Regression with Structured Features at Multiple Scales to the Study of General Cognition in Children

Rene Gutierrez*‡ Rajarshi Guhaniyogi*§ Aaron Scheffler*¶

May 5, 2024

Abstract

This article is motivated by an application, where we aim to comprehend the neural underpinnings of general cognition, a pivotal indicator of healthy brain development, by examining the relationship between structural task-based brain activation maps and resting-state brain connectivity graphs in children aged 9-10 years old. While prior studies have identified certain brain regions linked to general cognition, these findings predominantly rely on analyses focusing on a single image modality, such as the resting-state graph alone. Moreover, no structured regression technique currently exists to assess the collective impact of both structural and graph features on general cognition while preserving linkage between their topology. To address this gap, this article focuses on developing a regression model with a scalar outcome and two sets of imaging features obtained at different scales: (a) a *graph*-valued feature with "labelled" nodes at a coarse scale, quantifying interconnections between nodes in the form of a brain connectome graph from resting state functional magnetic resonance imaging (fMRI); and (b) *structural* features at a finer scale *nested* within each graph node in the form of task-based brain activation maps. We introduce a novel flexible Bayesian regression framework that harnesses the relational information of nodes in the graph-valued

‡Rene Gutierrez, Postdoctoral Researcher, Department of Statistics, Texas A & M, TAMU 3143, College Station, Texas 77843-3143 (E-mail: rgutie17@ucsc.edu).

§Rajarshi Guhaniyogi, Associate Professor, Department of Statistics, Texas A & M University, TAMU 3143, College Station, Texas 77843-3143 (E-mail: rajguhaniyogi@tamu.edu).

¶Aaron Scheffler, Assistant Professor, Department of Epidemiology and Biostatistics, UC San Francisco, 550 16th Street, San Francisco, CA 94158 (E-mail: Aaron.Scheffler@ucsf.edu).

feature and the nested architecture between graph and structural features through a novel joint prior structure on coefficients. We refer to the proposed framework as Bayesian Multi-Object Feature Regression (BMFR). The framework enables inference on significant nodes in the graph predictive of the outcome, coefficients for features at both scales, and predictive inference for the outcome, each accompanied by precise characterization of uncertainty. The implementation utilizes an efficient Markov Chain Monte Carlo algorithm. Results from simulations showcase the framework's excellent performance in terms of influential node inference, regression coefficient estimation, and outcome prediction, outperforming popular competitors such as high-dimensional regression approaches, tree-based models, and deep neural networks. Application of BMFR to the multi-modal imaging data identifies two parieto-frontal resting state networks and constituent structural regions activated during a working memory task that provide new evidence to support existing theories of neuronal integration.

Key Words: Bayesian inference; brain connectome; functional magnetic resonance imaging; graph features; multi-modal imaging; structural features; spike and slab prior.

1 Introduction

Aided by technological advances, the last decade has witnessed an extraordinary growth in the collection of structured data from multiple modalities to investigate a scientific question of importance in neuroscience (Sui *et al.*, 2012; Hagler *et al.*, 2019; Wakeman and Henson, 2015; Zhang and Shi, 2020), biology (Graw *et al.*, 2021; Kang *et al.*, 2022), and social sciences (Guhaniyogi and Rodriguez, 2020; Zhang *et al.*, 2022). Viewing the collection of data from each modality as multiple objects, the successful integration of multi-object data produces a sum of information greater than the individual parts, but this integration can be difficult due to the complexity induced through different topological structure of the objects. This is the case in our motivating multi-modal brain imaging study which seeks to understand the neural underpinnings of general cognition in children aged 9-10 years old as a function of structural task-based brain activation maps and resting state brain connectivity graphs. General cognition, also called the g-factor, is considered a critical transdiagnostic domain that informs diagnosis for several developmental and mental disorders including de-

pression, attention-deficit/hyperactivity disorder, and schizophrenia (Morris and Cuthbert, 2012). Despite the importance of the g-factor in characterizing cognitive disorders, the association between general cognition and both brain structure and function is still not well understood (Pat *et al.*, 2022) due to the challenge of integrating information from multiple brain images which fundamentally exist at different scales as detailed below.

Most models have focused on modeling association between general cognition and brain structure or function via a single brain imaging modality at a time and implicate the parieto-frontal pathways as fundamental in explaining variation among individuals (Deary *et al.*, 2010). Specifically, association of general cognition has been studied separately with grey matter content via MRI (Narr *et al.*, 2007), resting state functional connectivity via fMRI (Dubois *et al.*, 2018; Pamplona *et al.*, 2015; Sripada *et al.*, 2020), and task-based activation via fMRI (Gray *et al.*, 2003; Waiter *et al.*, 2009). Recently, several works have begun to examine the relationship between general cognition and multiple brain imaging modalities simultaneously and have demonstrated improved prediction of general cognition as well as detection of informative interactions between imaging modalities (Jiang *et al.*, 2020; Rasero *et al.*, 2021; Pat *et al.*, 2022). Except for Zhao *et al.* (2023), most multimodal approaches have focused on prediction of general cognition from multiple brain images via vectorization of multimodal images (Jiang *et al.*, 2020; Sui *et al.*, 2020) or ad hoc model stacking (Rasero *et al.*, 2021; Pat *et al.*, 2022). Specifically, maintenance of the structured information shared between task-based fMRI and resting state connectivity graphs is vital to better understand how focal activation and resting state brain connectivity jointly describe human cognition (Thiele *et al.*, 2022; Cole *et al.*, 2016; Rabini *et al.*, 2023). Yet, there has been no exploration of this relationship utilizing a structured regression approach that investigates the connection between general cognition and both localized task-based brain activation and subject-specific connectivity graphs.

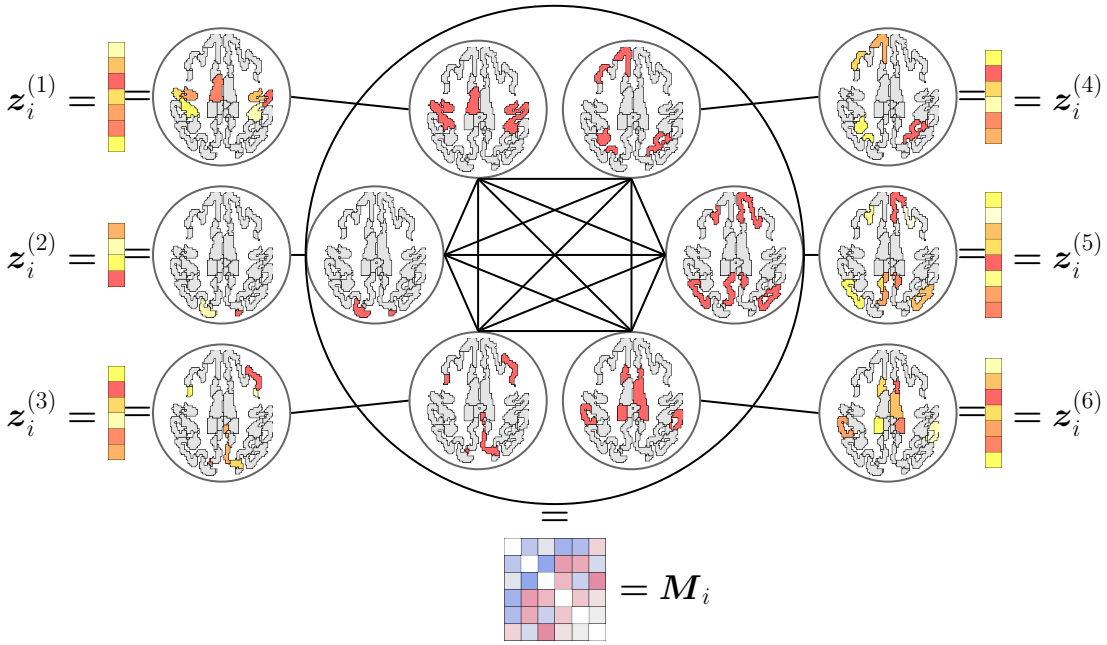
Motivated by this gap, this article focuses on developing a novel regression framework to predict a general cognition score from two types of features obtained at two scales: (a) a graph-valued feature over a set of common nodes quantifying interconnection between nodes at the coarse scale; and (b) a multivariate feature nested within each node of the graph at the fine scale, henceforth referred to as *structural features*. Specifically, for each

subject, at the coarse scale, a graph is constructed, representing the interconnections between functional networks in the brain. This graph is estimated using resting state functional magnetic resonance imaging (rs-fMRI). At the fine scale, structural information in the human brain is obtained over brain regions of interest (ROIs), using task-based fMRI (t-fMRI) data which measures focal task-based brain activation in each ROI. The graphical and structural images are collected and aligned such that spatially indexed ROIs are nested within nodes of predefined brain functional networks. **Figure 1** shows an illustration of the two sets of features with a toy example. In this figure, a graph is constructed over six nodes representing six functional networks. In each node of the graph, fine scale features over ROIs are observed. The goal of the proposed regression framework is to effectively model the relationship between the outcome variable and these two types of features. By incorporating information from both the graph-object and the nested structural features, the framework aims to capture the complex interplay between features at different scales to provide valuable insights into the regression relationship and nodes significantly related to the outcome, each with uncertainty.

Bayesian framework is a naturally suitable tool for modeling the linkage between features and the topology of individual features through careful construction of prior distributions. This framework also has the added advantage of incorporating uncertainty in inference and prediction, crucial for neuroimaging studies characterized by limited sample sizes, low signal-to-noise ratios, and high dimensionality. However, the endeavor to construct hierarchical Bayesian models with structured prior distributions is significantly limited with multi-object features. Particularly, there is a lack of prior work addressing the challenges with a graph-valued feature and structural features nested within each graph node.

To achieve this objective, we introduce a Bayesian multi-object feature regression (BMFR) framework. This framework constructs a novel joint prior distribution on graphical and structural coefficients, incorporating three constraints arising from the topological structure of graphs and the nested architecture of the features: the *nesting constraint*, *edge constraint*, and *transitivity effect* (refer to Section 2.2.2 for details). The joint prior formulation serves to complement and reinforce information from both structural and graphical features. This enhances the reliability and accuracy of inference on both sets of coefficients. Additionally, it facilitates the identification of influential nodes related to the outcome and enables

Figure 1: Data diagram: The figure shows a toy example to represent features at two different scales. It shows a graph-valued feature at the coarse scale over six nodes representing six functional networks. At each node, it shows a vector of fine-scale features, observed over ROIs nested within that particular node. In this picture the graph is represented by a symmetric matrix $\boldsymbol{M}_i$, and fine-scale feature vectors at six different nodes are represented by $z_i^{(1)}, ..., z_i^{(6)}$. The two sets of features are jointly regressed on the outcome.



accurate predictive inference for the outcome variable. BMFR demonstrates superior empirical performance compared to ordinary high-dimensional regressions, tree-based non-linear regression methods and deep neural network approaches.

Rest of the manuscript proceeds as follows. Section 1.1 briefly encapsulates our contribution, while a detailed comparisons of the existing literature with our proposed approach is provided in Section 1.2. Section 3 describes the novel prior framework to draw inference and prediction with the structural and graphical features, and Section 4 discusses posterior computation of the proposed model. Empirical investigations with data generated under various simulation settings are reported in Section 5. Section 6 analyzes the multi-modal imaging dataset, offering predictive inference and scientific findings on functional subnetworks in the brain predictive of a cognitive outcome. Finally, Section 7 summarizes the idea laid out in this article and highlights some of the extensions of our model to be explored in the near future.

1.1 Our Contribution

The main contribution of this article are listed as follows.

1. We present a innovative prior formulation for structural and graphical coefficients that reflect properties of our multi-modal brain imaging data. This proposed prior leverages information regarding the connectivity among graph nodes, as well as the topological characteristic that structural features are nested within a graph node, while conducting inference on the model. The linkage between the structural and graphical features, specifically the former being nested within the nodes of the graph, is enforced in the prior development through several constraints that reflect neuroscience domain knowledge, known as the edge constraint, nesting constraint, and transitivity effect. For more details on these constraints, please refer to Section **??**.

2. The proposed model and joint prior formulation achieve efficient computation and accurate predictive inference of the outcome. Additionally, the framework facilitates estimation of regression coefficients and identification of influential nodes. Furthermore, our proposed method incorporates uncertainty in identifying influential nodes and generates well-calibrated interval estimates for the structural and graphical co-

efficients. Previous works exist on Bayesian model-based identification of influential nodes, focusing solely on a graph-valued feature (Guha and Rodriguez, 2021, 2023). However, none of these prior methods focus on regression with graph-valued features and structural features at its nodes to deliver inference on graph nodes, as demonstrated in our approach. As a result, the prior structure enables novel questions to be asked and answered in the context of modeling general cognition as a function of multiple brain images.

3. BMFR is robust to the challenges of real data, and provide domain-relevant insights for real data analysis. In particular, BMFR facilitates the analysis and interpretation of multi-modal brain imaging data to better understand the neural underpinning of human cognition, reinforcing and expanding previous findings that link task-based activation and resting state connectivity in the parieto-frontal regions of the brain with fluid intelligence.

1.2 Related Work

In regressions involving object features, a common approach is to restructure the objects to fit them into an ordinary high-dimensional regression framework. As an example, in an object-data regression with a graph-valued feature, popular approaches often select a few summary measures from the graph (Bullmore and Sporns, 2009) or vectorize the graph object into a high-dimensional collection of edge weights (Craddock *et al.*, 2009; Richiardi *et al.*, 2011). Downstream inference is then drawn utilizing the latest developments in the ordinary high-dimensional regression architecture (Tibshirani, 1996; Park and Casella, 2008; Carvalho *et al.*, 2010), tree-based regression approaches (Sparapani *et al.*, 2021; Denison *et al.*, 1998) or deep neural network models (Polson and Sokolov, 2018; Dinh and Ho, 2020). However, restructuring the features in this manner may not adequately capture either the topology of the graph-valued feature, or the nested architecture of graph-valued and structural features to draw downstream inference, potentially compromising the inferential accuracy and interpretability of the regression model. Additionally, there is a growing field of study in graph representation learning, which concentrates on effectively encoding high-dimensional sparse graph-structured data into low-dimensional dense vectors (Choi *et al.*, 2017; Chen *et al.*,

2020). While much of this research is focused on unsupervised learning, it is feasible to regress the low-dimensional features derived from a graph against an outcome. However, it is important to note that after the extraction of low-dimensional features, it is not clear how to estimate influential nodes that are significantly linked to the outcome while also providing uncertainty estimates.

Of late, there are developments of scalar-on-object regression approaches exploiting the special structure of the object features. This include regression with functional features (Goldsmith *et al.*, 2014; Li *et al.*, 2015; Feng *et al.*, 2019; Kang *et al.*, 2018; Huang *et al.*, 2013), a tensor-valued feature (Zhou *et al.*, 2013; Guhaniyogi *et al.*, 2017; Guhaniyogi, 2017; Fan *et al.*, 2019) and a graph-valued feature (Guha and Rodriguez, 2021, 2023). These methods establish the importance of exploiting the topology of object-type features for better inference and prediction, but the referenced works mainly concentrate on developing regressions with a single object feature. Direct application of these approaches is inadequate in our setting since model and prior development should ideally consider the nested structure of the structural features within the nodes of the graph feature. By neglecting the structure of an object feature and the cross information between two types of features, model-based inference with multi-source data can suffer from various drawbacks, including lower detection power and sensitivity of results to noise (Calhoun and Sui, 2016).

There is an emerging literature on regression with multi-object features and a scalar outcome, mostly motivated from multi-modal imaging data applications, which is distinct from our approach in terms of modeling and inferential objectives. To this end, Zhao *et al.* (2023) model general fluid intelligence in children from multiple structurally indexed images using a multi-resolution approach that incorporates spatial information to capture main effects from individual images and interactions among images to predict the outcome. In another vein, Xue *et al.* (2018) proposes regression on disease status on low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel based morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. In the same vein Li and Li (2021) develops a factor analysis-based linear regression model, and Dai and Li (2021) extends this framework to account for non-linear association between a scalar response and multi-modal predictors. There is also an active and emerging literature on imaging genetics (Peng *et al.*,

2016; Du *et al.*, 2018) where imaging features as well as genetic features are used to predict an outcome. While these regression approaches exploit the linkages among multiple object features, they do not specifically address our methodological needs to incorporate graphical features and structural features at graph nodes to draw simultaneous inference on coefficients, influential network nodes, and prediction as we aim to do in our study. Furthermore, our proposed approach offers uncertainty estimation in all these inference, which is particularly crucial in clinical applications where sample sizes are often limited.

2 General cognition in children: case study

This article considers a clinical application derived from multimodal imaging studies conducted in children aged 9-10 years from the Adolescent Brain Cognitive Development (ABCD) study, the largest study on brain development in children and adolescents in the United States (Casey *et al.*, 2018). Our interest lies in characterizing the association between general cognition (Flynn, 2007; Pat *et al.*, 2022), a measure of cognitive ability that spans tasks such as language, memory, and mental flexibility, and multiple brain images as described in Section 1. Specifically, we consider structural and graph images of focal task-based functional activity and resting state connectivity, respectively.

Clinical images and language evaluation:

We utilized the ABCD Study 5.0 Tabulated Release Data (`http://dx.doi.org/10.15154/8873-zj65`) focusing specifically on baseline imaging and cognitive measures data collected on a random subsample of 395 children aged 9-10 years old collected form a single study site. Imaging data is collected and processed from the following imaging modalities as described thoroughly in (Hagler *et al.*, 2019): task-based fMRI (t-fMRI) data which measures focal task-based brain activation; and task-free resting state fMRI (rs-fMRI) to measure brain activation via neuronal oxygen consumption in subjects at rest.

All images are registered to the Montreal Neurological Institute (MNI) template space. Working memory was measured using t-fMRI activation (**Figure 1a**) from the 'N-back' task via the 2-back vs 0-back linear contrast as described in Hagler *et al.* (2019) for 145 spatially-indexed ROIs defined by the Destrieux atlas (Destrieux *et al.*, 2010). Resting state functional fMRI data was collected for 333 cortical-surface ROIs defined by the Gordon atlas (Gordon

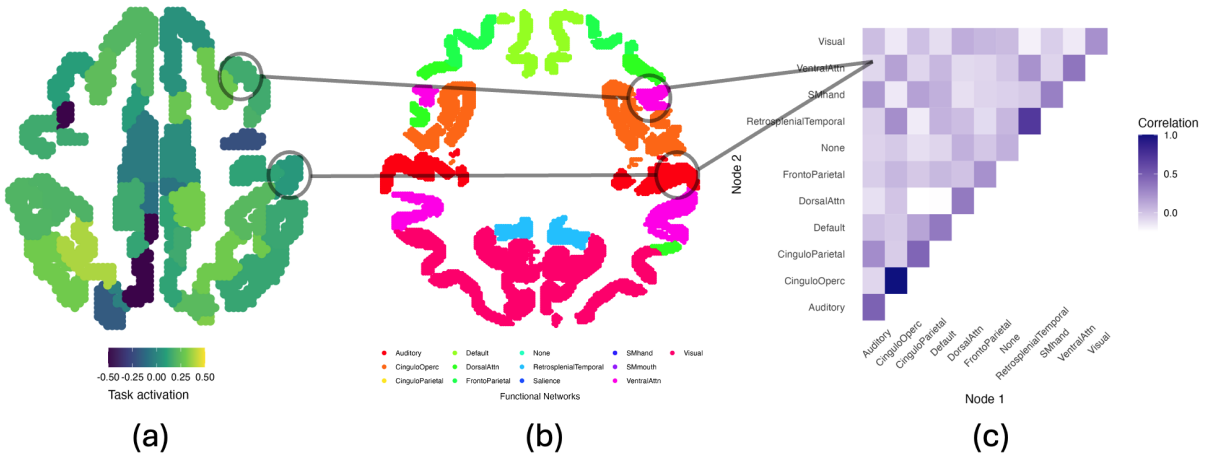


Figure 2: Schematic of the multi-object brain imaging data structure for a sample subject. (a) Structural image encoding ROI-level task activation, (b) Gordon functional atlas parcellation of the brain into functional networks, (c) Graph image obtained by calculating the pairwise Pearson correlation Z-score for the average within and between networks. Grey circles and lines connect (a) structural and (c) graphical information from images via the (b) parcellated atlas. Thus, the atlas provides an organizing hierarchy that links together structural information (task activation) at the ROI-level with graphical information indexed by pairs of functional networks.

et al., 2017) and ROIs were subsequently grouped into 11 predefined functional networks according to Gordon *et al.* (2017) (*auditory, cingulo-opercular, cingulo parietal, default mode, dorsal attention, frontoparietal, none, retrosplenial temporal, salience, sensorimotor-hand, sensorimotor-mouth, ventral attention* and *visual networks*). A symmetric adjacency matrix is obtained as in Hagler *et al.* (2019) by considering rows and columns of this matrix corresponding to different networks and entries corresponding to the Z-scores obtained by Fisher Z-transforming the Pearson correlation the average correlation of each unique pairwise combination of ROIs in one network versus another (**Figure 1c**). Given the lack of spatial alignment between the Destrieux and Gordon atlases, each ROI in the Destrieux atlas was nested within the predefined Gordon subnetworks by minimizing the Euclidean distance between the centroid of a given Destrieux ROI and corresponding centroids for the Gordon atlas ROIs such that images across modalities and subjects can be directly compared and each Destrieux ROI is nested in functionally defined Gordon subnetwork **(Figure 1b)**. The choice of brain atlas is an important decision in the analysis pipeline as it impacts subsequent results and interpretations.

We focus our analysis of general cognition, a measure of general intelligencce. Using a

Bayesian principal components analysis with varimax rotation on the ABCD Study data, Thompson *et al.* (2019) determined that general cognition as a latent factor is best measured by three items from the ABCD NIH Toolbox of cognition measures: picture vocabulary item, list sort working memory item, and oral reading test item. Therefore, we take an average of the items via their raw measures standardized to mean zero and standard deviation one to form a measure of general cognition. One of our primary goals is to extend the study of general cognition to consider multi-modal image data spanning the whole brain to better understand how structural and graphical changes simultaneously impact general cognition.

We apply our proposed regression framework to regress general cognition scores on sophisticated multi-modal images, specifically task-based activation maps which capture local brain activity during a working memory task, and fMRI brain connectivity graphs which capture resting state brain connectivity among a priori functional networks. Importantly, the ROI-level structural imaging features are nested within the graph nodes represented by functional networks, leading to regression of a scalar outcome with features at multiple scales. The next section describes the regression framework to achieve these scientific goals.

3 Bayesian Regression with Graphical and Structural Features at Multiple Scales

This section details out the model development and prior formulation, including the hyperparameter specification.

3.1 Model Development

For the ith subject, we observe a continuous scalar response $y_i \in \mathbb{R}$, a vector of ordinary unstructured features $\boldsymbol{x}_i = (x_{i,1}, ..., x_{i,P})^T$ and two sets of structured features at two scales: (a) a graph-valued feature at a coarse scale; and (b) features at a fine scale in every node of the graph. Let the graph-valued feature $\mathcal{G}_i$, for all subjects, be defined on the same set of labelled nodes/vertices $\mathcal{N} = \{\mathcal{N}_1, ..., \mathcal{N}_V\}$, with the number of nodes $|\mathcal{N}| = V$. The graph-valued feature $\mathcal{G}_i$ is expressed in the form of a $V \times V$ adjacency matrix $\boldsymbol{M}_i \in \mathbb{R}^{V \times V}$ over the common set of nodes $\mathcal{N}$, with the (v, v')-th entry $m_{i,[vv']}$ signifying the strength of association between the nodes $\mathcal{N}_v$ and $\mathcal{N}_{v'}$, where $v, v' = 1, ..., V$. This paper specifically

focuses on graphs that contain no self relationship, i.e., $m_{i,[vv]} \equiv 0$, and are undirected ($m_{i,[vv']} = m_{i,[v'v]}$). Additionally, $\boldsymbol{z}_i^{(1)}, ..., \boldsymbol{z}_i^{(V)}$ are features of dimensions $N_1, ..., N_V$, observed at the V graph nodes $\mathcal{N}_1, ..., \mathcal{N}_V$, respectively. These features observed at a fine scale are referred to as *structural features* throughout the article. Exploiting the *nested architecture* between graph and structural features is instrumental for drawing meaningful inference and efficient computation as detailed in the upcoming methodological development.

For $i = 1, .., n$, we propose a regression relationship between the outcome, structural and graphical features after accounting for ordinary unstructured features as below,

$$y_i = \beta_0 + \boldsymbol{\beta}_{\boldsymbol{x}}^T \boldsymbol{x}_i + \sum_{v=1}^{V} \boldsymbol{\beta}_v^T \boldsymbol{z}_i^{(v)} + \langle \boldsymbol{M}_i, \boldsymbol{\Lambda} \rangle / 2 + \epsilon_i, \;\; \epsilon_i \sim N(0, \tau^2), \tag{1}$$

where $\boldsymbol{\beta}_{\boldsymbol{x}}$ represents coefficients of the ordinary unstructured features in $\mathbb{R}^P$, and τ^2 is the variance of the observational error. Here $\boldsymbol{\beta}_v$ is the coefficient of dimension $N_v \times 1$ corresponding to the structural features observed at a fine scale in the node $\mathcal{N}_v$, and $\langle \cdot, \cdot \rangle$ is the Frobenius inner product between two matrices. Similar to $\boldsymbol{M}_i$, the $V \times V$ coefficient matrix $\boldsymbol{\Lambda} = ((\lambda_{[vv']}))$ is assumed to be symmetric with zero diagonal entries, so that $\langle \boldsymbol{M}_i, \boldsymbol{\Lambda} \rangle = 2 \sum_{1 \le v < v' \le V} m_{i,[vv']} \lambda_{[vv']}$. Such a simplification is useful in drawing connection between the multi-modal regression model (1) to a linear regression framework given by,

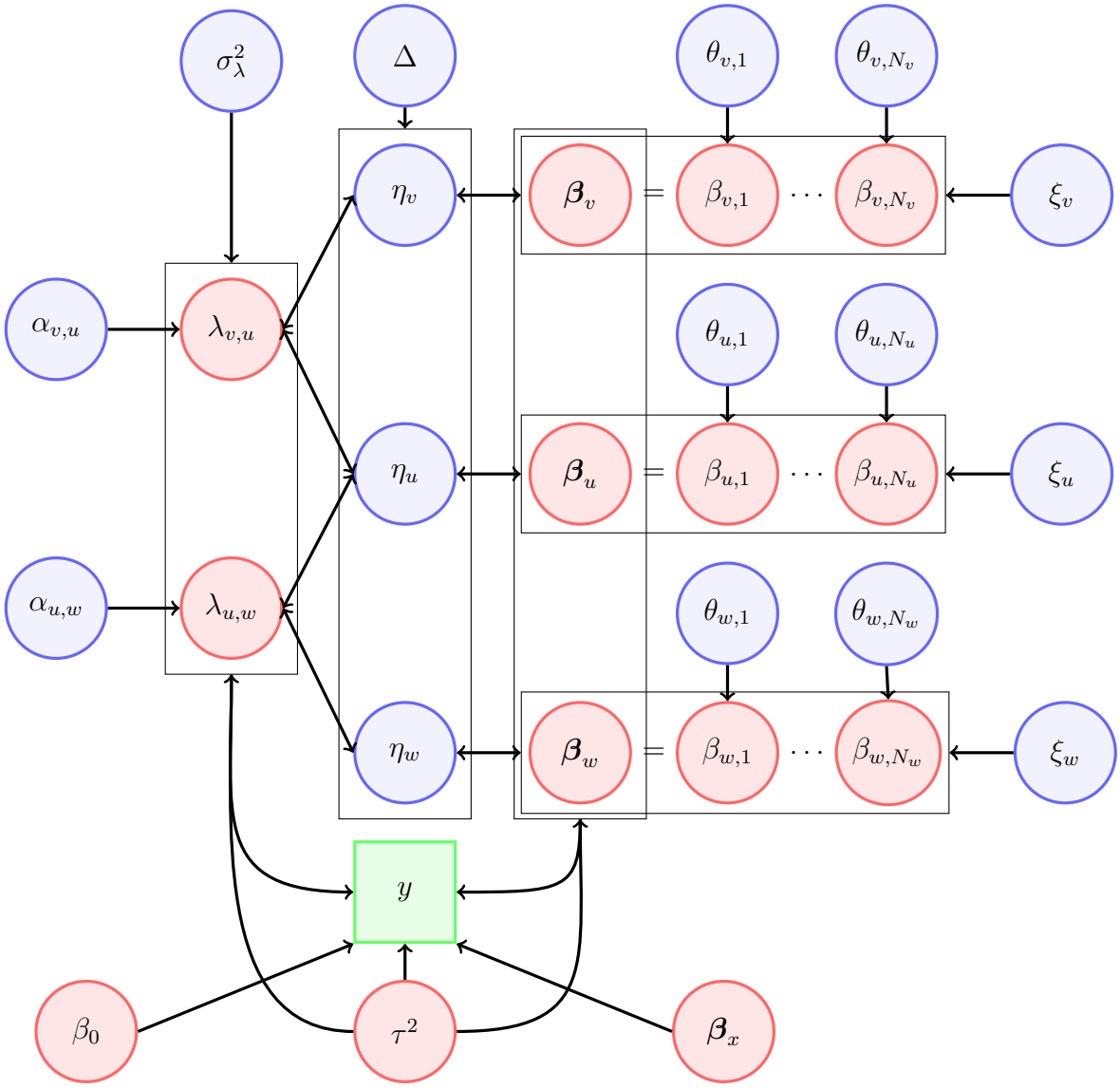
$$y_i = \beta_0 + \boldsymbol{\beta}_{\boldsymbol{x}}^T \boldsymbol{x}_i + \sum_{v=1}^{V} \boldsymbol{\beta}_v^T \boldsymbol{z}_i^{(v)} + \sum_{1 \le v < v' \le V} m_{i,[vv']} \lambda_{[vv']} + \epsilon_i, \;\; \epsilon_i \sim N(0, \tau^2). \tag{2}$$

(2) allows linking the information on node level data from the two types of features to draw inference on nodes influential in predicting the outcome through prior construction on model coefficients as described in the next section.

3.2 Prior Structure

This section introduces a novel joint prior for the structural and graphical coefficients, considering the nested architecture between the two sets of features at different scales. The joint prior construction on graphical and structural coefficients is aimed to achieve several objectives simultaneously: (a) inference on nodes $\mathcal{N}$ significantly related to the outcome; (b)

Figure 3: Plate diagram of the hierarchical model for Bayesian Multi-object Feature Regression (BMFR). The picture demonstrates hierarchical model with three graph nodes $\mathcal{N}_v$, $\mathcal{N}_u$ and $\mathcal{N}_w$. The three nodes have activation indicators η_v, η_u and η_w, respectively. The structural coefficient vectors corresponding to the three nodes are given by $\boldsymbol{\beta}_v$, $\boldsymbol{\beta}_u$ and $\boldsymbol{\beta}_w$, respectively.



inference on structural and graphical coefficients; (c) predictive inference of the outcome; (d) exploiting the nested architecture between the structural and graphical coefficients; and (e) guaranteeing efficient computation of the posterior for the proposed prior. We set out to accomplish (a)-(e) by envisioning (2) as a high-dimensional regression problem and formulate structured variable selection prior distributions on $\{\boldsymbol{\beta}_v : v = 1, ..., V\}$ and $\boldsymbol{\Lambda}$ after accounting for a few science-driven constraints imposed by the nested architecture of the features at two scales.

1. *Edge constraint*: If either node $\mathcal{N}_v$ or node $\mathcal{N}_{v'}$ is not influential in predicting the response, the corresponding edge coefficient $\lambda_{[vv']}$ that represents the edge between $\mathcal{N}_v$ and $\mathcal{N}_{v'}$ is deemed unimportant in predicting the outcome. This constraint ensures that only edges between influential nodes contribute significantly to the prediction of the outcome and focuses the analysis of general cognition on connectivity among distributed brain functional networks (Dubois *et al.*, 2018).

2. *Nesting constraint*: In line with the nested architecture of the structural and graphical coefficients, if node $\mathcal{N}_v$ is not influential in predicting the outcome, the structural features $\boldsymbol{z}^{(v)}$ observed at a fine-scale within node $\mathcal{N}_v$ do not have any effect on the prediction of the response. This constraint respects the hierarchical relationship between the graphical and structural features, ensuring that fine-scale features at the non-influential nodes do not contribute to the prediction. The nesting constraint reflects the assumption that human cognition can be understood by leveraging the the interplay of focal task-based activation and resting state brain connectivity (Thiele *et al.*, 2022; Cole *et al.*, 2016; Rabini *et al.*, 2023).

3. *Transitivity effects*: If both the edge between $\mathcal{N}_v$ and $\mathcal{N}_{v'}$ and the edge between $\mathcal{N}_{v'}$ and $\mathcal{N}_{v''}$ are influential in predicting the outcome, the edge between $\mathcal{N}_v$ and $\mathcal{N}_{v''}$ is likely to be influential. This constraint captures the transitivity effects within the graph, where the influence of an edge between two nodes extends to the indirect connection between other nodes (Hoff, 2005). It helps identify influential edges based on the presence of influential neighboring edges. The property of transitivity reflects the assumption of small-world network structure which is a popular model for neural organization that

achieves the simultaneous demands of network integration and segregation (Rubinov and Sporns, 2010; Bassett and Bullmore, 2006; Pamplona *et al.*, 2015).

The constraints introduce additional structure and dependencies among the coefficients that cannot be captured by an ordinary variable selection prior (Carvalho *et al.*, 2010; Park and Casella, 2008) on $\boldsymbol{\beta}_v$ and $\lambda_{[vv']}$ to draw inference on influential nodes in $\mathcal{N}$.

We propose a prior construction that exploits ideas from both discrete and continuous mixture variable selection priors to address our inferential goals (a)-(e) while incorporating the constraints 1-3. To achieve this, we introduce binary indicator variables $\eta_1, ..., \eta_V \in \{0, 1\}$ that signify the importance of the V nodes in predicting the outcome. Specifically, $\eta_v = 0$ indicates that the vth node $\mathcal{N}_v$ has no effect on the response from all features. The graph edge coefficient $\lambda_{[vv']}$ is then endowed with a variable selection prior given by

$$\lambda_{[vv']}|\alpha_{[vv']}, \sigma_\lambda, \tau \overset{ind.}{\sim} \begin{cases} N(0, \tau^2\sigma_\lambda^2\alpha_{[vv']}^2), & \text{if } \eta_v = \eta_{v'} = 1 \\ \delta_0, & \text{o.w.,} \end{cases} \qquad \sigma_\lambda \sim C^+(0,1),\ \alpha_{[vv']} \overset{ind.}{\sim} C^+(0,1), \tag{3}$$

where δ_0 corresponds to the Dirac-delta function, $\alpha_{[vv']}$ is the local parameter corresponding to the edge between $\mathcal{N}_v$ and $\mathcal{N}_{v'}$, σ_λ is the global parameter for the graph coefficient, and $C^+(0,1)$ denotes a half-Cauchy distribution. The prior construction in (3) satisfies the *edge constraint* by enforcing $\lambda_{[vv']} = 0$ when either $\eta_v = 0$ or $\eta_{v'} = 0$. The formulation by definition imposes *transitivity effects* in the network coefficient $\boldsymbol{\Lambda}$. Marginalizing over σ_λ and $\alpha_{[vv']}$ in (3) yields a mixture distribution for $\lambda_{[vv']}|\tau$ with mixture components as Dirac-delta function and the popular horseshoe prior (Carvalho *et al.*, 2010). The construction offers a flexible prior structure for precise estimation of nonzero graph edge coefficients a posteriori.

The structural coefficient vector $\boldsymbol{\beta}_v = (\beta_{v,1}, ..., \beta_{v,N_v})^T \in \mathbb{R}^{N_v}$ corresponding to the node $\mathcal{N}_v$ is modeled using,

$$\beta_{v,j}|\theta_{v,j}, \xi_v, \tau \overset{ind.}{\sim} \begin{cases} N(0, \tau^2\xi_v^2\theta_{v,j}^2), & \text{if } \eta_v = 1 \\ \delta_0, & \text{o.w.,} \end{cases} \qquad \theta_{v,j} \overset{i.i.d.}{\sim} C^+(0,1),\ \xi_v \overset{i.i.d.}{\sim} C^+(0,1), \tag{4}$$

for $j = 1, ..., N_v;\ v = 1, ..., V$. The construction ensures $\boldsymbol{\beta}_v = 0$ when $\eta_v = 0$, enforcing the

nesting constraint. Additionally, (4) induces approximate sparsity in structural coefficients by shrinking the less influential components toward zero while retaining the true signals (Polson and Scott, 2010) using the horseshoe shrinkage mechanism. Finally, the binary inclusion indicators are assigned Bernoulli prior distribution $\eta_v \overset{i.i.d.}{\sim} Ber(\Delta)$ with $\Delta \sim Beta(a_\Delta, b_\Delta)$ to account for multiplicity correction (Scott and Berger, 2010). Notably, an estimate of the posterior probability of the event $\{\eta_v = 1\}$ shows the uncertainty in identifying $\mathcal{N}_v$ to be influential. A posterior probability close to 1 or 0 indicates strong evidence in favor of or against the influence of node $\mathcal{N}_v$ in predicting the response. The prior specification is completed by assigning a normal prior on coefficients of $\boldsymbol{\beta}_x$, β_0 and $IG(a_\tau, b_\tau)$ on the error variance τ^2. A plate diagram of model and prior structure with only three graph nodes can be seen in Figure 3.

4 Posterior Computation

Although summaries of the posterior distribution cannot be computed in closed form, full conditional distributions for all the parameters are available and mostly correspond to standard families (available in Appendix). Thus, posterior computation can proceed through a Markov chain Monte Carlo algorithm. While a naive implementation of such an algorithm to jointly update $\{\boldsymbol{\beta}_v : v = 1, .., V\}$ and $(\lambda_{[vv']} : 1 \leq v < v' \leq V)^T$ is viable for small values of V and $N_1, ..., N_V$, it entails complexity of $\sim Q^3$, where $Q = \sum_{v=1}^{V} N_v + V(V-1)/2$, which may lead to intractable computation for moderately large values of V and $N_1, ..., N_V$. To address this issue, we follow the procedure outlined in Guha and Rodriguez (2021) which allows computation at $\sim n^3$ complexity. Since in biomedical applications, often n is much smaller than Q, this approach leads to substantial computational savings.

The MCMC sampler is run for 10000 iterations, with the first 5000 iterations discarded as burn-in. All posterior inference is based on post burn-in samples. The average effective sample size (ESS) as a fraction of the total post burn-in iterations averaged over all $\boldsymbol{\Lambda}$ and $\boldsymbol{\beta}_v$'s for each case show fairly uncorrelated MCMC samples (see Table 1).

Table 1: Average effective sample size as a fraction of post burn-in iterations by case for our proposed approach Bayesian Multi-Object Feature Regression (BMFR) and the Horseshoe competitor.

Method	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 5*	*Case 6*	*Case 7*
BMFR	**0.81**	**0.78**	**0.73**	**0.7**	**0.72**	**0.64**	**0.56**
	(0.098)	(0.078)	(0.068)	(0.064)	(0.072)	(0.054)	(0.051)
Horseshoe	0.4	0.37	0.34	0.33	0.32	0.27	0.26
	(0.063)	(0.055)	(0.052)	(0.042)	(0.044)	(0.051)	(0.045)
Method	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 5*	*Case 6*	*Case 7*
BMFR	**0.47**	**0.61**	**0.48**	**0.34**	**0.31**	**0.35**	**0.35**
	(0.064)	(0.049)	(0.049)	(0.034)	(0.033)	(0.035)	(0.033)
Horseshoe	0.29	0.25	0.27	0.32	0.31	0.32	0.32
	(0.045)	(0.043)	(0.049)	(0.03)	(0.02)	(0.026)	(0.026)

5 Simulation Studies

In this section we evaluate and compare inferential and out-of-sample predictive performance of the proposed BMFR approach to two sets of competitors, (a) high-dimensional linear and non-linear regressions; (b) object data regressions.

High-dimensional linear and non-linear regressions. Broadly, these competitors treat the edges between nodes in the undirected graph feature $\boldsymbol{M}_i$ as a "long vector of features" and regress response y_i on vectors $\boldsymbol{m}_i = (m_{i,[vv']} : 1 \leq v < v' \leq V)^T$ and $\boldsymbol{z}_i = (\boldsymbol{z}_i^{(1)T}, ..., \boldsymbol{z}_i^{(V)T})^T$, thereby ignoring the relational nature of $\boldsymbol{M}_i$. To this end, horseshoe prior (Carvalho *et al.*, 2010) is employed on the regression coefficients which we refer to as "Horseshoe" and show its convergence behavior in terms of ESS averaged over all parameters in Table 1. A frequentist high dimensional regression competitor has also been constructed by adopting a penalized optimization framework with the minimax concave penalty (MCP) on the feature coefficients (Zhang, 2010). MCP is implemented using the `ncvreg` (Breheny and Huang, 2011) package in R, with the penalty parameter of MCP chosen through ten-fold cross validation technique. Additionally, non-linear relationship between the outcome and vectorized high-dimensional feature is accommodated in the competitors by fitting Bayesian additive regression tree (BART) model and deep neural network (NN) model. BART and NN models are fitted using the `BART` and `neuralnet` packages in R, respectively, and we allow variable selection within the BART implementation. NN is fitted in two layers, with the size of the

first layer being the sum of number of structural features and the number of edges, given by $V(V-1)/2 + \sum_{v=1}^{V} N_v$. The size of the second layer is chosen to be half of the first layer. Horseshoe, MCP and BART are compared with BMFR in terms of inferential and predictive performance. In contrast, NN is employed only to assess predictive performance of BMFR.

Object data regressions. This approach fits a regression framework as in (1) with horseshoe prior on each component of $\boldsymbol{\beta}_v$ and Bayesian network lasso (BNL) prior (Guha and Rodriguez, 2021) on the coefficient $\boldsymbol{\Lambda}$. Unlike ordinary high-dimensional competitors, this approach accounts for graph topology of $\boldsymbol{M}_i$ in the regression, but construction of priors on $\boldsymbol{\beta}_v$ and $\boldsymbol{\Lambda}$ separately (as opposed to jointly in BMFR) does not adequately leverage interconnections between the two sets of features.

5.1 Data Generation

In all our simulations, we generate outcome from the following model,

$$y_i = \beta_{0,t} + \sum_{v=1}^{V} \boldsymbol{\beta}_{v,t}^T \boldsymbol{z}_i^{(v)} + \langle \boldsymbol{M}_i, \boldsymbol{\Lambda}_t \rangle / 2 + \epsilon_i, \;\; \epsilon_i \sim N(0, \tau_t^2), \tag{5}$$

where the subscript t indicates the true data generating parameters. We set the number of graph nodes to be equal to $V = 20$ and the sample size $n = 150$ in all simulations. The number of fine scale features nested in each graph node is taken to be equal, i.e., $N_1 = \cdots = N_V = N$ in all simulations. We present simulation cases by varying N, as discussed later.

Simulating true coefficients $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\beta}_{v,t}$. To simulate the true coefficients $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\beta}_{v,t}$, binary variables $\eta_{1,t}, ..., \eta_{V,t} \overset{i.i.d.}{\sim} Ber(\Delta_t)$ are generated, and $\eta_{v,t} = 1$ indicates that the v-th node is influential in predicting the outcome. Since $(1 - \Delta_t)$ is the probability of a region not being "influential," it is referred to as the node sparsity parameter. The coefficient corresponding to the edge connecting the v-th and v'-th node is drawn from the following distribution,

$$\lambda_{[vv'],t} \overset{ind.}{\sim} \begin{cases} N(1,1), & \text{if } \eta_{v,t} = \eta_{v',t} = 1 \\ \delta_0, & \text{o.w.,} \end{cases} \qquad \lambda_{[vv'],t} = \lambda_{[v'v],t}; \;\; v < v'. \tag{6}$$

Table 2: It presents the different simulation cases. Here Δ_t is the probability of a region being active and N is the number of cells per region. Cases 1-12 represent dense graphical features with all edges present and referred to as *Scenario 1*, where as Cases 13 and 14 use graphical features generated from different stochastic block models. Thus these two cases are referred to as *Scenario 2*.

Cases	*Node sparsity* $(1-\Delta_t)$	*Features per node* (N)	*Cases*	*Node sparsity* $(1-\Delta_t)$	*Features per node* (N)
Case 1	0.9	10	Case 8	0.8	25
Case 2	0.9	15	Case 9	0.7	10
Case 3	0.9	20	Case 10	0.7	15
Case 4	0.9	25	Case 11	0.7	20
Case 5	0.8	10	Case 12	0.7	25
Case 6	0.8	15	Case 13	0.7	20
Case 7	0.8	20	Case 14	0.7	20

(6) ensures *Edge constraint* whereby any edge connecting node $\mathcal{N}_v$ in the graph-valued feature is un-influential if $\mathcal{N}_v$ is un-influential, i.e., $\eta_{v,t} = 0 \Rightarrow \lambda_{[vv'],t} = 0$ for all $v' \in \{1, .., V\}$. Corresponding to each un-influential node $\mathcal{N}_v$, the $N \times 1$ dimensional structural feature coefficient $\boldsymbol{\beta}_{v,t}$ is set at $\mathbf{0}$. When $\eta_{v,t} = 1$, indicating that node $\mathcal{N}_v$ is influential, only a proportion $\upsilon_t = 0.4$ of the elements in the structural feature vector $\boldsymbol{z}_i^{(v)}$ are considered to be influentially related to the outcome. Thus, only υ_t proportion of $\boldsymbol{\beta}_{v,t}$ is non-zero, and these non-zero values are simulated from a N(1,1) distribution.

Simulation cases. For a comprehensive simulation study, we consider 14 cases after varying N and the node sparsity parameter $(1-\Delta_t)$, as summarized in Table 2. These 14 cases include two different scenarios to simulate the graph-valued feature, as we describe below.

Scenario 1: In Cases 1–12, the upper triangular entries of the undirected graph-valued feature $\boldsymbol{M}_i$ is simulated from a standard normal distribution, resulting in a dense graph-valued predictor (i.e., there is an edge between any pair of nodes). These 12 cases are together referred to as *Scenario 1*.

Scenario 2: To further assess the performance of competitors, the graph feature is generated following a stochastic block-model in Cases 13 and 14. In Case 13, we assume that each graph has three local clusters with high within-cluster and low between-cluster connectivity. More specifically, the matrices $\boldsymbol{M}_i$ consist of three symmetric block diagonal matrices of

dimensions 6×6, 7×7, and 7×7, respectively. Elements in these matrices are drawn from $\mathrm{N}(j, j^2)$ where $j \in \{1, 2, 3\}$, for the j-th block diagonal. The off-diagonal blocks are highly sparse, with very few non-sparse elements denoting connections between nodes in different clusters, randomly chosen from N(0, 1). In Case 14, each graph feature consists of 3 block diagonal matrices of dimensions 6×6, 7×7, and 7×7. As before, the elements in these matrices have been drawn from $\mathrm{N}(j, j^2)$ where $j \in \{1, 2, 3\}$, for the j-th block diagonal. However, in this case the elements in the off-diagonal matrices have been drawn from $\mathrm{N}(2, 1)$, $\mathrm{N}(3, 1)$, and $\mathrm{N}(4, 1)$. These two cases are referred to as *Scenario 2* to differentiate them from the cases in *Scenario 1*. The error variance τ_t^2 is fixed at 1 under all simulation settings.

5.1.1 Identification of Influential Regions

Table 3 presents the true positive rate (TPR) and false positive rate (FPR) of correctly identifying the truly influential nodes by the five competing models, averaged over 100 simulations. While both BMFR and BNL approaches allow identification of influential nodes from the node-specific latent binary indicators in a principled Bayesian manner as described in Section 4, ordinary MCP, BART and Horseshoe are not designed for influential node identification. To compare the performance of ordinary MCP, BART, and Horseshoe with BMFR in terms of identifying influential nodes, a post-processing strategy is devised. For MCP, a node $\mathcal{N}_v$ is considered influential if at least one of the structural features nested in $\mathcal{N}_v$ or one of the edges connected to $\mathcal{N}_v$ in the graph-valued feature is found to be statistically significant in the regression. In other words, $\mathcal{N}_v$ is identified as influential if any element of $\boldsymbol{\beta}_v$, $\{\lambda_{[vv']} : v' > v\}$, or $\{\lambda_{[v'v]} : v' < v\}$ (referring to (2)) is estimated to be non-zero. The same technique is applied to identify influential nodes in BART. For the ordinary Horseshoe prior, a post-processing step is applied following the approach in Guha and Rodriguez (2021) to distinguish between signal features and noise features. Specifically, we consider $\mathcal{N}_v$ to be influential if at least one of the coefficients in $\boldsymbol{\beta}_v$, $\{\lambda_{[vv']} : v' > v\}$, or $\{\lambda_{[v'v]} : v' < v\}$ is estimated to be non-zero in the post-processing step of Horseshoe. It is important to note that the neural network (NN) model is not designed to explicitly draw inference on influential nodes.

The TPR values close to 1 and FPR values close to 0 in most cases (see Table 3), except

Table 3: True Positive Rates (TPR) and False Positive Rates (FPR) for identifying the truly influential nodes for BMFR, BNL, Horseshoe (HS), BART and MCP are presented under all simulation cases. Highest TPR and lowest FPR are boldfaced in each case. Results are averaged over 100 replications.

Node Sparsity $(1-\Delta_t) = 0.9$, Scenario 1											
		True Positive Rate					*False Positive Rate*				
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*
1	10	**1.00**	1.00	0.87	0.13	0.06	0.00	0.82	0.02	0.02	0.06
2	15	**0.99**	1.00	0.84	0.07	0.03	0.00	0.86	0.03	0.00	0.04
3	20	**1.00**	1.00	0.87	0.03	0.11	0.00	0.84	0.04	0.01	0.04
4	25	**0.99**	1.00	0.79	0.05	0.09	0.00	0.95	0.04	0.01	0.05
Node Sparsity $(1-\Delta_t) = 0.8$, Scenario 1											
		True Positive Rate					*False Positive Rate*				
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*
5	10	**1.00**	1.00	0.99	0.02	0.07	0.00	0.80	0.04	0.00	0.06
6	15	**0.99**	1.00	0.99	0.04	0.05	0.00	0.87	0.04	0.01	0.06
7	20	**0.99**	1.00	0.99	0.08	0.06	0.01	0.84	0.03	0.02	0.08
8	25	**1.00**	1.00	0.91	0.05	0.10	0.10	0.86	0.08	0.03	0.04
Node Sparsity $(1-\Delta_t) = 0.7$, Scenario 1											
		True Positive Rate					*False Positive Rate*				
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*
9	10	**1.00**	1.00	0.91	0.03	0.07	0.00	0.82	0.03	0.02	0.06
10	15	**1.00**	1.00	0.88	0.04	0.08	0.00	0.77	0.02	0.02	0.04
11	20	**1.00**	1.00	0.79	0.05	0.09	0.14	0.90	0.08	0.03	0.04
12	25	**1.00**	1.00	0.54	0.05	0.05	0.74	0.98	0.03	0.03	0.04
Node Sparsity $(1-\Delta_t) = 0.7$, Scenario 2											
		True Positive Rate					*False Positive Rate*				
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*	*BMFR*	*HS*	*MCP*	*BART*	*BNL*
13	20	**1.00**	1.00	0.79	0.03	0.01	0.14	0.87	0.08	0.01	0.00
14	20	**0.99**	1.00	0.83	0.02	0.01	0.11	0.90	0.19	0.01	0.02

for Case 12, indicate that BMFR achieves highly accurate detection of influential nodes. MCP is the second best performer, also exhibiting high TPR and low FPR in scenarios with high sparsity. In contrast, Horseshoe shows high TPR and FPR, identifying almost all nodes as influential. BNL and BART are least competitive in terms of influential node identification. As sparsity decreases and the number of nodes V increases, the performance of all models tends to deteriorate, with case 12 demonstrating high FPR for BMFR and low TPR for MCP. However, it is worth noting that BMFR provides uncertainty estimates in

identifying $\mathcal{N}_v$ as influential through $P(\eta_v = 1|\text{Data})$, which is not offered by its competitors. The performance does not seem to be significantly affected by the different structures of the network predictor in *Scenario 1* and *Scenario 2*.

5.1.2 Predictive Inference

The predictive inference of different models is compared using $n^* = 100$ out-of-sample observations. Point prediction is evaluated using the mean squared prediction error (MSPE), which measures the average squared difference between the true and predicted outcome. Uncertainty quantification is assessed through the coverage and length of 95% predictive intervals. Table 4 provides the results of predictive inference for all competing models.

Table 4: Mean Squared Prediction Error (MSPE) are presented for all the competitors under 14 different cases, with the lowest MSPE in each case in boldfaced. Additionally, average coverage and average length of 95% predictive intervals for BMFR, Horseshoe (HS) and BNL are presented for all simulation scenarios. Results are averaged over 100 replications.

Node Sparsity $(1-\Delta_t) = 0.9$, Scenario 1													
		MSPE						*Avg. Coverage*			*Avg. length*		
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*NN*	*BNL*	*BMFR*	*HS*	*BNL*	*BMFR*	*HS*	BNL
1	10	**1.12**	1.14	1.14	3.90	85.30	2.918	0.72	0.88	0.909	1.38	2.62	5.697
2	15	**1.20**	1.24	1.22	6.09	109.35	2.676	0.77	0.92	0.941	1.71	3.17	6.248
3	20	**1.21**	1.38	1.29	9.87	157.22	2.832	0.81	0.95	0.962	2.03	3.89	6.984
4	25	1.45	1.52	**1.44**	18.58	148.36	3.298	0.99	0.99	0.970	0.02	0.17	7.740
Node Sparsity $(1-\Delta_t) = 0.8$, Scenario 1													
		MSPE						*Avg. Coverage*			*Avg. length*		
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*NN*	*BNL*	*BMFR*	*HS*	*BNL*	*BMFR*	*HS*	*BNL*
5	10	**1.24**	1.49	1.43	13.84	149.34	11.545	0.74	0.87	0.913	2.25	4.00	11.841
6	15	**1.42**	1.89	1.68	32.23	193.45	11.903	0.79	0.92	0.961	2.83	5.07	14.312
7	20	**1.61**	2.16	2.32	59.65	159.78	16.702	0.84	0.96	0.969	3.47	6.08	17.855
8	25	**1.99**	3.12	7.03	81.98	220.75	20.652	0.78	0.91	0.978	4.57	7.80	21.388
Node Sparsity $(1-\Delta_t) = 0.7$, Scenario 1													
		MSPE						*Avg. Coverage*			*Avg. length*		
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*NN*	*BNL*	*BMFR*	*HS*	*BNL*	*BMFR*	*HS*	*BNL*
9	10	**1.60**	2.26	2.01	62.64	241.59	27.538	0.81	0.92	0.920	3.18	5.74	18.915
10	15	**1.97**	3.37	8.55	98.04	248.33	32.869	0.88	0.96	0.959	4.27	8.22	23.797
11	20	**6.01**	22.74	52.69	147.02	284.98	50.923	0.92	0.96	0.971	7.64	18.20	30.384
12	25	**39.29**	68.21	89.46	187.21	308.31	76.176	0.95	0.92	0.967	21.79	28.03	36.520
Node Sparsity $(1-\Delta_t) = 0.7$, Scenario 2													
		MSPE						*Avg. Coverage*			*Avg. length*		
Cases	*N*	*BMFR*	*HS*	*MCP*	*BART*	*NN*	*BNL*	*BMFR*	*HS*	*BNL*	*BMFR*	*HS*	*BNL*
13	20	**6.11**	24.09	49.28	52.23	141.98	82.37	0.92	0.94	0.964	7.61	17.92	37.260
14	20	**6.72**	27.28	53.35	737.76	2114.94	75.417	0.92	0.94	0.967	7.17	18.44	36.593

In cases 1-4 with high sparsity, BMFR, MCP and Horseshoe exhibit similar performance

in terms of MSPE. BMFR starts outperforming its competitors as sparsity decreases, i.e., in cases 5-14, and the performance gap widens as sparsity decreases and N increases. Digging a bit further, we observe that with moderate degree of sparsity and larger values of N, both Horseshoe and MCP over-shrink nonzero coefficients in order to estimate zero coefficients, which leads to poor performance of these methods. In terms of predictive uncertainty, BMFR exhibits slight under-coverage when sparsity is high and N is small, but the coverage improves and becomes close to the nominal level as sparsity decreases and N increases. Horseshoe, on the other hand, offers slightly better coverage than BMFR, but at the cost of larger predictive intervals, which are almost twice the size of those in BMFR. BNL and BART show higher MSPE compared to BMFR in all simulation cases. NN performs the worst among the competitors, likely due to its inability to fully exploit the nested structure between the two sets of features. Overall, BMFR appears to be a much better performer than its competitors in terms of prediction under a variety of simulation settings.

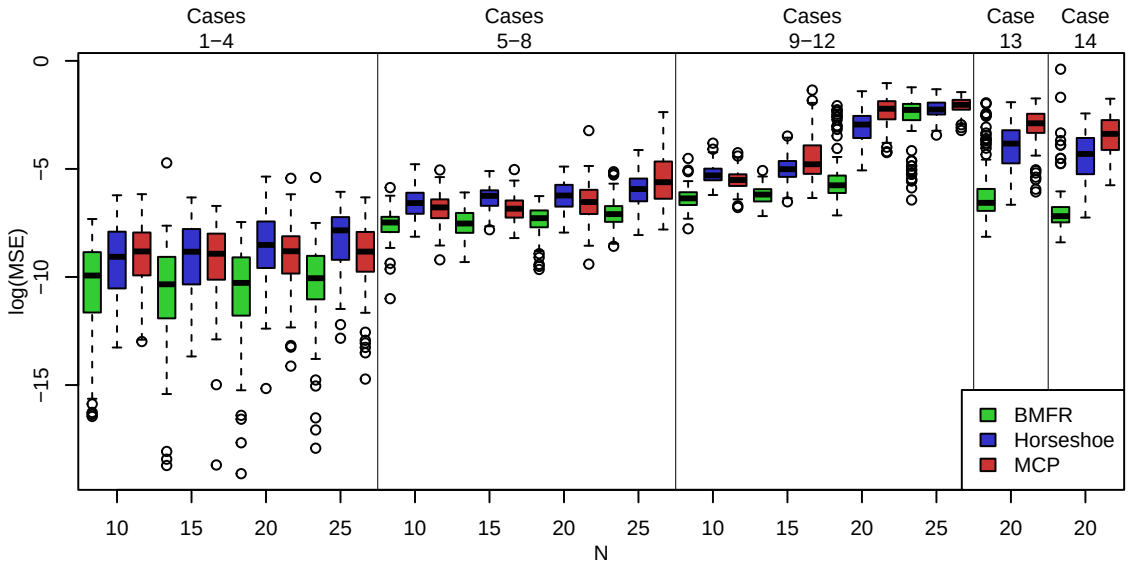
5.1.3 Estimation of Parameters $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\beta}_{v,t}$

True regression coefficients cannot be estimated from BART and NN since they fit non-linear regression functions. On the other hand, Horseshoe and MCP emerge as two best performing competitors of BMFR in Sections 5.1.1 and 5.1.2 vastly outperforming BNL. Hence, we show comparison with these two competitors in terms of estimating regression coefficients. Figures 4 and 5 present point estimation along with uncertainty quantification in estimating $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_{1,t}^T, ..., \boldsymbol{\beta}_{V,t}^T)^T$, respectively. The point estimation of every competitor is assessed using mean squared errors (MSE) of estimating the coefficients corresponding to the network- and the structural features. Since both $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_t$ are symmetric with zero diagonals, the MSE for the graphical coefficient is given by $2\sum_{v<v'}(\lambda_{[vv'],t} - \widehat{\lambda}_{[vv']})^2/V(V-1)$, where $\widehat{\lambda}_{[vv']}$ is the point estimate of $\lambda_{[vv']}$. Similarly, we compute and present MSE for the structural coefficients given by $\sum_{v=1}^{V} ||\boldsymbol{\beta}_{v,t} - \widehat{\boldsymbol{\beta}}_v||^2/NV$, with $\widehat{\boldsymbol{\beta}}_v$ representing the point estimate of $\boldsymbol{\beta}_v$. The point estimates are taken to be the posterior medians for the Bayesian competitors.

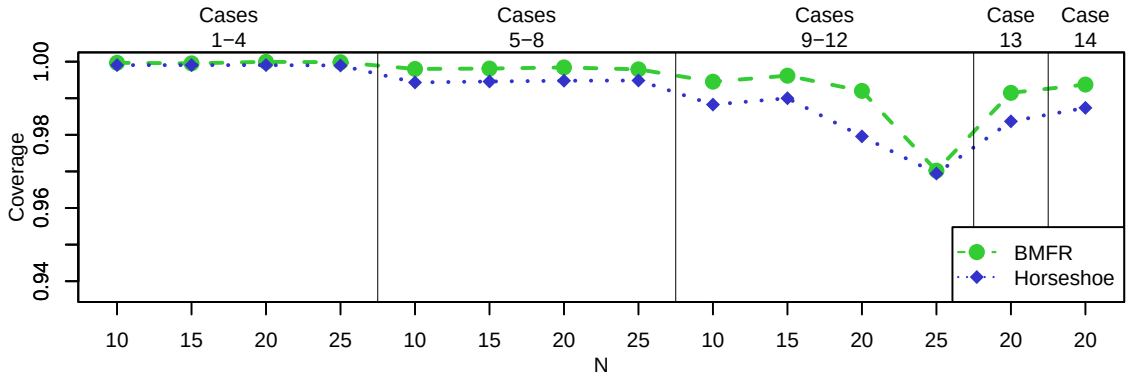
Both Figures 4 and 5 show that BMFR outperforms Horseshoe and MCP in all 14 cases. When both N is moderately large and true sparsity level is moderate, i.e. in cases 11,

Figure 4: Estimation of $\boldsymbol{\Theta}$: Figures present mean squared error, coverage and length of 95% credible interval for $\boldsymbol{\Theta}$.

(a) Mean Squared Error



(b) Coverage of 95% CI



(c) Length of 95% CI

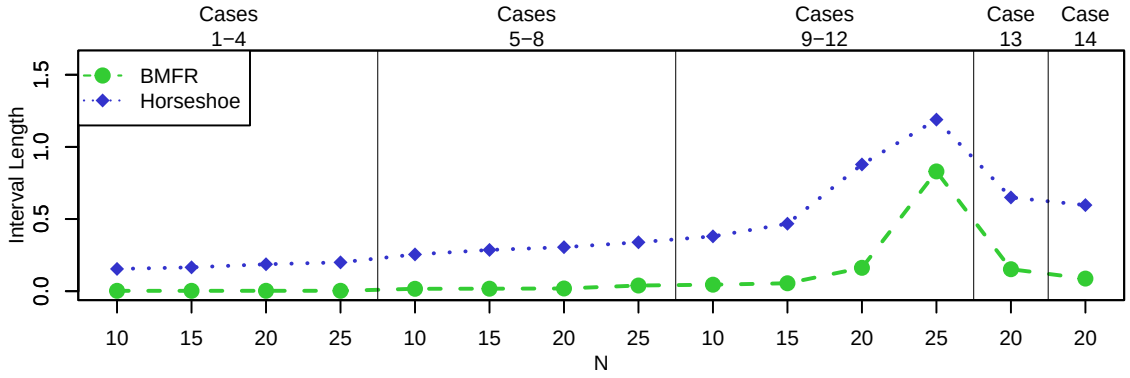
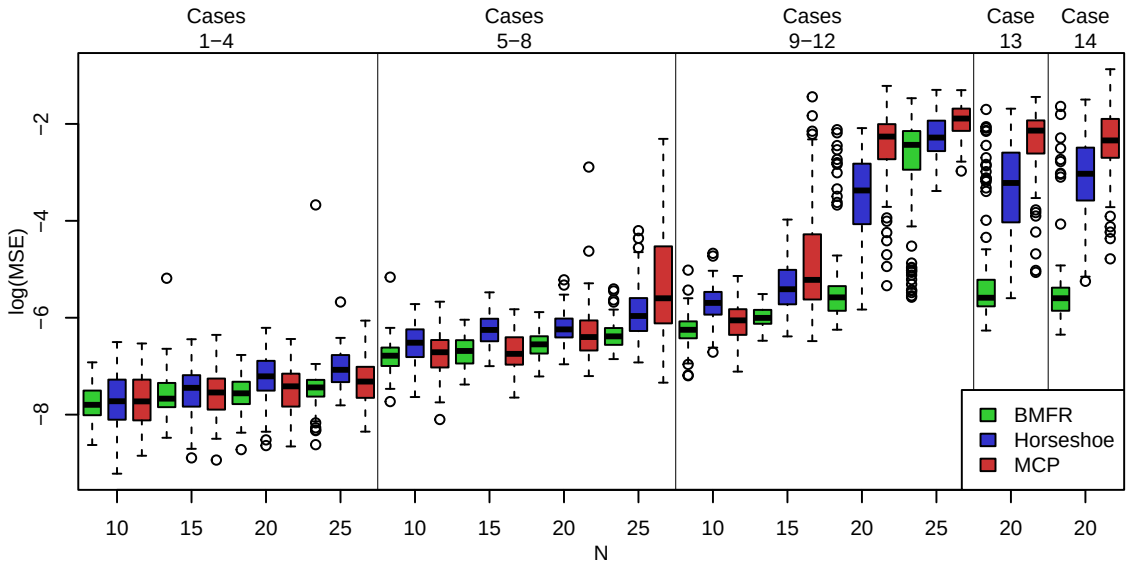
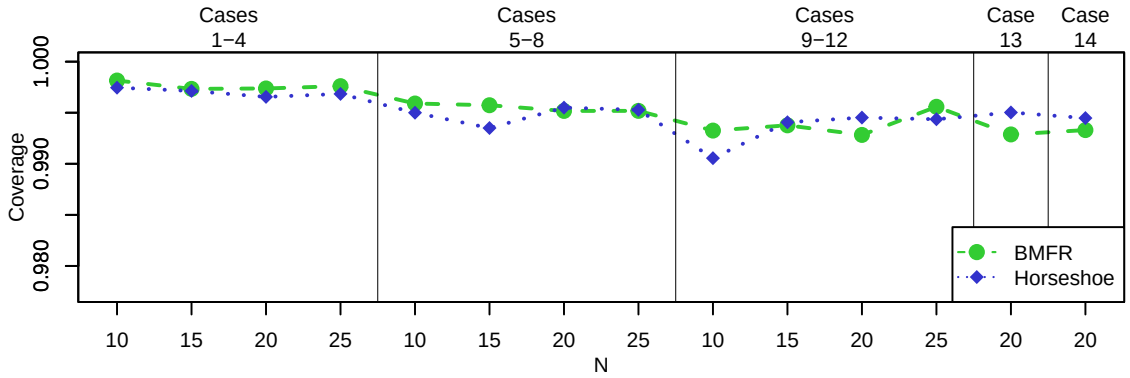


Figure 5: Estimation of $\boldsymbol{B}$: Figures present mean squared error, coverage and length of 95% credible interval for $\boldsymbol{B}$.

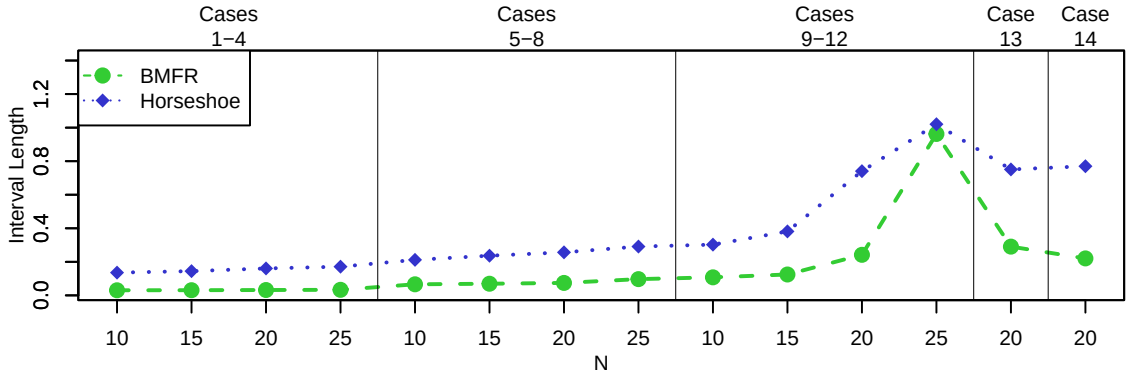
(a) Mean Squared Error



(b) Coverage of 95% CI



(c) Length of 95% CI



12, 13, 14, we perform overwhelmingly better than both competitors due to exploiting the graphical information and linkage between the graphical and the structural predictor. Since an overwhelming number of coefficients are set to zero in cases 1-8, we modestly outperform our competitors. This might be attributed to the fact that very high degree of sparsity in the truly influential nodes leads to high degree to sparsity in the regression coefficients in the truth, which is conducive for ordinary high dimensional regression which treats graph edges as one set of predictors. As we decrease sparsity or increase N keeping sample size fixed, the performance of all competitors deteriorate significantly, with BMFR continuing to show superior performance over the other two.

While both Bayesian competitors BMFR and Horseshoe provide automatic characterization of uncertainty, the resulting confidence intervals may not have the correct frequentist coverage in high dimensional regressions (Szabó *et al.*, 2015). Thus, in order to assess uncertainty in estimating $\boldsymbol{\Theta}_t$ from Bayesian competitors, we evaluate the length and coverage of 95% credible intervals averaged across coefficients in $\boldsymbol{\Theta}$ and present them in Figure 4 for all cases. Similar quantities are presented for $\boldsymbol{B}$ in Figure 5. Both figures show close to nominal coverage of BMFR under all cases. As sparsity decreases and N increases, the uncertainty associated with BMFR seems to be more which results in an increase in the length of the credible intervals. Importantly, under all cases, BMFR enjoys similar coverage as Horseshoe with much narrower credible intervals. This improved precision in uncertainty quantification can be attributed to BMFR's incorporation of the nested arrangement between structural and graph features, which allows for a more informed and precise estimation of the regression coefficients.

6 Analysis of General Cognition in Children in ABCD Study

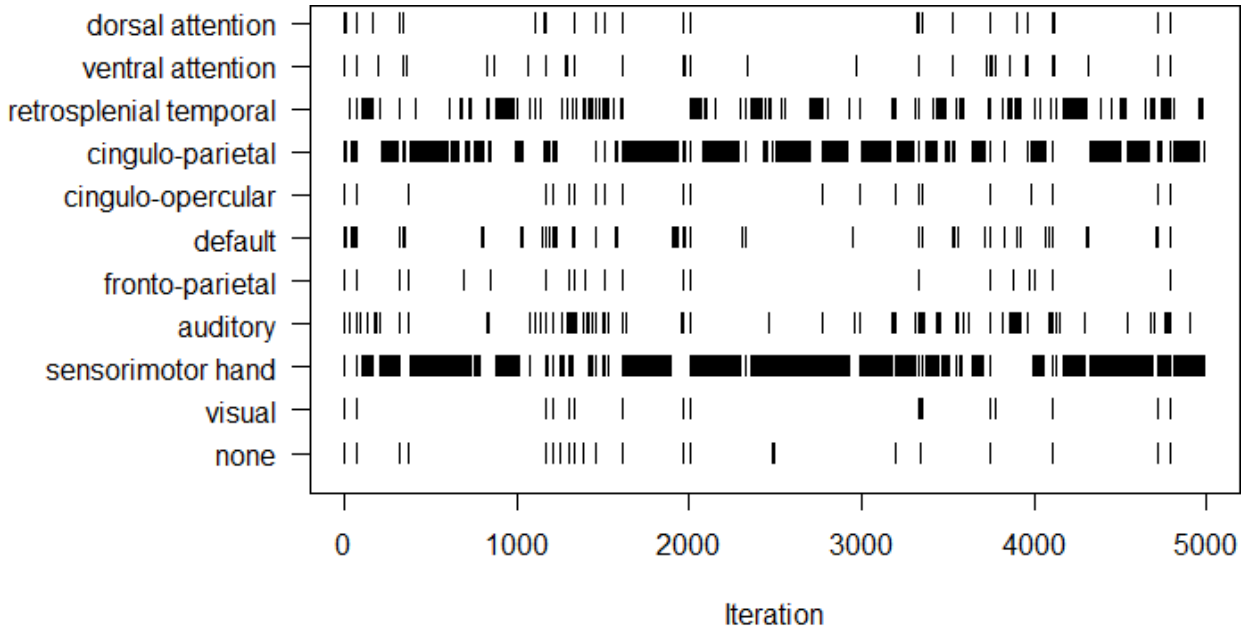
In this section, we analyze the data on general cognition of children extracted from the ABCD study to address the scientific questions outlined in Section 2. During our data analysis, we perform regression of the cognitive score y on both the graph-valued feature and the structural features for a total of $n = 395$ subjects. The graph-valued feature for each

subject is denoted by a symmetric matrix $\boldsymbol{M} \in \mathbb{R}^{11\times 11}$, which estimates the interconnections between functional networks in the brain. Additionally, the structural features, represented by $\{\boldsymbol{z}_i^{(v)} \in \mathbb{R}^{N_v}\}$, capture the structural information of the brain across ROIs obtained using task-based functional magnetic resonance imaging (t-fMRI). The dimension of the structural features N_v represents the number of ROIs in the vth functional network. They vary between 2 to 28 ROIs.

Figure 6 illustrates the plot of activation indicators η_v's across 5000 post-burn-in iterations corresponding to the 11 functional networks. In each specific iteration, a black vertical line corresponding to the vth network indicates that particular network is influentially related to the outcome for that particular iteration. Among the 11 networks, the *sensorimotor-hand* and *cingulo-parietal* networks are included in 72% and 59% of the iterations, respectively, suggesting the estimated posterior probabilities of them being influentially related to the response are 0.72 and 0.59, respectively. The *retrosplenial temporal* network has a posterior probability of 0.30 of being included, while the remaining functional networks show posterior probabilities of inclusion close to 0. The connection between two identified functional networks have been previously identified as correlates of human intelligence (Hearne *et al.*, 2016; Pat *et al.*, 2022). The findings from BMFR also generally agree with the dominant neuroscience paradigm for explaining variation in human intelligence which focuses on the parieto-frontal regions of the brain (Deary *et al.*, 2010). Crucially, BMFR enables quantifying uncertainty in the identification of networks by offering posterior probabilities of networks being selected. In contrast, the only other method capable of identifying influential functional networks (using only the graph-valued feature), BNL, estimates very similar probabilities (between 45% and 49%) to every functional network, making it challenging to derive any meaningful insights.

BMFR also facilitates inference on regions of interest (ROIs) within influential networks. To illustrate this, Figure 8 displays the 95% posterior credible intervals (CIs) and medians for the coefficients of structural features within the *sensorimotor-hand* and *cingulo-parietal* networks. Within the *sensorimotor-hand* network, four ROIs exhibit significant associations with the cognitive score, with one ROI demonstrating a significantly positive effect. Conversely, within the *cingulo-parietal* network, two ROIs display significant associations with

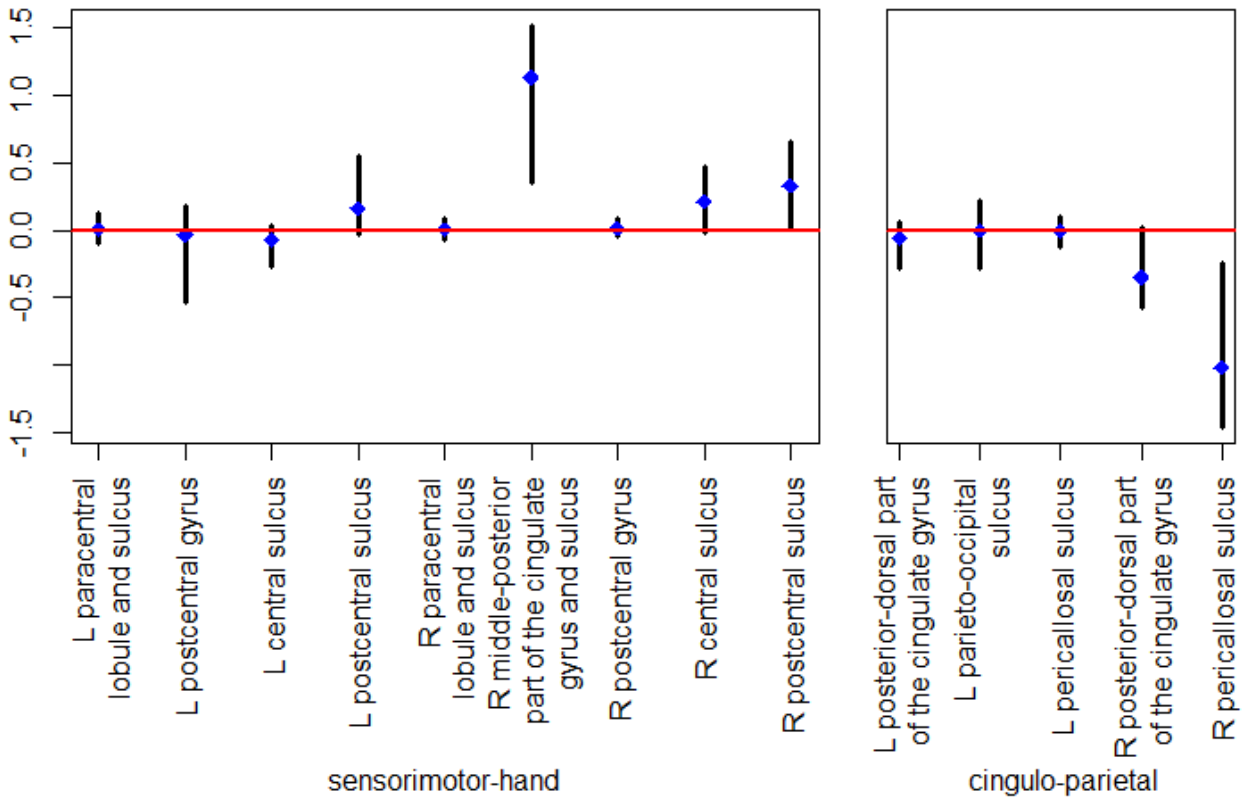
Figure 6: Activation indicators η_v's corresponding to 11 functional networks for 5000 post burn-in iterations. A vertical line in an iteration indicates the corresponding network being active in that iteration.



the cognitive score, one of which demonstrates a significantly negative effect.The association of task-based activation with general cognition in the two identified ROIs reinforces previous findings that link regions in the cingulo-parietal network (Vazquez-Trejo *et al.*, 2022) and posterior cingulate cortex (Leech and Sharp, 2014) with working memory. With respect to general cognition, BMFR estimates a positive association between activation in the right middle-posterior part of the cingulate gyrus and sulcus and a negative association between activation in the right pericallosal sulcus though it is difficult to situate the direction of these associations within the literature which presents mixed findings (Leech and Sharp, 2014). It is noteworthy that the remaining ROIs in both networks have coefficients with 95% CIs encompassing zero, indicating them being un-influential in the current sample though a larger sample may provide greater precision.

To assess the out-of-sample predictive performance of the competing methods, we conducted a leave-one-out analysis and calculated the mean squared prediction error (MSPE), along with the coverage and length of 95% predictive intervals. Regarding point estimation, BMFR notably outperforms other alternatives in terms of MSPE, with BART exhibiting the second-best performance (see Table 5). In terms of predictive uncertainty, BMFR ex-

Figure 7: Posterior median and 95% CIs for structural coefficients $\boldsymbol{\beta}_v$'s for ROIs within influential networks.

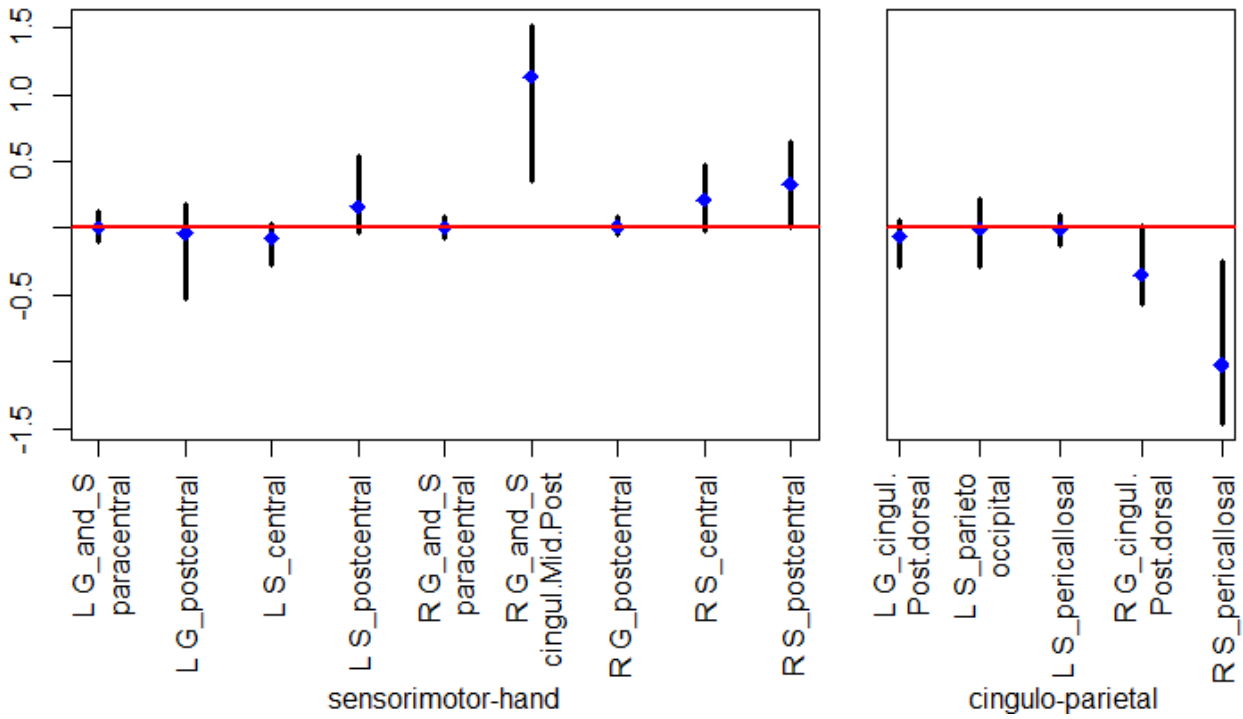


hibits slight under-coverage with narrower CIs compared to BNL and HS. BART provides the shortest CIs, albeit at the expense of significant under-coverage. MCP and NN are not designed to offer predictive uncertainty.

6.1 Sensitivity to Hyper-parameter Choice

Our hierarchical Bayesian modeling framework notably involves two sets of user-dependent hyperparameters: those for the prior distribution of η_v's denoted by a_Δ and b_Δ, and those for the prior distribution of error variance, denoted by a_τ and b_τ. While our analysis sets these parameters as $a_\Delta = b_\Delta = 1$ and $a_\tau = b_\tau = 1$, we explore the posterior probability of each functional network being influential by fitting BMFR with different combinations of $a_\tau, b_\tau, a_\Delta, b_\Delta$, as depicted in Table 6. As shown in Table 6, the posterior probabilities of functional networks being active remain robust despite changes in hyperparameters. Moreover, predictive inference for BMFR appears to be robust (see Table 6), with minimal perturbation in MSPE, coverage and length of 95% PIs observed across different combinations of

Figure 8: Posterior median and 95% CIs for structural coefficients $\boldsymbol{\beta}_v$'s for ROIs within influential networks.



hyper-parameters.

7 Conclusion and Future Work

This article presents a novel hierarchical Bayesian regression framework designed for a scalar outcome and features obtained across multiple scales: (a) a graph-valued feature at a coarse scale, and (b) fine-scale structural features at each node in the graph. We introduce a novel prior distribution on structural and feature coefficients jointly, considering the nested relationship between these two sets of features. This framework enables simultaneous inference on influential graph nodes related to the outcome, feature coefficients, and predictive inference on the outcome, while incorporating uncertainty in all aspects of the inference. Compared to conventional high-dimensional regression methods, tree-based nonlinear regression approaches, and neural networks that overlook graph topology or the nested structure between feature sets, our method demonstrates superior performance. It also outperforms regression techniques that leverage the graph topology but fail to effectively exploit the nested architecture between feature sets to predict the outcome. Our approach provides

Table 5: Mean squared prediction error (MSPE) for the competing methods are presented where point prediction is computed using leave-one-out analysis. We also compute 95% credible intervals for Bayesian competitors using leave-one-out analysis and present the length and coverage of 95% CIs for all competitors.

Competitors	*MSPE*	*Coverage*	*Length*
BMFR	**0.495**	0.942	2.608
BNL	0.562	**0.965**	3.034
HS	0.564	0.949	2.815
MCP	0.622	-	-
NN	6.698	-	-
BART	0.550	0.762	**1.687**

Table 6: Sensitivity analysis for the hyper-parameters. We show estimated posterior probability of a functional network being active with different combinations of hyper-parameters.

	$a_\Delta = b_\Delta = 0.1$			$a_\Delta = b_\Delta = 1$			$a_\Delta = b_\Delta = 10$		
$a_\tau = b_\tau =$	0.1	1	10	0.1	1	10	0.1	1	10
sensorimotor-hand	0.73	0.72	0.70	0.75	0.72	0.74	0.77	0.74	0.73
auditory	0.18	0.18	0.18	0.18	0.19	0.18	0.17	0.18	0.18
cingulo-parietal	0.59	0.59	0.57	0.61	0.59	0.60	0.62	0.59	0.59
retrosplenial temporal	0.27	0.28	0.27	0.29	0.30	0.28	0.32	0.32	0.31
max(Other)	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.08	0.08
MSPE	0.464	0.464	0.464	0.465	0.464	0.465	0.465	0.465	0.466
Interval Coverage	0.946	0.947	0.953	0.945	0.946	0.952	0.943	0.944	0.950
Interval Length	2.579	2.586	2.627	2.570	2.574	2.617	2.571	2.576	2.609
R^2	0.228	0.228	0.228	0.226	0.227	0.227	0.226	0.225	0.224

valuable insights into brain ROIs and functional networks critical for understanding general cognition in children.

Our proposed approach captures linear associations between the two sets of features at different scales and the outcome. We foresee an immediate extension of our approach to capture nonlinear associations between the outcome and features, facilitating inference on influential nodes, regression functions, and predictive inference on the outcome. Such methods are likely to be computationally intensive, prompting us to focus on developing approximate Bayesian algorithms to achieve efficient inference.

8 Acknowledgements

Rajarshi Guhaniyogi acknowledges funding from National Science Foundation Grant DMS-2220840 and DMS-2210672; and Office of Naval Research Grant N00014-18-1-274. Aaron Scheffler acknowledges funding from National Science Foundation Grant DMS-2210206. Aaron Scheffler and Rajarshi Guhaniyogi acknowledge funding from the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number R01NS131604.

9 ABCD Data Acknowledgement

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at `https://abcdstudy.org/federal-partners.html`. A listing of participating sites and a complete listing of the study investigators can be found at `https://abcdstudy.org/consortium_members`. ABCD consortium investigators designed

and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from doi: 10.15154/8873-zj65.

A Full Conditionals

Bayesian estimation of the model is performed through Gibbs sampling which cycles through the following steps.

1. β_0 and $\boldsymbol{\beta}_x$ jointly from the full conditional $p(\beta_0, \boldsymbol{\beta}_x | \tau^2, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_v, \boldsymbol{\Lambda}, \boldsymbol{y})$.

2. For each $v \in 1, \ldots, V$ sample jointly η_v, $\boldsymbol{\beta}_v$ and $\boldsymbol{\Lambda}_{-v,v}$ from:

$$\begin{aligned} &p(\eta_v, \boldsymbol{\Lambda}_{-v,v}, \boldsymbol{\beta}_v | \boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v' \neq v}, \Delta, \tau^2, \boldsymbol{A}^2_{-v,v}, \sigma^2_\lambda, \boldsymbol{\theta}_v, \xi^2_v, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y}) \\ &\quad = p(\boldsymbol{\Lambda}_{-v,v}, \boldsymbol{\beta}_v | \boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v' \neq v}, \tau^2, \boldsymbol{A}^2_{-v,v}, \sigma^2_\lambda, \boldsymbol{\theta}_v, \xi^2_v, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y} \\ &\qquad \times p(\eta_v | \boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v' \neq v}, \Delta, \tau^2, \boldsymbol{A}^2_{-v,v}, \sigma^2_\lambda, \boldsymbol{\theta}_v, \xi^2_v, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y}). \end{aligned}$$

3. Sample Δ from the full conditional: $p(\Delta | \boldsymbol{\eta})$.

4. Sample τ^2 from the full conditional: $p(\tau^2 | \beta_0, \boldsymbol{\beta}_x, \{\boldsymbol{\beta}_v\}_{v=1}^V, \boldsymbol{\Lambda}, \boldsymbol{A}^2, \sigma^2_\lambda, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_V, \boldsymbol{\xi}^2, y)$.

5. Sample the horseshoe parameters $(\boldsymbol{A}^2, \sigma^2_\lambda)$ and $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_V, \boldsymbol{\xi}^2)$ for $\boldsymbol{\Lambda}$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_V$, repectively, using the latent variable approach as in (Makalic and Schmidt, ????).

Next we describe all the relevant conditional distributions in steps 1-5. Note that:

$$(\beta_0, \boldsymbol{\beta}'_x)' | \tau^2, \{\boldsymbol{\beta}_v\}_{v=1}^V, \boldsymbol{\Lambda}, \boldsymbol{y} \sim N(\hat{\boldsymbol{\beta}}_X, \tau^2 (\boldsymbol{X}'\boldsymbol{X})^{-1})$$

where

$$\hat{\boldsymbol{\beta}}_X = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{r}_x,$$

$$\boldsymbol{X} = (\boldsymbol{1}_n, (x'_1, \ldots, x'_n)'),$$

$$\boldsymbol{r}_x = (r_{x,1}, \ldots, r_{x,n})',$$

$$r_{x,i} = y_i - \sum_{v=1}^{V} \boldsymbol{\beta}_v^T \boldsymbol{z}_i^{(v)} - \langle \boldsymbol{M}_i, \boldsymbol{\Lambda} \rangle / 2.$$

To sample $\boldsymbol{\Lambda}_{-v,v}$ and $\boldsymbol{\beta}_v$ set them to 0 if $\eta_v = 0$, if $\eta_v = 1$ there are 2 cases. In case 1, if $\eta_{v'} = 0 \quad \forall v'$, then set $\boldsymbol{\Lambda}_{-v,v} = 0$ and sample $\boldsymbol{\beta}_v$ from:

$$\boldsymbol{\beta}_v | \boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v' \neq v}, \tau^2, \boldsymbol{\theta}_v, \xi_v^2, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y} \sim N(\hat{\boldsymbol{b}}_v, \tau^2 (\boldsymbol{Z}_v' \boldsymbol{Z}_v + \boldsymbol{L})^{-1})$$

where

$$\hat{\boldsymbol{b}}_v = (\boldsymbol{Z}_v' \boldsymbol{Z}_v + \boldsymbol{L}_v)^{-1} \boldsymbol{Z}_v' \boldsymbol{r}_v$$

$$\boldsymbol{Z}_v = (\boldsymbol{z}_1^{(v)'}, \ldots, \boldsymbol{z}_n^{(v)'})',$$

$$\boldsymbol{L}_v = \text{diag}(\xi_v^2 \boldsymbol{\theta}_v^2),$$

$$\boldsymbol{r}_v = (r_{v,i}, \ldots, r_{v,n})',$$

$$r_{v,i} = y_i - \beta_0 - \boldsymbol{\beta}_{\boldsymbol{x}}^T \boldsymbol{x}_i - \sum_{v' \neq v} \boldsymbol{\beta}_{v'}^T \boldsymbol{z}_i^{(v')} - \langle \boldsymbol{M}_{-v,-v,i}, \boldsymbol{\Lambda}_{-v,-v} \rangle / 2$$

In case 2, if $\exists v'$ such that $\eta_{v'} \neq 0$, then sample $\boldsymbol{\Lambda}_{-v,v}[\eta_{-v} = 1]$ and $\boldsymbol{\beta}_v$ from:

$$\boldsymbol{\Lambda}_{-v,v}[\eta_{-v} = 1], \boldsymbol{\beta}_v | \boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v' \neq v}, \tau^2, \boldsymbol{A}_{-v,v}, \sigma_\lambda^2, \boldsymbol{\theta}_v, \xi_v^2, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y} \sim N(\hat{\boldsymbol{b}}_v, \tau^2 (\boldsymbol{Z}_v' \boldsymbol{Z}_v + \boldsymbol{L})^{-1})$$

where

$$\hat{\boldsymbol{b}}_v = (\boldsymbol{Z}_v' \boldsymbol{Z}_v + \boldsymbol{L}_v)^{-1} \boldsymbol{Z}_v' \boldsymbol{r}_v$$

$$\boldsymbol{Z}_v = [(\boldsymbol{M}'_{-v,v,1}[\eta_{-v} = 1], \ldots, \boldsymbol{M}'_{-v,v,n}[\eta_{-v} = 1])', (\boldsymbol{z}_1^{(v)'}, \ldots, \boldsymbol{z}_n^{(v)'})'],$$

$$\boldsymbol{L}_v = \text{diag}(\boldsymbol{A}_{-v,v}[\eta_{-v} = 1], \xi_v^2 \boldsymbol{\theta}_v^2),$$

$$\boldsymbol{r}_v = (r_{v,i}, \ldots, r_{v,n})',$$

$$r_{v,i} = y_i - \beta_0 - \boldsymbol{\beta}_{\boldsymbol{x}}^T \boldsymbol{x}_i - \sum_{v' \neq v} \boldsymbol{\beta}_{v'}^T \boldsymbol{z}_i^{(v')} - \langle \boldsymbol{M}_{-v,-v,i}, \boldsymbol{\Lambda}_{-v,-v} \rangle / 2$$

And the indicators functions η_v sampled from:

$$\eta_v|\boldsymbol{\Lambda}_{-v,-v}, \{\boldsymbol{\beta}_{v'}\}_{v'\neq v}, \Delta, \tau^2, \boldsymbol{A}^2_{-v,v}, \sigma^2_\lambda, \boldsymbol{\theta}_v, \xi^2_v, \beta_0, \boldsymbol{\beta}_x, \boldsymbol{y} \sim Bernoulli(\hat{\Delta}_v)$$

, where

$$\hat{\Delta}_v = o_v/(1+o_v),$$

$$o_v = \exp\left(\frac{1}{2}C\right)\frac{\Delta}{1-\Delta},$$

$$C = \frac{\hat{\boldsymbol{b}}_v'(\boldsymbol{Z}_v'\boldsymbol{Z}_v + \boldsymbol{L}_v)\hat{\boldsymbol{b}}_v}{\tau^2} - |\boldsymbol{Z}_v'\boldsymbol{Z}_v + \boldsymbol{L}_v| - |\boldsymbol{L}_v|$$

with $\hat{\boldsymbol{b}}_v$, $\boldsymbol{Z}_v$ and $\boldsymbol{L}_v$ according to the cases 1 and 2. For Δ, sample from the full conditional:

$$\Delta|\boldsymbol{\eta} \sim Beta(\hat{a}_\Delta, \hat{b}_\Delta),$$

where

$$\hat{a}_\Delta = a_\Delta + \sum_{v=1}^{V}\eta_v,$$

$$\hat{b}_\Delta = b_\Delta + V - \sum_{v=1}^{V}\eta_v$$

Finally, sample τ^2 from the full conditional τ^2 from the full conditional:

$$p(\tau^2|\beta_0, \boldsymbol{\beta}_x, \{\boldsymbol{\beta}_v\}_{v=1}^V, \boldsymbol{\Lambda}, \boldsymbol{A}^2, \sigma^2_\lambda, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_V, \boldsymbol{\xi}^2, y \sim Gamma(\hat{a}_\tau, \hat{b}_\tau),$$

where

$$\hat{a}_\tau = a_\tau + \frac{1}{2}\left(n + \sum_{v=1}^{V}N_v + \frac{Q(Q-1)}{2}\right),$$

$$\hat{b}_\tau = b_\tau + \frac{1}{2}\left(\boldsymbol{r}'\boldsymbol{r} + \sum_{v=1}^{V}\sum_{j=1}^{N_v}\frac{\beta^2_{v,j}}{2\theta_{v,j}\xi^2_v} + \sum_{v=1}^{V}\sum_{v'\neq v}\frac{\lambda^2_{v,v'}}{\alpha^2_{v,v'}\sigma^2_\lambda}\right),$$

$$\boldsymbol{r} = (r_1, \ldots, r_n)',$$

$$r_i = y_i - \beta_0 - \boldsymbol{\beta}_{\boldsymbol{x}}^T\boldsymbol{x}_i - \sum_{v=1}^{V}\boldsymbol{\beta}_v^T\boldsymbol{z}_i^{(v)} - \langle\boldsymbol{M}_i, \boldsymbol{\Lambda}\rangle/2$$

References

Bassett, D. S. and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, **12**(6), 512–523.

Breheny, P. and Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, **5**(1), 232–253.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, **10**(3), 186–198.

Calhoun, V. D. and Sui, J. (2016). Multimodal fusion of brain imaging data: A key to finding the missing link (s) in complex mental illness. *Biological psychiatry: cognitive neuroscience and neuroimaging*, **1**(3), 230–244.

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

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., Chaarani, B., Mejia, M. H., Hagler, Jr, D. J., Daniela Cornejo, M., Sicat, C. S., Harms, M. P., Dosenbach, N. U. F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J. R., Kuperman, J. M., Fair, D. A., Dale, A. M., and ABCD Imaging Acquisition Workgroup (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.*, **32**, 43–54.

Chen, F., Wang, Y.-C., Wang, B., and Kuo, C.-C. J. (2020). Graph representation learning: a survey. *APSIPA Transactions on Signal and Information Processing*, **9**, e15.

Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., and Sun, J. (2017). Gram: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 787–795.

Cole, M. W., Ito, T., Bassett, D. S., and Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nat. Neurosci.*, **19**(12), 1718–1726.

Craddock, R. C., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, **62**(6), 1619–1628.

Dai, X. and Li, L. (2021). Orthogonal statistical inference for multimodal data analysis. *arXiv preprint arXiv:2103.07088*.

Deary, I. J., Penke, L., and Johnson, W. (2010). The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.*, **11**(3), 201–211.

Denison, D. G., Mallick, B. K., and Smith, A. F. (1998). A bayesian cart algorithm. *Biometrika*, **85**(2), 363–377.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, **53**(1), 1–15.

Dinh, V. C. and Ho, L. S. (2020). Consistent feature selection for analytic deep neural networks. *Advances in Neural Information Processing Systems*, **33**, 2420–2431.

Du, L., Liu, K., Yao, X., Risacher, S. L., Han, J., Guo, L., Saykin, A. J., and Shen, L. (2018). Fast multi-task scca learning with feature selection for multi-modal brain imaging genetics. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 356–361. IEEE.

Dubois, J., Galdi, P., Paul, L. K., and Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **373**(1756).

Fan, J., Gong, W., and Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of econometrics*, **212**(1), 177–202.

Feng, X., Li, T., Song, X., and Zhu, H. (2019). Bayesian scalar on image regression with nonignorable nonresponse. *Journal of the American Statistical Association*, pages 1–24.

Flynn, J. R. (2007). *What Is Intelligence?: Beyond the Flynn Effect*. Cambridge University Press.

Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *Journal of Computational and Graphical Statistics*, **23**(1), 46–64.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, **95**(4), 791–807.e7.

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., and Byrum, S. D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular omics*, **17**(2), 170–185.

Gray, J. R., Chabris, C. F., and Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nat. Neurosci.*, **6**(3), 316–322.

Guha, S. and Rodriguez, A. (2021). Bayesian Regression With Undirected Network Predictors With an Application to Brain Connectome Data. *Journal of the American Statistical Association*, **116**(534), 581–593.

Guha, S. and Rodriguez, A. (2023). High-dimensional bayesian network classification with network global-local shrinkage priors. *Bayesian Analysis*, **1**(1), 1–30.

Guhaniyogi, R. (2017). Convergence rate of Bayesian supervised tensor modeling with multiway shrinkage priors. *Journal of Multivariate Analysis*, **160**, 157–168.

Guhaniyogi, R. and Rodriguez, A. (2020). Joint modeling of longitudinal relational data and exogenous variables.

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *The Journal of Machine Learning Research*, **18**(1), 2733–2763.

Hagler, Jr, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B. J., Barch, D. M., Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Bartsch, H., Xue, F., Heitzeg, M. M., Laird, A. R., Trinh, T. T., Gonzalez, R., Tapert, S. F., Riedel, M. C., Squeglia, L. M., Hyde, L. W., Rosenberg, M. D., Earl, E. A., Howlett, K. D., Baker, F. C., Soules, M., Diaz, J., de Leon, O. R., Thompson, W. K., Neale, M. C., Herting, M., Sowell, E. R., Alvarez, R. P., Hawes, S. W., Sanchez, M., Bodurka, J., Breslin, F. J., Morris, A. S., Paulus, M. P., Simmons, W. K., Polimeni, J. R., van der Kouwe, A., Nencka, A. S., Gray, K. M., Pierpaoli, C., Matochik, J. A., Noronha, A., Aklin, W. M., Conway, K., Glantz, M., Hoffman, E., Little, R., Lopez, M., Pariyadath, V., Weiss, S. R., Wolff-Hughes, D. L., DelCarmen-Wiggins, R., Feldstein Ewing, S. W., Miranda-Dominguez, O., Nagel, B. J., Perrone, A. J., Sturgeon, D. T., Goldstone, A., Pfefferbaum, A., Pohl, K. M., Prouty, D., Uban, K., Bookheimer, S. Y., Dapretto, M., Galvan, A., Bagot, K., Giedd, J., Infante, M. A., Jacobus, J., Patrick, K., Shilling, P. D., Desikan, R., Li, Y., Sugrue, L., Banich, M. T., Friedman, N., Hewitt, J. K., Hopfer, C., Sakai, J., Tanabe, J., Cottler, L. B., Nixon, S. J., Chang, L., Cloak, C., Ernst, T., Reeves, G., Kennedy, D. N., Heeringa, S., Peltier, S., Schulenberg, J., Sripada, C., Zucker, R. A., Iacono, W. G., Luciana, M., Calabro, F. J., Clark, D. B., Lewis, D. A., Luna, B., Schirda, C., Brima, T., Foxe, J. J., Freedman, E. G., Mruzek, D. W., Mason, M. J., Huber, R., McGlade, E., Prescot, A., Renshaw, P. F., Yurgelun-Todd, D. A., Allgaier, N. A., Dumas, J. A., Ivanova, M., Potter, A., Florsheim, P., Larson, C., Lisdahl, K., Charness, M. E., Fuemmeler, B., Hettema, J. M., Maes, H. H., Steinberg, J., Anokhin, A. P., Glaser, P., Heath, A. C., Madden, P. A., Baskin-Sommers, A., Constable, R. T., Grant, S. J., Dowling, G. J., Brown, S. A., Jernigan, T. L., and Dale, A. M. (2019). Image processing and analysis methods for the adolescent brain cognitive development study. *Neuroimage*, **202**, 116091.

Hearne, L. J., Mattingley, J. B., and Cocchi, L. (2016). Functional brain networks related to individual differences in human intelligence at rest. *Sci. Rep.*, **6**, 32328.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, **100**(469), 286–295.

Huang, L., Goldsmith, J., Reiss, P. T., Reich, D. S., and Crainiceanu, C. M. (2013). Bayesian scalar-on-image regression with application to association between intracranial dti and cognitive outcomes. *NeuroImage*, **83**, 210–223.

Jiang, R., Calhoun, V. D., Cui, Y., Qi, S., Zhuo, C., Li, J., Jung, R., Yang, J., Du, Y., Jiang, T., and Sui, J. (2020). Multimodal data revealed different neurobiological correlates of intelligence between males and females. *Brain Imaging Behav.*, **14**(5), 1979–1993.

Kang, J., Reich, B. J., and Staicu, A.-M. (2018). Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*, **105**(1), 165–184.

Kang, M., Ko, E., and Mersha, T. B. (2022). A roadmap for multi-omics data integration using deep learning. *Briefings in Bioinformatics*, **23**(1), bbab454.

Leech, R. and Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain*, **137**(Pt 1), 12–32.

Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L., and Coan, J. A. (2015). Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *The Annals of Applied Statistics*, **9**(2), 687–713.

Li, Q. and Li, L. (2021). Integrative factor regression and its inference for multimodal data analysis. *Journal of the American Statistical Association*, pages 1–15.

Makalic, E. and Schmidt, D. F. (????). A simple sampler for the horseshoe estimator. **23**(1), 179–182. Conference Name: IEEE Signal Processing Letters.

Morris, S. E. and Cuthbert, B. N. (2012). Research domain criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin. Neurosci.*, **14**(1), 29–37.

Narr, K. L., Woods, R. P., Thompson, P. M., Szeszko, P., Robinson, D., Dimtcheva, T., Gurbani, M., Toga, A. W., and Bilder, R. M. (2007). Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cereb. Cortex*, **17**(9), 2163–2171.

Pamplona, G. S. P., Santos Neto, G. S., Rosset, S. R. E., Rogers, B. P., and Salmon, C. E. G. (2015). Analyzing the association between functional connectivity of the brain and intellectual performance. *Front. Hum. Neurosci.*, **9**, 61.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, **103**(482), 681–686.

Pat, N., Wang, Y., Anney, R., Riglin, L., Thapar, A., and Stringaris, A. (2022). Longitudinally stable, brain-based predictive models mediate the relationships between childhood cognition and socio-demographic, psychological and genetic factors. *Hum. Brain Mapp.*, **43**(18), 5520–5542.

Peng, J., An, L., Zhu, X., Jin, Y., and Shen, D. (2016). Structured sparse kernel learning for imaging genetics based alzheimer's disease diagnosis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 70–78. Springer.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian Statistics*, **9**, 501–538.

Polson, N. G. and Sokolov, V. O. (2018). Deep learning. *arXiv preprint arXiv:1807.07987*.

Rabini, G., Ubaldi, S., and Fairhall, S. L. (2023). Task-based activation and resting-state connectivity predict individual differences in semantic capacity for complex semantic knowledge. *Commun Biol*, **6**(1), 1020.

Rasero, J., Sentis, A. I., Yeh, F.-C., and Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLoS Comput. Biol.*, **17**(3), e1008347.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., and Van De Ville, D. (2011). Decoding brain states from fMRI connectivity graphs. *Neuroimage*, **56**(2), 616–626.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, **52**(3), 1059–1069.

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, **38**(5), 2587–2619.

Sparapani, R., Spanbauer, C., and McCulloch, R. (2021). Nonparametric machine learning and efficient computation with Bayesian additive regression trees: The BART R package. *Journal of Statistical Software*, **97**(1), 1–66.

Sripada, C., Rutherford, S., Angstadt, M., Thompson, W. K., Luciana, M., Weigard, A., Hyde, L. H., and Heitzeg, M. (2020). Prediction of neurocognition in youth from resting state fMRI. *Mol. Psychiatry*, **25**(12), 3413–3421.

Sui, J., Adali, T., Yu, Q., Chen, J., and Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of neuroscience methods*, **204**(1), 68–81.

Sui, J., Jiang, R., Bustillo, J., and Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises. *Biol. Psychiatry*, **88**(11), 818–828.

Szabó, B., Van Der Vaart, A. W., and van Zanten, J. (2015). Frequentist coverage of adaptive nonparametric bayesian credible sets. *The Annals of Statistics*, **43**(4), 1391–1428.

Thiele, J. A., Faskowitz, J., Sporns, O., and Hilger, K. (2022). Multitask brain network reconfiguration is inversely associated with human intelligence. *Cereb. Cortex*, **32**(19), 4172–4182.

Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B. J., Nixon, S. J., and Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Dev. Cogn. Neurosci.*, **36**, 100606.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**(1), 267–288.

Vazquez-Trejo, V., Nardos, B., Schlaggar, B. L., Fair, D. A., and Miranda-Dominguez, O. (2022). Use of connectotyping on task functional MRI data reveals dynamic network level cross talking during task performance. *Front. Neurosci.*, **16**, 951907.

Waiter, G. D., Deary, I. J., Staff, Roger T., Murray, A. D., Fox, H. C., Starr, J. M., and Whalley, L. J. (2009). Exploring possible neural mechanisms of intelligence differences using processing speed and working memory tasks: An fMRI study. *Intelligence*, **37**(2), 199–206.

Wakeman, D. G. and Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific data*, **2**(1), 1–10.

Xue, W., Bowman, F. D., and Kang, J. (2018). A bayesian spatial model to predict disease status using imaging data from various modalities. *Frontiers in neuroscience*, **12**, 184.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, **38**(2), 894–942.

Zhang, T. and Shi, M. (2020). Multi-modal neuroimaging feature fusion for diagnosis of alzheimer's disease. *Journal of Neuroscience Methods*, **341**, 108795.

Zhang, X., Xu, G., and Zhu, J. (2022). Joint latent space models for network data with high-dimensional node variables. *Biometrika*, **109**(3), 707–720.

Zhao, Y., Wu, B., and Kang, J. (2023). Bayesian interaction selection model for multimodal neuroimaging data analysis. *Biometrics*, **79**(2), 655–668.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, **108**(502), 540–552.