling of Multiple Brain Functional Networks

Joshua Lukemire, Suprateek Kundu, Giuseppe Pagnoni & Ying Guo

View supplementary material

Published online: 01 Sep 2020.

Submit your article to this journal

Article views: 355

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Taylor & Francis
Taylor & Francis Group

Check for updates

Bayesian Joint Modeling of Multiple Brain Functional Networks

Joshua Lukemire[a], Suprateek Kundu[a], Giuseppe Pagnoni[b], and Ying Guo[a]

[a]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; [b]Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

ABSTRACT

Investigating the similarity and changes in brain networks under different mental conditions has become increasingly important in neuroscience research. A standard separate estimation strategy fails to pool information across networks and hence has reduced estimation accuracy and power to detect between-network differences. Motivated by an fMRI Stroop task experiment that involves multiple related tasks, we develop an integrative Bayesian approach for jointly modeling multiple brain networks that provides a systematic inferential framework for network comparisons. The proposed approach explicitly models shared and differential patterns via flexible Dirichlet process-based priors on edge probabilities. Conditional on edges, the connection strengths are modeled via Bayesian spike-and-slab prior on the precision matrix off-diagonals. Numerical simulations illustrate that the proposed approach has increased power to detect true differential edges while providing adequate control on false positives and achieves greater network estimation accuracy compared to existing methods. The Stroop task data analysis reveals greater connectivity differences between task and fixation that are concentrated in brain regions previously identified as differentially activated in Stroop task, and more nuanced connectivity differences between exertion and relaxed task. In contrast, penalized modeling approaches involving computationally burdensome permutation tests reveal negligible network differences between conditions that seem biologically implausible. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

1. Introduction

The Stroop task (Stroop 1935) is one of the most reliable psychometric tests (MacLeod 1991) that is widely used as an index of attention and executive control. It has been extensively employed to investigate the neural underpinnings of mental effort that is defined as the deployment of mental resources in a demanding task that needs to be willfully maintained. Stroop task experiments have been successfully adopted in neuroimaging studies targeting child development (Levinson et al. 2018) and aging (Duchek et al. 2013), addictive behavior (Wang et al. 2018), psychiatric conditions (Woodward et al. 2016), and several other areas. Neuroimaging studies have shown differential activation in several brain regions related to the Stroop task (Gruber et al. 2002; Shan et al. 2018). However, there are limited advances for understanding the neural circuitry changes related to differences in the capacity to exert mental effort (Levinson et al. 2018). Existing connectivity studies have focused on independent component analysis or ICA (Wang et al. 2018), seed region based correlation analysis (Levinson et al. 2018), and pairwise correlation analysis (Peterson et al. 1999). While useful, ICA based studies do not provide edge-level interpretations necessary for graph theoretical insights, whereas seed region-based analysis are subjective and do not use whole brain information. Moreover, pairwise correlations fail to account for spurious effects of third party nodes (Smith

et al. 2011) that may lead to misleading connectivity findings. In addition, none of existing methods investigated connectivity differences related to varying mental effort in Stroop task, although recent evidence point to significant brain activation differences when the task is performed by voluntarily engaging a maximum or a minimum of mental effort (Khachouf et al. 2017).

In one of the first such efforts to our knowledge, we investigate how the brain network reorganizes under different cognitive conditions corresponding to passive fixation and task performance, as well as between effortful and relaxed task performance, under a Stroop task experiment. The scientific hypothesis based on previous studies is that considerable neurobiological and connectivity differences should be present between the different cognitive conditions. The investigation of brain network differences may be performed on a single subject or, as in our case, at a group level which is expected to average out subject-specific idiosyncrasies (Kim, Pan, and Alzheimer's Disease Neuroimaging Initiative 2015). Under a graph theoretical approach, edges featuring differential strengths correspond to brain connections that are more activated or suppressed during one experimental condition as compared to others. On the other hand, connections shared across networks may represent an *intrinsic* functional network architecture which is common across experimental conditions (Fox et al. 2007).

However, separate estimation of multiple networks may not have enough power to accurately detect shared and differential features between networks due to the inherent noise in fMRI data. Separate network estimation may also be inadequate in terms of comparing multiple networks (a central question of interest in our applications), due to a lack of systematic inferential tools to test significant connectivity differences between experimental conditions. The above factors could potentially result in a loss of biological interpretability, as illustrated via our Stroop task data analysis. These critical issues can be potentially resolved via a joint learning approach for multiple networks that pools information across experimental conditions to learn shared and differential features. Such an approach is motivated by the success of recent data fusion methods for multiple datasets in literature (Lahat, Adah, and Jutten 2015).

There has been a limited development of approaches for the joint estimation of multiple networks. Penalized approaches for the joint estimation of multiple Gaussian graphical models (GGMs) (Guo et al. 2011; Danaher, Wang, and Witten 2014; Zhu, Shen, and Pan 2014) typically smooth over the strength of connections across networks to enforce shared edges, which is a useful modeling assumption but may not be supported in practical brain network applications. Further, they often require a careful choice of more than one tuning parameter that results in an increase in computational burden, and they do not provide measures of uncertainty which are often desirable in characterizing heterogeneity in group level analyses. With the exception of a recent penalized neighborhood selection approach by Belilovsky, Varoquaux, and Blaschko (2016), few penalized methods have been vetted for the joint estimation of multiple brain networks. Unfortunately, the approach by Belilovsky, Varoquaux, and Blaschko (2016) cannot be used to obtain positive definite precision matrices that are necessary for quantification of edge strengths via partial correlations. Moreover, a major difficulty under penalized approaches arises when comparing multiple networks, since the estimated network differences may be artifacts resulting from estimation errors under point estimates (Kim, Pan, and Alzheimer's Disease Neuroimaging Initiative 2015). Penalized methods for comparing networks rely on permutation tests that are computationally burdensome and hence not scalable, or they construct null distributions to conduct hypothesis testing (Higgins et al. 2019) that may be restrictive when the associated assumptions are not satisfied. Hence, penalized approaches may not be adequate for inferring network differences between multiple experimental conditions, which is a central objective in this article.

Several Bayesian approaches, including spike-and-slab methods (Peterson, Stingo, and Vannucci 2015; Yu and Dauwels 2016), and continuous shrinkage methods (Carvalho, Polson, and Scott 2010; Polson and Scott 2010; Li, Craig, and Bhadra 2017; Piironen and Vehtari 2017) have been proposed for individual precision matrix estimation. Though Bayesian approaches have proven extremely useful in estimating brain networks (Mumford and Ramsey 2014), few attempts have been made to develop Bayesian methods for the joint estimation of multiple networks. Some existing Bayesian methods for joint network estimation include the approach by Yajima

et al. (2012), who focused on multiple directed acyclic graphs, and the Bayesian Markov random field approach by Peterson, Stingo, and Vannucci (2015) for estimating multiple protein–protein interaction networks. The former cannot be used to obtain undirected brain networks which is the focus of this article, while the latter is only applicable to examples involving a small number of nodes and cannot be scaled up to whole brain network analysis considered in this study. There is also some recent work on jointly estimating multiple temporally dependent networks (Qiu et al. 2016; Lin et al. 2017), but these approaches cannot be directly generalized for the integrative analysis of multiple brain networks across different experimental conditions. The above discussion suggests a clear need for developing flexible and scalable Bayesian approaches for joint estimation of multiple brain networks which pool information across experimental conditions to provide more accurate estimation and inferences. An appealing feature of Bayesian joint modeling approaches is that they provide a rigorous inferential framework for comparing networks at multiple scales using Markov chain Monte Carlo (MCMC) samples, which precludes the need for computationally involved permutation tests or constructing test statistics based on heuristic null distributions. While a separate Bayesian estimation for multiple networks also enables one to test for network differences using MCMC samples, it is unable to pool information across experimental conditions and cannot account for dependencies across multiple related networks.

In this article, we develop a Bayesian GGM approach for jointly estimating multiple networks. This approach models the probability of a connection as a parametric function of a baseline component shared across networks and differential components unique to each network. The shared and differential effects are modeled under a Dirichlet process (DP) mixture of Gaussians prior (Müller, Erkanli, and West 1996), and the edge probabilities are estimated by pooling information across experimental conditions, thereby resulting in the joint estimation of multiple networks. An exploratory analysis of the Stroop task data, which involved deriving the subject-specific network for each of the 45 subjects under the task and rest conditions using the graphical lasso (Friedman, Hastie, and Tibshirani 2008), and then estimating the group level probability for each edge by combining the edge sets across all subjects, followed by a K-means algorithm on the edge probabilities, revealed clearly defined and well separated clusters for these probabilities. This provides a strong motivation for a DP mixture approach to cluster the edge probabilities under BJNL. The role of the edge probabilities is 2-fold—they characterize uncertainty in network estimation and allows one to pool information across networks. The connection strengths are encapsulated via network specific precision matrices, which are modeled separately for each network under a spike-and-slab Bayesian graphical lasso prior conditional on the above edge probabilities. Adopting a joint modeling approach that involves a combination of a parametric link function with flexible DP priors on the shared and differential components within the edge probabilities results in an interpretable and flexible approach. It also enables more accurate estimation of edge strengths and provides improved network comparisons (greater power to detect true differential connections while ensuring adequate control

for false positives), as demonstrated via extensive numerical experiments. Another important advantage in using the DP prior on the components is the robustness to the specification of the parametric link function, as evident from simulation results. The approach, denoted as Bayesian joint network learning (BJNL), is implemented via a fully Gibbs posterior computation scheme.

Our BJNL analysis of the Stroop task data confirmed the hypothesis that brain connections as well as global and local topological characteristics of the brain network are considerably different when subjects are actively engaged in the task as compared with the rest condition, which is not surprising given the difference in the cognitive requirements of the two conditions. The connectivity differences between task and passive fixation also aligned with the theory of global workspace (Gießing et al. 2013), which confirms the biological interpretability of the BJNL findings. Subtler network differences were observed between effortful and relaxed task conditions that is somewhat expected, since these conditions differ only in the mental attitude voluntarily applied to the performance of the same task. The BJNL connectivity differences were concentrated in brain regions previously shown to be differentially activated by a varying degree of willfully applied mental effort (Khachouf et al. 2017), which supports the plausibility of BJNL findings and provides important evidence supporting the scientific hypothesis of connectivity differences formed under the different cognitive conditions. In addition to the BJNL analysis, we also performed a comparative analysis using penalized approaches such as the graphical lasso and the joint graphical lasso (JGL) (Danaher, Wang, and Witten 2014). The analysis under the penalized approaches revealed negligible network differences between effortful and relaxed task performance, and very limited network differences between task and passive fixation, which did not always involve brain regions implicated in previous activation studies (Khachouf et al. 2017). Hence, these findings appeared to be inconsistent with our scientific hypothesis.

In addition to the advantage in terms of biological interpretability of findings, the BJNL analysis of Stroop task data also provides another major advantage over penalized methods for joint modeling in terms of computation time. Unlike BJNL that provided a robust inferential framework for comparing networks via MCMC samples, network comparisons under penalized approaches involved computationally burdensome permutation tests. In fact, the permutation tests under JGL using a full tuning parameter search procedure requires an hour to run per permutation (compared to an overall run time of a few hours for BJNL), which makes the approach impractical in terms of comparing brain networks. Due to this prohibitive computational burden, tuning parameters for JGL were chosen in an iterative manner across permutations, which reduced the computation time to around 27 hr but also potentially resulted in suboptimal performance in brain network analysis. The computation efficiency of BJNL represents an important practical advantage in terms of comparing multiple whole brain networks, compared to other methods for joint estimation of multiple networks that may be hindered by computational bottlenecks.

2. Methodology

2.1. Description of the fMRI Dataset

Forty-five volunteers participated in the study. All subjects were right handed with an average age of 21.9 (SD = 2.2) years. MRI scanning was performed at the N.O.C.S.A.E Hospital in Baggiovara (MO), Italy, using a 3T Philips Achieva scanner. For each subject, the imaging session consisted of the collection of 6 echo-planar imaging (EPI) runs (112 volumes each, TR = 2.5 sec, 25 axial slice, $3 \times 3 \times 3$ mm voxels) and a T1-weighted high-resolution volume (180 sagittal slices, 1mm isotropic voxels) for anatomical reference. While in the scanner, subjects performed a 4-color version of the Stroop task with a button-press response modality (Gianaros et al. 2005). In this task, subjects are presented with a color word displayed in colored fonts in the center of a computer screen and are asked to press a button on a response device corresponding to the font color of the stimulus. There are two types of trials: *congruent* trials, where the font color matches the text (e.g., the word "RED" in red fonts), and *incongruent* trials, where the font color does not match the text (e.g., the word "RED" in green fonts). The "Stroop effect" refers to a significant slowing of response times to the incongruent trials compared to the congruent ones (Stroop 1935). Figure 1 illustrates the Stroop task experiment.

Stimuli were presented in (task) blocks of 30 sec containing six congruent and six incongruent trials appearing in a pseudo-random order with a 2.5 sec intertrial interval. Each task block was alternated with 25 sec-blocks of passive fixation on a centrally presented cross. Six fMRI runs were collected for each subject, with each run consisting of four blocks of task and five blocks of passive fixation appearing in ABABABABA
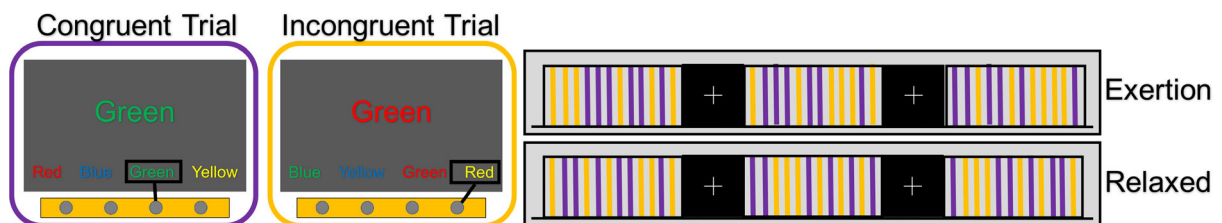
Figure 1. An illustration of the Stroop task involving task blocks of congruent and incongruent trials, indicated by purple bars and yellow bars, respectively, and fixation blocks denoted by a centrally fixated cross. The purple and yellow bars are expanded into two boxes, and the correct button presses are indicated with a rectangle within each box. Subjects were instructed to perform odd-numbered runs "with maximum exertion" (EXR condition) and even-numbered runs "as relaxed as possible" (RLX condition).

order (A = passive fixation, B = task). Crucially, subjects were instructed to perform odd-numbered runs "with maximum exertion" (EXR condition) and even-numbered runs "as relaxed as possible" (RLX condition). This scheme was reversed for a subset of volunteers to check for potential order effects. A major aim of the study was to compare the brain connectivity under REST (passive fixation) and the two TASK conditions (RLX and EXR) (Khachouf et al. 2017).

2.2. Bayesian Modeling of Multiple Networks

We develop a novel Bayesian approach for jointly estimating multiple group-level brain functional networks from multi-subject fMRI data. For each subject, the data are demeaned and prewhitened across time points, where the prewhitened fMRI observations are considered statistically independent. The prewhitened fMRI data over p nodes or regions of interest (ROI) for the ith subject and gth experimental condition at time point t is denoted by $\mathbf{y}_{it}(g) = (y_{it1}(g), \ldots, y_{itp}(g))$, $i = 1, \ldots, n$, $t = 1, \ldots, T_{ig}$, $g = 1, \ldots, G$. Our goal is to jointly estimate multiple networks denoted by $\mathcal{G}_1, \ldots, \mathcal{G}_G$ using GGMs characterized by sparse inverse covariance matrices. The graph $\mathcal{G}_g$ is defined by the vertex set $\mathcal{V} = \{1, \ldots, p\}$ containing p nodes and the edge set $\mathcal{E}_g$ containing all edges/connections in the graph $\mathcal{G}_g$, $g = 1, \ldots, G$.

The prewhitened fMRI measurements for gth experimental condition are modeled as $\mathbf{y}_{it}(g) \sim N_p(\mathbf{0}, \mathbf{\Omega}_g^{-1})$, $i = 1, \ldots, n$, $t = 1, \ldots, T_{ig}$, $g = 1, \ldots, G$, where

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

where $\pi(\cdot)$ denotes the prior distribution, $\omega_{g,kl}$ and $w_{g,kl}$ denote the strength and probability of the functional connection between nodes k and l for network $\mathcal{G}_g$, respectively, M^+ denotes the space of all positive definite matrices, $I(\cdot)$ denotes the indicator function, C_g is the intractable normalizing constant for the prior on the precision matrix, $N_p(\cdot; \mathbf{0}, \mathbf{\Sigma})$ denotes a p-variate Gaussian distribution with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}$, and $E(\alpha)$ and $DE(\lambda)$ denote the exponential and double exponential distributions with scale parameters α^{-1} and λ^{-1}, respectively. Small values of the scale parameters $\tau_{g,kl} \sim \pi(\tau_{g,kl})$ and λ_0^{-1} in Equation (3) result in a spike-and-slab prior (George and McCulloch 1993) on the precision off-diagonals, so that $\mathbf{\Omega}_g \sim \pi(\mathbf{\Omega}_g)$ is denoted as the *spike-and-slab Bayesian graphical lasso*. The spike-and-slab prior shrinks the values corresponding to absent edges toward zero and encourages values away from zero for important connections. The slab component is modeled under a Gaussian distribution having thick tails under small values of the precision parameter, while the spike component is modeled under a double exponential distribution having a sharp spike at zero under a large value of λ_0. It is straightforward to show that $C_g < \infty$ so that the prior in model (1) is proper using the results in Wang (2012).

2.2.1. Pooling Information Across Experimental Conditions

Information is pooled across experimental conditions to estimate the edge weights $w_{g,kl}$, $k \neq l, k, l = 1, \ldots, p$, leading to joint estimation of multiple networks. Note that by pooling information to model the edge probabilities instead of the edge strengths, we are able to jointly model multiple brain networks without constraining the edge strengths in separate networks to be similar. The prior weights represent the unknown probabilities of having functional connections, and are modeled via a parametric link function comprising unknown shared and differential effects as described below

$$w_{g,kl} = h(\eta_{0,kl}, \eta_{g,kl}), \quad \eta_{0,kl} \sim f_0, \quad \eta_{g,kl} \sim f_g,$$
$$f_0 \sim DP(MP_0), \quad f_g \sim DP(MP_0), \tag{2}$$

for $k \neq l, k, l = 1, \ldots, p, g = 1, \ldots, G$, where $h(\cdot)$ is the parametric link function relating the probability for edge (k, l) in network $\mathcal{G}_g$ to the network specific differential effect $(\eta_{g,kl})$ and common effect $(\eta_{0,kl})$ across all networks, and $DP(MP_0)$ denotes a Dirichlet process mixture prior defined by the precision parameter M and base measure $P_0 \equiv N(0, \sigma_\eta^2)$. The Dirichlet process mixture prior induces a flexible class of distributions on the edge probabilities and also results in clusters of edges having the same prior inclusion probabilities, enforcing parsimony in the number of model parameters. The number of clusters and the cluster sizes are unknown and controlled via the precision parameter M (Antoniak 1974).

Under specification (2), the baseline effect $\eta_{0,kl}$ represents the shared feature for edge (k, l) which is estimated by pooling information across experimental conditions, resulting in the joint estimation of multiple networks. The baseline effect controls the overall probability of having an edge across all networks, while the differential effects contribute to the network specific variations which are estimated using the information from individual experimental conditions. For example, large differences between $\eta_{g,kl}$ and $\eta_{g',kl}$, $g \neq g'$ potentially imply a differential status for edge (k, l) between $\mathcal{G}_g$ and $\mathcal{G}_{g'}$. On the other hand when $\eta_{g,kl} = \eta_{g',kl}$, $g \neq g'$, the model specifies equal probability for edge (k, l) in networks $\mathcal{G}_g$ and $\mathcal{G}_{g'}$. For ease in interpretability, we choose a logistic form link in (2) as $h(\eta_{0,kl}, \eta_{g,kl}) = \exp\{\eta_{0,kl} + \eta_{g,kl}\}/[1 + \exp\{\eta_{0,kl} + \eta_{g,kl}\}]$, $g = 1, \ldots, G$, so that $\eta_{0,kl} + \eta_{g,kl}$ can be interpreted as the log odds of having the edge (k, l) in the network $\mathcal{G}_g$, and the log odds ratio of having edge (k, l) in the brain network $\mathcal{G}_g$ versus $\mathcal{G}_{g'}$ can be expressed as $\eta_{g,kl} - \eta_{g',kl}$ $(g \neq g')$. A schematic representation of the proposed model is illustrated in Figure 2.

Note that the parameters $\eta_{0,kl}, \eta_{g,kl}$, in (2) are not identifiable since $h(\eta_{0,kl}, \eta_{g,kl}) = h(\eta_{0,kl} + c, \eta_{g,kl} - c)$ for any real constant c. However, the functionals of interest such as the log odds $(\eta_{0,kl} + \eta_{g,kl})$, the log-odds ratio $(\eta_{g,kl} - \eta_{g',kl})$, and the edge probabilities themselves are clearly identifiable, which is adequate for our purposes. The proposed specification (2) is purposely overcomplete, which is an issue routinely arising in Bayesian models. By "overcomplete," we mean that we include $G + 1$ parameters in the weights model when G parameters would suffice. Such overcompleteness allows us to pool information in a systematic manner, and ensures computational efficiency and interpretability in terms of shared and differential group effects and is designed to avoid any problems in identifiability
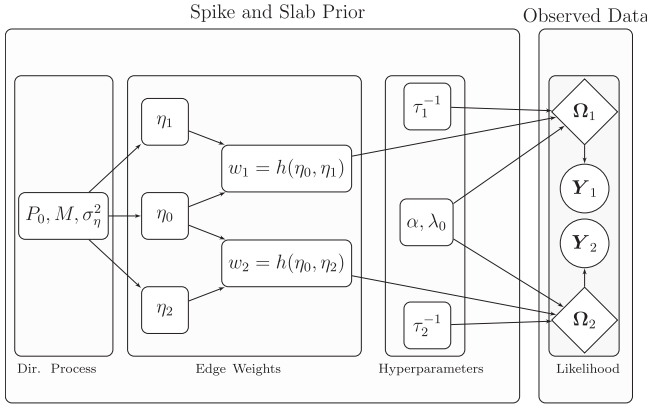
Figure 2. Directed graph illustrating the relationships between the model parameters for the case of two experimental conditions represented by fMRI data matrices Y_1 and Y_2. Rectangular nodes correspond to parameters which are updated or tuned, diamond-shaped nodes correspond to parameters involved in the likelihood, and the circular nodes correspond to the observed data.

of functionals of interest—refer, for example, to Ghosh and Dunson (2009).

Our treatment of the edge weights is motivated by existing literature on modeling binary or ordered categorical responses using mixture distributions (Kottas, Müller, and Quintana 2005; Jara, García-Zattera, and Lesaffre 2007; Gill and Casella 2009; Canale and Dunson 2011). Specifically we are able to achieve both the interpretability discussed above and a high degree of flexibility while also reducing the sensitivity to the link function and enabling straightforward posterior computation. A similar approach was taken by Durante, Dunson, and Vogelstein (2017) who modeled structural connections in a population of networks via a mixture of Bernoulli distributions, although they did not focus on joint estimation of multiple networks.

3. Posterior Computation

We design a block Gibbs sampler to fit the proposed model (1). The sampler enables data adaptive shrinkage by introducing latent scale parameters to sample the precision matrix off-diagonals corresponding to the spike component under a scale mixture representation of Gaussians while defining conjugate priors on the precision parameters in the slab component. Define edge inclusion indicators as $\delta_{g,kl} = 1$ if edge (k,l) is included in $\mathcal{G}_g$, and $\delta_{g,kl} = 0$ otherwise, where $P(\delta_{g,kl} = 1) = w_{g,kl}$. The augmented likelihood for Equation (1) can be written as

$$\pi(\boldsymbol{\Omega}_g \mid \lambda_0, \boldsymbol{\tau}_g, \boldsymbol{\tau}_g^*) \tag{3}$$

$$= C_{\tau,g}^{-1} \mathbb{1}(\boldsymbol{\Omega}_g \in M^+) \prod_{l=1}^p \text{Exp}(\omega_{g,ll}; \alpha/2)$$

$$\times \prod_{l=1}^p \prod_{k<l} w_{g,kl}^{\delta_{g,kl}} (1 - w_{g,kl})^{1-\delta_{g,kl}}$$

$$\times \prod_{l=1}^p \prod_{k<l} [N(\omega_{g,kl}; 0, \tau_{g,kl}^{-1})]^{\delta_{g,kl}}$$

$$\left[\int N(\omega_{g,kl}; 0, \tau_{g,kl}^{-1}) \text{Exp}(\tau_{g,kl}; \frac{\lambda_0^2}{2}) d\tau_{g,kl}\right]^{1-\delta_{g,kl}},$$

with

$$\pi(\boldsymbol{\tau}_g, \boldsymbol{\tau}_g^*) \propto C_{\tau,g} \left(\prod_{l=1}^p \prod_{k<l} \text{Ga}(\tau_{g,kl}; a_\tau, b_\tau) \times \text{Exp}(\tau_{g,kl}^*; \lambda_0^2/2) \right),$$

where $\boldsymbol{\tau}_g = \{\tau_{g,kl}, k \neq l, k, l = 1, \ldots, p\}$, $\boldsymbol{\tau}_g^* = \{\tau_{g,kl}^*, k \neq l, k, l = 1, \ldots, p\}$, $\text{Ga}(\cdot; a_\tau, b_\tau)$ corresponds to a Gamma distribution with mean a_τ/b_τ, and $C_{\tau,g}$ is the intractable normalizing constant which cancels out in the expression for $\pi(\boldsymbol{\Omega}_g, \lambda_0, \boldsymbol{\tau}_g, \boldsymbol{\tau}_g^*)$ to yield a marginal prior $\pi(\boldsymbol{\Omega}_g, \lambda_0, \boldsymbol{\tau}_g)$ as in (1) after integrating out $\boldsymbol{\tau}_g^*$. In our implementation, we prespecify $\lambda_0 = 100$ to ensure a sharp spike at zero leading to strong shrinkage for precision off-diagonals corresponding to absent edges. On the other hand, we choose a_τ and b_τ such that a_τ/b_τ is small, enabling adaptive thick tails for the Gamma prior on the latent scale parameters corresponding to the slab component.

We choose a logistic link function in (2) for our purposes, although more general link functions can also be used. For implementing a fully Gibbs sampler, we rely on an approximation to the logistic function using a probit link, which employs a data augmentation scheme as in O'brien and Dunson (2004). In particular,

$$\frac{e^{\mu*}}{(1 + e^{\mu*})} \approx \int_0^\infty \mathbf{t}(u; \mu*, \frac{\pi^2(\phi - 2)}{3\phi}) du$$

$$= \int_0^\infty N(u; \mu*, \frac{\pi^2(\phi - 2)}{3\phi}\sigma_\phi^2)\pi(\sigma_\phi^2; \frac{\phi}{2}, \frac{\phi}{2}) du,$$

where $\mathbf{t}(\cdot)$ denotes a t-distribution, $\pi(\sigma_\phi^2)$ corresponds to an inverse Gamma distribution, $\phi = 7.3$, and u is the Gaussian latent variable used for data augmentation. This approximation results in sampling from a posterior that is approximately equal to the posterior under specification (1) and (2) using a logistic link function. Although such an approximation is used, we note that the resulting posterior computation is fully Gibbs since all MCMC samples are drawn from exact posterior distributions. Alternatively, one could adapt the Polya-gamma data augmentation in Polson, Scott, and Windle (2013) for Bayesian logistic regression. However, the approximation in O'brien and Dunson (2004) works reasonably well in a wide variety of numerical studies in our experience. Moreover, the stick-breaking representation (Sethuraman 1994) is used for the Dirichlet process mixture prior in (2), which facilitates posterior computation and can be written as

$$\eta_{g,kl} \sim f_g, f_g = \sum_{h=1}^\infty v_{g,h} \delta_{\eta_{g,h}^*}, \eta_{g,h}^* \sim N(0, \sigma_\eta^2),$$

$$v_{g,h} = (v_{g,h} \prod_{l<h} [1 - v_{g,l}]), v_{g,h} \sim \text{Beta}(1, M),$$

$$g = 0, \ldots, G, \tag{4}$$

where Beta($\cdot$) denotes a Beta distribution. The slice sampling technique (Walker 2007) is used to sample the atoms from the infinite mixture in (2), which significantly expedites computation. See Section 1 of the supplementary materials for posterior computation details.

3.1. Edge Detection

The important network edges (and hence the network structure) can be estimated by either including edges with high marginal inclusion probabilities or those with nonnegligible absolute values for the precision off-diagonals, lying above a chosen threshold. We propose a strategy to choose such thresholds in a manner which controls the false discovery rate (FDR). Denoting $\zeta_{g,kl}$ as the marginal posterior exclusion probability for edge (k, l) in network $\mathcal{G}_g$, one can compute the FDR as in Peterson, Stingo, and Vannucci (2015) as

$$\text{FDR} = \frac{\sum_{g=1}^{G} \sum_{k<l} \zeta_{g,kl} 1(\zeta_{g,kl} < \kappa)}{\sum_{g=1}^{G} \sum_{k<l} 1(\zeta_{g,kl} < \kappa)}, \text{ or}$$

$$\text{FDR} = \frac{\sum_{g=1}^{G} \sum_{k<l} \zeta_{g,kl} 1(|\hat{\omega}_{g,kl}| > \kappa^*)}{\sum_{g=1}^{G} \sum_{k<l} 1(|\hat{\omega}_{g,kl}| > \kappa^*)}, \quad (5)$$

depending on whether the edges are included based on posterior inclusion probabilities or edge strengths. Clearly the FDR increases with κ/κ^*, and one can choose a suitable threshold to control the FDR. In our numerical experiments, we found that choosing the edges based on whether the absolute precision off-diagonals were greater than 0.1 results in overall good numerical performance and FDR values which are less than 0.03 across a wide spectrum of scenarios. Hence, we recommend this as a default threshold under our approach, and we note that the corresponding threshold for posterior probability for edge selection can be obtained as one which yields similar FDR as computed using (5).

3.2. Inferring Network Differences

In addition to network estimation, the proposed BJNL provides a natural framework for testing network differences between experimental conditions at multiple scales. In particular, for our Stroop task data analysis, we use MCMC samples under BJNL to obtain the posterior distribution for differences in edge level partial correlations as well as global and local network metrics. At the edge-level, T-tests of the Fisher Z-transformed partial correlation differences for all MCMC samples (after burn-in) were used to infer differences in edge strengths across networks. Similarly, the differences in the graph metrics across conditions were computed at each MCMC iteration, and the central tendency and dispersion of their distributions were statistically assessed by T-tests and Kolmogorov–Smirnov (KS) tests. The p-values of these tests were used to assess significance after controlling for false discoveries (Benjamini and Hochberg 1995).

4. Numerical Studies

4.1. Simulation Setup

We conducted a series of simulations to compare group level network estimation between BJNL and competing methods. These approaches include the graphical horseshoe estimator (HS) (Carvalho, Polson, and Scott 2010; Li, Craig, and Bhadra 2017) which extends the horseshoe prior in regression settings to graphical model estimation, and the graphical lasso approach (GL) (Friedman, Hastie, and Tibshirani 2008) which imposes

L_1 penalty on the off-diagonals to impose sparsity, as well as the joint graphical lasso (JGL) (Danaher, Wang, and Witten 2014) which uses a fused lasso penalty to pool information across graphs while encouraging sparsity via a L_1 penalty. While both the HS and GL approaches estimate individual networks separately, the JGL approach is designed to jointly estimate multiple networks. The HS was implemented using Matlab codes provided on the author's website. The JGL and the graphical lasso were implemented using the *JGL* and *glasso* packages in R, respectively. Our method was implemented in Matlab, version 8.3.0.532 (R2014a), and a GUI implementing the method has been submitted as a supplementary materials.

The data for the simulation study was generated under a GGM for $n = 60$ subjects with $T = 300$ time points each and for dimensions $p = 40, 100$. Each subject had data corresponding to two experimental conditions having networks with shared and differential patterns. We considered three different network structures: (a) Erdos–Renyi networks which randomly generate edges with equal probabilities, (b) small-world networks generated under the Watts–Strogatz model (Watts and Strogatz 1998), and (c) scale-free networks generated using the preferential attachment model (Barabási and Albert 1999) resulting in a hub network. For each type of network, we obtained an adjacency matrix corresponding to the first experimental condition, and then flipped a proportion of the edges in this adjacency matrix to obtain the second network, adding edges where there were no edges and removing an equal number of edges. The proportion of flipped edges was set to 25% (low), 50% (medium), and 75% (high), which correspond to varying levels of discordance between the experimental conditions.

After generating the networks, the corresponding precision matrices were constructed as follows. For each edge, we generated the corresponding off-diagonal element from a Uniform$(-1, 1)$ distribution and fixed the diagonal elements to be one and the off-diagonals corresponding to absent edges as zero. To ensure that the resulting precision matrices were positive definite, we subtracted the minimum of the eigenvalues from each diagonal element of the generated precision matrix. To enable a group level comparison for each scenario, all subjects had the same network across all time points within each experimental condition and the same precision matrices for each network.

4.1.1. Tuning

We used BJNL with 1000 burn-in iterations and 5000 MCMC iterations. We specified the tuning parameters as follows. We chose $\lambda_0 = 100$ and $\tau_{g,kl} \sim \text{Ga}(a_\tau, b_\tau)$ with $a_\tau = 0.1$ and $b_\tau = 1$ in prior specification (3) to enforce a sharp spike at zero and thick tails for the slab component. The stick breaking weights in the mixture distribution in (4) were modeled as $\nu_{g,h} \sim \text{Be}(1, M)$, where $M \sim \text{Ga}(a_m, b_m)$, and we choose $a_m = 1, b_m = 1$, to encourage a small number of edge clusters for a parsimonious representation. We could increase a_m to encourage a larger number of clusters. However, we have observed that varying a_m has a limited effect on the final estimated network, as demonstrated through simulations in Section 2 of the supplementary materials. Our experience in extensive numerical studies suggests that the performance of

the approach is not overly sensitive to the choice of λ_0 as long as it is large enough (>100); however, extremely large values of λ_0 can result in numerical instability. Moreover, performance is fairly robust to the choice of the hyperparameters in the prior for the precision parameter of the slab component in (3), as long as the ratio $a_\tau/b_\tau < 1$. For the joint graphical lasso that depends on two tuning parameters (a lasso penalty and a fused lasso penalty), we searched a 30×30 grid over $[0.01, 0.1]$ for both parameters to find the best combination using a AIC criteria as recommended in Danaher, Wang, and Witten (2014). The graphical lasso was run independently for each network over a grid of regularization parameter values, and the optimal graph was selected for each network using a BIC criteria as described in Yuan and Lin (2007).

4.1.2. Performance Metrics

We assessed the performance of the three algorithms in terms of the ability to estimate the individual networks, as measured by the area under the receiver operating characteristic (ROC) curve (AUC), the accuracy in estimating the strength of connections, as measured by the L_1 error in estimating the precision matrix (L_1 error), the power to detect true differential edges as measured via sensitivity (TPR) and control over false positives for differential edges which is computed as 1 − specificity (FPR). For all the metrics, we performed pairwise comparisons using Wilcoxon signed rank tests to assess whether one approach performed significantly better than the others. For edge detection, point estimates for the penalized networks were obtained by choosing the threshold for the absolute off-diagonal elements as 0.005, while for BJNL we computed thresholds controlling for false discoveries as described in Section 3.

4.2. Simulation Results

Figure 3 displays the ROC curves for the 100 node simulations, Figure 4 displays boxplots of the reported metrics for the Erdos–Renyi case, and Table 1 reports results for the 100 node simulations. The boxplots for the other networks and the results for the 40 node case are reported in the supplementary materials due to space constraints. The results across the three network types are relatively consistent. First, we note that the degree of dissimilarity between the networks does not appear to have a major effect on the relative performance of the algorithms, although

we conjecture that the differences could be more pronounced for smaller sample sizes. For all settings involving Erdős–Renyi graphs, the proposed BJNL approach outperformed the HS, JGL, and GL uniformly across all metrics under the Wilcoxon signed rank test. Notably, the proposed approach simultaneously achieved a significantly higher TPR and a significantly lower FPR for differential edges, indicating that it was both better able to detect significant differences and less likely to incorrectly classify an edge as differential. These, and the additional boxplots in the supplementary materials, suggest a greater power to detect true differential edges with an adequate control over false positives across all network types, under the BJNL. Further, an increased improvement of the TPR over competing approaches and relative stability of the FPR for differential edges for $p = 100$ versus $p = 40$ indicates a clear advantage of the proposed joint estimation approach for increasing dimensions. For the small-world and scale-free networks, the BJNL also had significantly improved AUC, TPR, and L_1 error metrics, and a comparable or lower FPR, compared to all other considered approaches.

On the other hand, the significantly higher L_1 error under the JGL potentially points to the perils of smoothing over edge strengths across networks under penalized approaches. In particular, assigning similar magnitudes for precision matrix off-diagonals for shared edges may adversely affect the identification of differential edges, as well as the estimation of varying edge strengths for common edges across networks. Moreover while HS has low FPR, it consistently exhibits the lowest AUC and TPR and the highest L_1 error for $p = 100$ across all scenarios, which is concerning. On the other hand, the GL had the highest FPR for both the small-world and scale-free network simulations, but has a reasonable TPR. These results under HS and GL illustrate the difficulties resulting from the separate estimation of individual networks which may result in exceedingly low power to detect true positives (as with HS), or an inflated number of false positives (as with GL).

To examine the sensitivity of the proposed approach with respect to the chosen link function, we performed additional simulation studies by fitting the proposed model to the 100 node data generated as above, but under a probit link. The results in Table 2 illustrate nonsignificant differences in the performance metrics for network estimation across the logit and the probit links, which illustrate the robustness of the proposed approach resulting from the specification of the DP prior on the shared and differential components in (2).
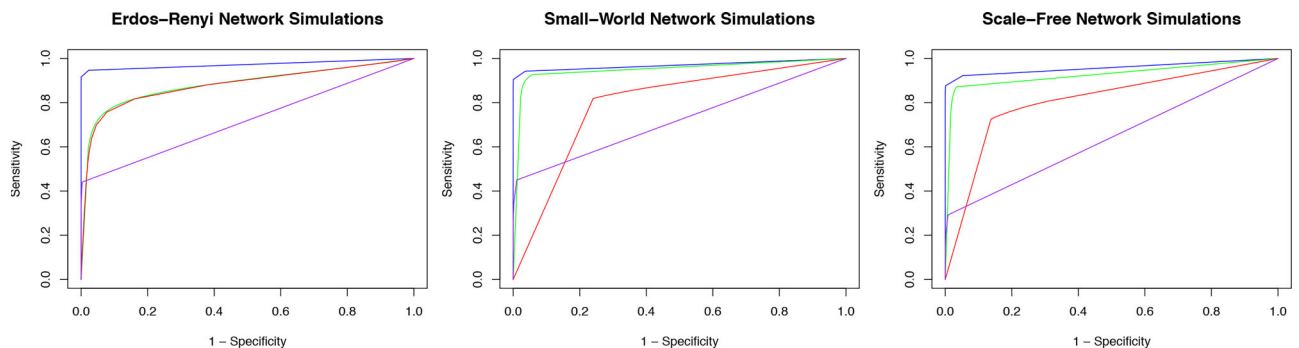
Figure 3. ROC curves for edge detection for the 100 node simulations. The blue, green, red, and purple solid lines correspond to BJNL, JGL, GL, and HS, respectively.

Figure 4. Boxplots of the AUC, L_1 error, and TPR/FPR for differential edge detection for the Erdős–Renyi simulations for Bayesian joint network learning (BJNL), the joint graphical lasso (JGL), graphical lasso (GL), and the graphical horseshoe estimator (HS). Within each approach, the boxplots are organized as: low difference, medium difference, and high difference in edges between experimental conditions, in that order.

Table 1. 100 node simulation results comparing Bayesian joint network learning (BJNL), the joint graphical lasso (JGL), graphical lasso (GL), and the graphical horseshoe estimator (HS).
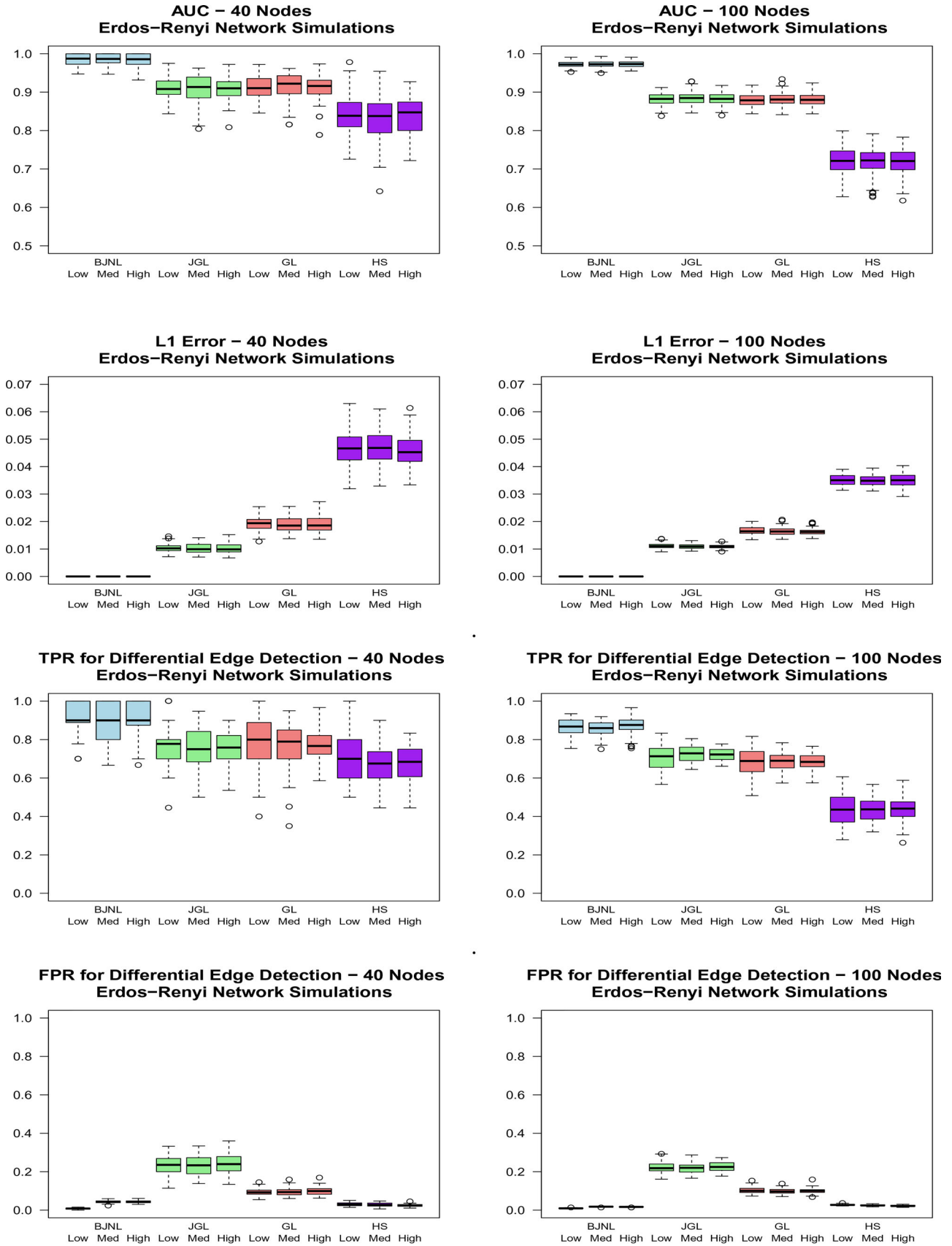
	AUC				L_1 Error $\times$ 100			
	BJNL	JGL	GL	HS	BJNL	JGL	GL	HS
Erdos–Renyi								
Low	**0.97 (0.01)**	0.88 (0.02)	0.88 (0.02)	0.72 (0.03)	**0.11 (0.01)**	1.11 (0.09)	1.66 (0.13)	3.51 (0.19)
Med	**0.97 (0.01)**	0.88 (0.02)	0.88 (0.02)	0.72 (0.04)	**0.11 (0.01)**	1.09 (0.09)	1.65 (0.14)	3.50 (0.20)
High	**0.97 (0.01)**	0.88 (0.02)	0.88 (0.02)	0.73 (0.03)	**0.11 (0.01)**	1.09 (0.07)	1.62 (0.11)	3.50 (0.23)
Small world								
Low	**0.97 (0.01)**	0.95 (0.01)	0.79 (0.01)	0.72 (0.04)	**0.25 (0.01)**	0.75 (0.12)	2.06 (0.08)	4.70 (0.15)
Med	**0.97 (0.01)**	0.95 (0.01)	0.80 (0.01)	0.72 (0.03)	**0.24 (0.01)**	0.77 (0.13)	2.07 (0.08)	4.65 (0.14)
High	**0.97 (0.01)**	0.95 (0.01)	0.79 (0.01)	0.73 (0.03)	**0.24 (0.01)**	0.78 (0.13)	2.06 (0.08)	4.65 (0.14)
Scale free								
Low	**0.96 (0.01)**	0.93 (0.01)	0.81 (0.01)	0.64 (0.03)	**0.20 (0.01)**	1.01 (0.20)	2.23 (0.10)	5.30 (0.23)
Med	**0.96 (0.01)**	0.92 (0.01)	0.81 (0.01)	0.64 (0.03)	**0.19 (0.01)**	1.02 (0.21)	2.24 (0.90)	5.26 (0.24)
High	**0.96 (0.01)**	0.92 (0.01)	0.81 (0.01)	0.64 (0.03)	**0.19 (0.01)**	1.00 (0.21)	2.20 (0.08)	5.23 (0.23)

	TPR				FPR			
	BJNL	JGL	GL	HS	BJNL	JGL	GL	HS
Erdos–Renyi								
Low	**0.87 (0.05)**	0.71 (0.07)	0.68 (0.07)	0.43 (0.08)	0.01 (0.001)	0.22 (0.03)	0.10 (0.02)	0.03(0.00)
Med	**0.88 (0.04)**	0.73 (0.04)	0.69 (0.05)	0.44 (0.06)	0.01 (0.001)	0.22 (0.03)	0.10 (0.01)	0.03(0.00)
High	**0.88 (0.02)**	0.72 (0.03)	0.69 (0.04)	0.44 (0.06)	0.01 (0.001)	0.23 (0.02)	0.10 (0.02)	0.02 (0.00)
Small world								
Low	**0.86 (0.04)**	0.47 (0.07)	0.66 (0.06)	0.44 (0.07)	0.02 (0.002)	0.02 (0.00)	0.36 (0.01)	0.06 (0.01)
Med	**0.86 (0.04)**	0.49 (0.04)	0.67 (0.04)	0.46 (0.05)	0.02 (0.002)	0.02 (0.00)	0.36 (0.01)	0.05 (0.01)
High	**0.86 (0.02)**	0.48 (0.04)	0.67 (0.03)	0.46 (0.05)	0.01 (0.002)	0.02 (0.00)	0.36 (0.01)	0.05 (0.01)
Scale free								
Low	**0.87 (0.05)**	0.39 (0.06)	0.63 (0.07)	0.25 (0.06)	0.02 (0.002)	0.02 (0.00)	0.24 (0.03)	0.04 (0.01)
Med	**0.87 (0.03)**	0.41 (0.05)	0.63 (0.04)	0.26 (0.05)	0.02 (0.002)	0.02 (0.00)	0.24 (0.02)	0.04 (0.01)
High	**0.87 (0.03)**	0.42 (0.04)	0.64 (0.04)	0.27 (0.05)	0.01 (0.002)	0.02 (0.00)	0.25 (0.02)	0.04 (0.01)

NOTE: Text in bold indicates a method was better than all other competing methods as assessed through Wilcoxon signed rank tests at $\alpha = 0.05$.

Table 2. Comparison of the 100 node simulation results using the probit link function to the simulation results using the logit link function.

	Erdos–Renyi			Small world			Scale free		
	AUC	TPR	FPR	AUC	TPR	FPR	AUC	TPR	FPR
Probit	0.97	0.88	0.01	0.96	0.86	0.02	0.97	0.87	0.02
Logit	0.97	0.88	0.01	0.97	0.87	0.02	0.96	0.86	0.02

5. Stroop Task Analysis

5.1. Description of Analysis

We applied the proposed BJNL to the fMRI Stroop task study to investigate similarities and differences in the brain network under the two experimental conditions and passive fixation (REST). The first analysis was aimed at comparing the mental states of task performance (TASK) and passive fixation (REST), with the hypothesis that the brain networks exhibit major differences between these two grossly different conditions. The TASK data consisted of the subject-wise concatenation of the prewhitened fMRI time courses acquired during the exertion (EXR) and relaxed (RLX) task blocks, while the REST data consisted of the subject-wise concatenation of the prewhitened fMRI time courses acquired during the passive fixation blocks. The second analysis aimed to detect finer differences in connectivity between the mental states of EXR and RLX task performance. The study hypothesized that the mental states should be similar between the two task conditions with some fine network differences. In this case, the subject-wise prewhitened fMRI time courses were concatenated for the EXR blocks and also separately for the RLX blocks for analysis.

We performed a brain network analysis based on region of interest (ROI) level data, adopting the 90 node Automated Anatomical Labeling (AAL) cortical parcellation scheme described in Tzourio-Mazoyer et al. (2002). For each ROI, we estimated the representative BOLD time series by performing a singular value decomposition on the time series of the voxels within the ROI and extracting the first principal time series. This resulted in 90 time courses of fMRI measurements, one for each ROI, which were then demeaned. We classified each ROI into one of nine functional modules as defined in Smith et al. (2009). We performed standard preprocessing including slice-timing correction, warping to standard Talairach space, blurring, demeaning, and prewhitening. The fMRI time series was prewhitened using an ARMA(1, 1) model, as is common in imaging toolboxes such as AFNI (Cox 1996). Further details are provided in Section 5 of the supplementary materials. The proposed BJNL was run using the same tuning parameters as in the simulations. Dickey–Fuller tests of stationarity (Dickey and Fuller 1979) were performed to assess convergence of the MCMC sampler (see Section 7 of the supplementary materials). We also examined the widths of the credible intervals in Section 8 of the supplementary materials, where Figure 7 of the supplementary materials demonstrates that the credible intervals

for absent-edges are much narrower than the credible intervals for present-edges. Finally, we performed chi-squared goodness of fit tests under BJNL (see Section 9 of the supplementary materials).

5.1.1. Graph Metrics

We analyzed the brain's connectivity structure during the different mental states in terms of four graph metrics: global efficiency, local efficiency, clustering coefficient, and characteristic path length. Efficiency and characteristic path length measure how effectively information in transmitted between nodes, while clustering coefficient measures the interconnectedness of the network—see Section 6 in the supplementary materials for a full description. All metrics were calculated using the Matlab Brain Connectivity Toolbox (Rubinov and Sporns 2010). In addition, we also examined differences in local graph metrics across experimental conditions corresponding to several brain regions (see Tables 2 and 3 in the supplementary materials) that were found to be differentially activated in a previous study using the same Stroop task experiment (Khachouf et al. 2017). Although distinct from earlier activation analysis, potential connectivity differences in these previously identified brain regions will bolster earlier activation based discoveries and also help illustrate the biological interpretability of the connectivity analysis.

5.2. Results

5.2.1. TASK Versus REST Conditions

The analysis produced a large contingent of edges with significantly different edge strengths in the two mental states—Figure 6 displays a heatmap of the significant edge counts by functional module. Our analysis revealed 1550 significantly different edges (under T-tests) that provide evidence supporting the study hypothesis that there are major differences in the

brain networks due to the manifest phenomenological and procedural dissimilarity of task performance and rest. Moreover, our examination revealed significant differences in the mean (under T-tests) and the posterior distributions (under KS tests) for all network metrics between the two conditions (Figure 5). Additional examination of local network differences between task and fixation conditions corresponding to 20 prespecified regions revealed larger clustering coefficients for REST in all implicated regions, and larger local efficiencies for REST in 18 of the 20 regions (see Table 2 in the supplementary materials for the brain regions and p-values).

5.2.2. EXR Versus RLX Conditions of Task Performance

Compared with the relatively large network differences between TASK and REST, the network structures corresponding to the EXR and RLX task conditions exhibited more nuanced differences. Our analysis revealed 226 significantly different edges between the EXR and RLX conditions—see Figure 6 for a heatmap of the significant edge counts by functional module. None of the graph metrics were significantly different between the EXR and RLX conditions, implying that the network differences did not manifest at a global level (Figure 5). However, more localized changes were discovered in the preselected regions that were previously shown to be differentially activated between EXR versus RLX (Khachouf et al. 2017). Significant differences were found in terms of mean local efficiency in the right inferior occipital node and the left caudate. Similarly, significant differences were found in mean and distribution for the clustering coefficient for the right inferior occipital node. Several borderline network differences were also identified—see Table 3 in the supplementary materials for reported p-values.

5.2.3. Interpretation of Findings

Our BJNL analysis identified strong connectivity differences between Stroop task performance and passive fixation in terms
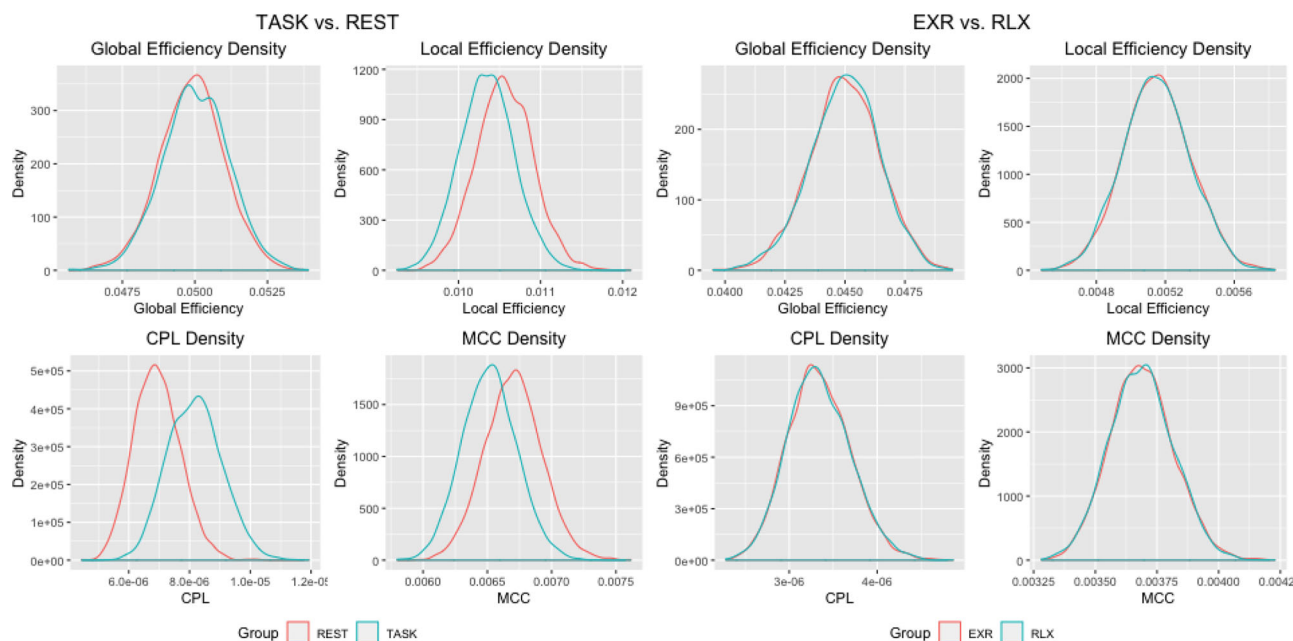

Figure 5. Estimated densities of graph metrics for the analysis of task versus passive fixation and maximum exertion (EXR) versus relaxed (RLX) task performance.
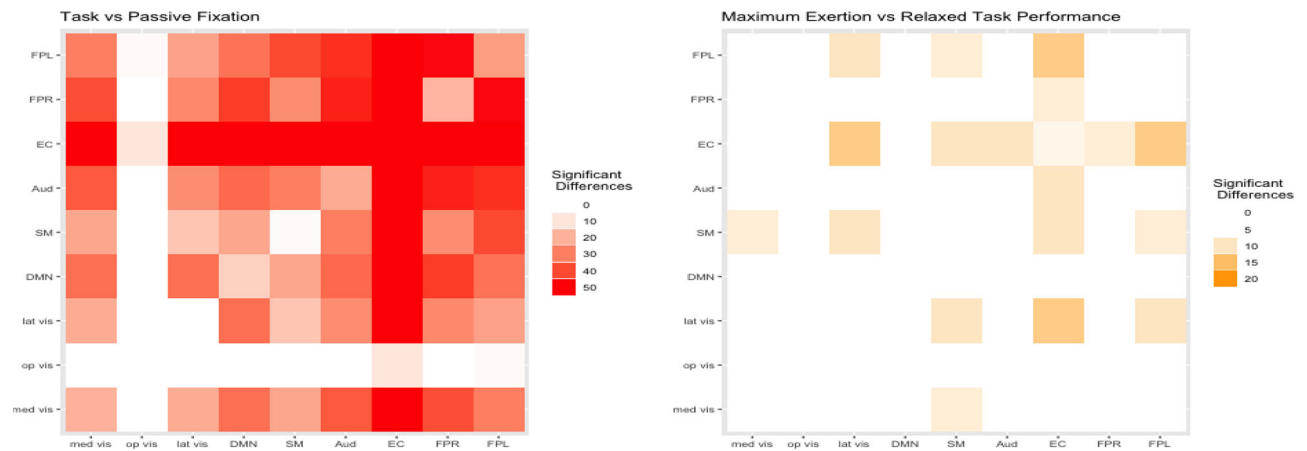
Figure 6. Heatmaps of the number of differential edges between conditions. The heatmap on the left corresponds to the analysis of task versus passive fixation, and the heatmap on the right corresponds to the analysis of maximum exertion (EXR) versus relaxed task performance (RLX).

of significantly higher efficiency and clustering, and lower characteristic path length for REST as well as stronger positive connections involving frontoparietal circuits, EC, DMN, sensorimotor, and visual cortices in the TASK condition compared to REST. Our findings also aligned with the widely used theory of global workspace where more difficult tasks are associated with increased connection distance, as well as reduced clustering (Gießing et al. 2013). More localized associations were also discovered in all regions identified as differentially activated in previous studies (Khachouf et al. 2017), which highlights the biological interpretability of our connectivity findings. Our analysis provides exciting new insights into the connectivity differences between passive fixation and the task-related network that requires a rearrangement of connections to perform the task.

On the other hand, fewer connectivity differences were discovered between EXR and RLX task performance, as expected due to the only difference between conditions being the level of voluntary effort invested in the task. While no global topological differences between the EXR and RLX conditions were discovered, the BJNL analysis did reveal some fine differences in the functional modules including the EC and FPL that are involved in high level cognitive function, as well as some limited localized connectivity differences in 23 prespecified brain regions that previously showed major activation differences in Khachouf et al. (2017). In general compared to TASK versus REST, the RLX task performance condition featured significantly more negative connections between regions compared to EXR, and there were fewer connectivity differences between nodes within the EC. Compared to a much larger number of connectivity differences in EC and other functional modules in TASK versus REST, the limited connectivity differences between EXR and RLX implies a restricted rearrangement of the network between EXR and RLX.

5.2.4. Comparison With Penalized Approaches

We are also interested in comparing the network differences under BJNL with those obtained under penalized methods. Hence, we performed an illustrative analysis for the Stroop task data using the GL and JGL approaches that involved permutation testing to infer significant network differences between experimental conditions. A permuted sample for two

experimental conditions was generated by randomly switching the labels across conditions multiple times. Then, the networks corresponding to these permuted samples were computed using JGL and GL. Subsequently, the network differences corresponding to distinct experimental conditions were computed. The above process was repeated 10,000 times, and the permutation distributions for between-network differences were constructed to compute p-values to test for significant differences.

Note that it was computationally infeasible to use AIC to select the tuning parameters for JGL for all 10,000 permutations since the run time for the best tuning parameter search over a grid took 1 hr per permutation. Hence, the starting values for the tuning parameters for JGL were selected as those values used for the JGL analysis for the original data without permutation. The tuning parameters were then adaptively searched on a permutation-by-permutation basis until the resulting edge density was within 20% of the edge density for the network corresponding to the original samples. While the process was required to make testing under the JGL computationally feasible, it could potentially result in misleading results under JGL due to possible mis-specification of network densities.

The analysis revealed that only one of the resulting edges for the EXR versus RLX network comparisons was significant under the GL, whereas only 62 edges were significant for the JGL. Similarly, for the analysis of TASK versus REST, the JGL identified 476 edges with differential strengths and 3873 common edges (versus 1550 differential edges and 1565 common edges under BJNL). In this case, GL was able to identify 51 edges with differential strengths, and 552 common edges. We believe that the low number of differential edges between EXR versus RLX conditions under the penalized approaches is unrealistic, and that more differences are to be expected between TASK and REST since it involves significant differences in brain activation across the brain (Khachouf et al. 2017). Further, only 5 of the 20 prespecified brain regions which were shown to be differentially activated had significant network differences between TASK and REST under the penalized approaches (see Table 1 in the supplementary materials). These results suggest the proposed BNJL method has much better statistical power to detect differences in brain networks under different cognitive states compared to penalized approaches for modeling networks.

6. Discussion

In this article, we introduced a novel Bayesian approach to joint estimation of multiple group level brain networks that pools information across networks to estimate shared and differential patterns in brain functional networks formed under different cognitive conditions. The proposed BJNL approach naturally enables a systematic inferential framework for comparing networks, which is a central question of interest in many connectome studies including our Stroop task application where the focus is to investigate connectivity differences between passive fixation and relaxed and exertion modes of Stroop task. Our analysis of Stroop task data revealed important dissimilarities between the task and rest conditions, but more subdued differences between the two task conditions, which aligns with the scientific hypotheses of the study. Moreover the connectivity differences were found to be concentrated in brain regions shown to be differentially activated for Stroop task in previous studies, which signifies that the connectivity differences are biological interpretability. In contrast, a separate estimation of networks using penalized approaches identified negligible or limited connectivity differences between varying modes of mental effort that seem biologically implausible. In addition, the joint estimation of multiple networks under a penalized approach is not naturally conducive for comparing networks and hence one had to use computationally prohibitive permutation tests that tend to give suboptimal results in terms of network accuracy and inferring between-network differences.

In this article, we demonstrated BJNL for estimating networks using fMRI data because they are the most prevalent type of functional images. However, our method can also be generalized to data from other imaging modalities in a straightforward manner. One advantage of our proposed approach for clustering the edge weights is that it allows for unsupervised estimation of the number of clusters. This means that in generalizing the method to other modalities, we do not have to laboriously tune the clustering parameters to each individual problem. Going beyond multiple experimental conditions, our approach can also be used to jointly model networks across multiple cohorts, such as healthy individuals, subjects with mild cognitive disorder, and those with Alzheimer's disease (Kundu et al. 2019). Future work should investigate the scalability of BJNL to larger numbers of conditions while taking into account the dynamic nature of the brain networks over time.

Supplementary Materials

The supplementary materials contain the detailed posterior computation steps, description of the network metrics used for Stroop task analysis, the results of the 40 node simulations, additional boxplots for performance metrics for simulations, additional details on the Stroop Task data analysis, and a Matlab GUI to implement the method.

Funding

References

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152–1174. [4]

Barabási, A.-L., and Albert, R. (1999), "Emergence of Scaling in Random Networks," *Science*, 286, 509–512. [6]

Belilovsky, E., Varoquaux, G., and Blaschko, M. B. (2016), "Testing for Differences in Gaussian Graphical Models: Applications to Brain Connectivity," in *Advances in Neural Information Processing Systems*, pp. 595–603. [2]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [6]

Canale, A., and Dunson, D. B. (2011), "Bayesian Kernel Mixtures for Counts," *Journal of the American Statistical Association*, 106, 1528–1539. [5]

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010), "The Horseshoe Estimator for Sparse Signals," *Biometrika*, 97, 465–480. [2,6]

Cox, R. W. (1996), "AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages," *Computers and Biomedical Research*, 29, 162–173. [9]

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, 76, 373–397. [2,3,6,7]

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431. [9]

Duchek, J. M., Balota, D. A., Thomas, J. B., Snyder, A. Z., Rich, P., Benzinger, T. L., Fagan, A. M., Holtzman, D. M., Morris, J. C., and Ances, B. M. (2013), "Relationship Between Stroop Performance and Resting State Functional Connectivity in Cognitively Normal Older Adults," *Neuropsychology*, 27, 516–528. [1]

Durante, D., Dunson, D. B., and Vogelstein, J. T. (2017), "Nonparametric Bayes Modeling of Populations of Networks," *Journal of the American Statistical Association*, 112, 1516–1530. [5]

Fox, M. D., Snyder, A. Z., Vincent, J. L., and Raichle, M. E. (2007), "Intrinsic Fluctuations Within Cortical Systems Account for Intertrial Variability in Human Behavior," *Neuron*, 56, 171–184. [1]

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse Inverse Covariance Estimation With the Graphical Lasso," *Biostatistics*, 9, 432–441. [2,6]

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889. [4]

Ghosh, J., and Dunson, D. B. (2009), "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320. [5]

Gianaros, P. J., Derbtshire, S. W., May, J. C., Siegle, G. J., Gamalo, M. A. and Jennings, J. R. (2005), "Anterior Cingulate Activity Correlates With Blood Pressure During Stress," *Psychophysiology*, 42, 627–635. [3]

Gießing, C., Thiel, C. M., Alexander-Bloch, A. F., Patel, A. X., and Bullmore, E. T. (2013), "Human Brain Functional Network Changes Associated With Enhanced and Impaired Attentional Task Performance," *Journal of Neuroscience*, 33, 5903–5914. [3,11]

Gill, J., and Casella, G. (2009), "Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation," *Journal of the American Statistical Association*, 104, 453–454. [5]

Gruber, S. A., Rogowska, J., Holcomb, P., Soraci, S., and Yurgelun-Todd, D. (2002), "Stroop Performance in Normal Control Subjects: An fMRI Study," *Neuroimage*, 16, 349–360. [1]

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011), "Joint Estimation of Multiple Graphical Models," *Biometrika*, 98, 1–15. [2]

Higgins, I. A., Kundu, S., Choi, K. S., Mayberg, H. S., and Guo, Y. (2019), "A Difference Degree Test for Comparing Brain Networks," *Human Brain Mapping*, 40, 4518–4536. [2]

Jara, A., García-Zattera, M. J., and Lesaffre, E. (2007), "A Dirichlet Process Mixture Model for the Analysis of Correlated Binary Responses," *Computational Statistics & Data Analysis*, 51, 5402–5415. [5]

Khachouf, O. T., Chen, G., Duzzi, D., Porro, C. A., and Pagnoni, G. (2017), "Voluntary Modulation of Mental Effort Investment: An fMRI Study," *Scientific Reports*, 7, 17191. [1,3,4,10,11]

Kim, J., Pan, W., and Alzheimer's Disease Neuroimaging Initiative (2015), "Highly Adaptive Tests for Group Differences in Brain Functional Connectivity," *NeuroImage: Clinical*, 9, 625–639. [1,2]

Kottas, A., Müller, P., and Quintana, F. (2005), "Nonparametric Bayesian Modeling for Multivariate Ordinal Data," *Journal of Computational and Graphical Statistics*, 14, 610–625. [5]

Kundu, S., Lukemire, J., Wang, Y., and Guo, Y. (2019), "A Novel Joint Brain Network Analysis Using Longitudinal Alzheimer's Disease Data," *Scientific Reports*, 9, 1–18. [12]

Lahat, D., Adah, T., and Jutten, C. (2015), "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," *Proceedings of the IEEE*, 103, 1449–1477. [2]

Levinson, O., Hershey, A., Farah, R., and Horowitz-Kraus, T. (2018), "Altered Functional Connectivity of the Executive Functions Network During a Stroop Task in Children With Reading Difficulties," *Brain Connectivity*, 8, 516–525. [1]

Li, Y., Craig, B. A., and Bhadra, A. (2017), "The Graphical Horseshoe Estimator for Inverse Covariance Matrices," arXiv no. 1707.06661. [2,6]

Lin, Z., Wang, T., Yang, C., and Zhao, H. (2017), "On Joint Estimation of Gaussian Graphical Models for Spatial and Temporal Data," *Biometrics*, 73, 769–779. [2]

MacLeod, C. M. (1991), "Half a Century of Research on the Stroop Effect: An Integrative Review," *Psychological Bulletin*, 109, 163. [1]

Müller, P., Erkanli, A., and West, M. (1996), "Bayesian Curve Fitting Using Multivariate Normal Mixtures," *Biometrika*, 83, 67–79. [2]

Mumford, J. A., and Ramsey, J. D. (2014), "Bayesian Networks for fMRI: A Primer," *Neuroimage*, 86, 573–582. [2]

O'brien, S. M., and Dunson, D. B. (2004), "Bayesian Multivariate Logistic Regression," *Biometrics*, 60, 739–746. [5]

Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., and Gore, J. C. (1999), "An fMRI Study of Stroop Word-Color Interference: Evidence for Cingulate Subregions Subserving Multiple Distributed Attentional Systems," *Biological Psychiatry*, 45, 1237–1258. [1]

Peterson, C., Stingo, F., and Vannucci, M. (2015), "Bayesian Inference of Multiple Gaussian Graphical Models," *Journal of the American Statistical Association*, 110, 159–174. [2,6]

Piironen, J., and Vehtari, A. (2017), "Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors," *Electronic Journal of Statistics*, 11, 5018–5051. [2]

Polson, N. G., and Scott, J. G. (2010), "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction," *Bayesian Statistics*, 9, 501–538. [2]

Polson, N. G., Scott, J. G., and Windle, J. (2013), "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349. [5]

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016), "Joint Estimation of Multiple Graphical Models From High Dimensional Time Series," *Journal of the Royal Statistical Society*, Series B, 78, 487–504. [2]

Rubinov, M., and Sporns, O. (2010), "Complex Network Measures of Brain Connectivity: Uses and Interpretations," *Neuroimage*, 52, 1059–1069. [10]

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650. [5]

Shan, Z. Y., Finegan, K., Bhuta, S., Ireland, T., Staines, D. R., Marshall-Gradisnik, S. M., and Barnden, L. R. (2018), "Brain Function Characteristics of Chronic Fatigue Syndrome: A Task fMRI Study," *NeuroImage: Clinical*, 19, 279–286. [1]

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., and Beckmann, C. F. (2009), "Correspondence of the Brain's Functional Architecture During Activation and Rest," *Proceedings of the National Academy of Sciences of the United States of America*, 106, 13040–13045. [9]

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., and Woolrich, M. W. (2011), "Network Modelling Methods for fMRI," *Neuroimage*, 54, 875–891. [1]

Stroop, J. R. (1935), "Studies of Interference in Serial Verbal Reactions," *Journal of Experimental Psychology*, 18, 643. [1,3]

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, 273–289. [9]

Walker, S. G. (2007), "Sampling the Dirichlet Mixture Model With Slices," *Communications in Statistics—Simulation and Computation*, 36, 45–54. [5]

Wang, H. (2012), "Bayesian Graphical Lasso Models and Efficient Posterior Computation," *Bayesian Analysis*, 7, 867–886. [4]

Wang, L., Zhang, Y., Lin, X., Zhou, H., Du, X., and Dong, G. (2018), "Group Independent Component Analysis Reveals Alternation of Right Executive Control Network in Internet Gaming Disorder," *CNS Spectrums*, 23, 300–310. [1]

Watts, D. J., and Strogatz, S. H. (1998), "Collective Dynamics of 'Small-World' Networks," *Nature*, 393, 440–442. [6]

Woodward, T. S., Leong, K., Sanford, N., Tipper, C. M., and Lavigne, K. M. (2016), "Altered Balance of Functional Brain Networks in Schizophrenia," *Psychiatry Research: Neuroimaging*, 248, 94–104. [1]

Yajima, M., Telesca, D., Ji, Y., and Muller, P. (2012), "Differential Patterns of Interaction and Gaussian Graphical Models." [2]

Yu, H., and Dauwels, J. (2016), "Variational Bayes Learning of Time-Varying Graphical Models," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, pp. 1–6. [2]

Yuan, M., and Lin, Y. (2007), "Model Selection and Estimation in the Gaussian Graphical Model," *Biometrika*, 94, 19–35. [7]

Zhu, Y., Shen, X., and Pan, W. (2014), "Structural Pursuit Over Multiple Undirected Graphs," *Journal of the American Statistical Association*, 109, 1683–1696. [2]