Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Bayesian Regression with Undirected Network Predictors with an Application to Brain Connectome Data

Sharmistha Guha & Abel Rodriguez

To cite this article: Sharmistha Guha & Abel Rodriguez (2020): Bayesian Regression with Undirected Network Predictors with an Application to Brain Connectome Data, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1772079

To link to this article: https://doi.org/10.1080/01621459.2020.1772079

Bayesian Regression with Undirected Network Predictors with an Application to Brain Connectome Data

Sharmistha Guha*[1] and Abel Rodriguez[2]

[1]Postdoctoral Associate, Department of Statistical Science, Duke University

[2]Professor, Department of Statistics, University of California, Santa Cruz

*Corresponding author: email: sharmistha.guha@duke.edu

Abstract

This article focuses on the relationship between a measure of creativity and the human brain network for subjects in a brain connectome dataset obtained using a diffusion weighted magnetic resonance imaging (DWI) procedure. We identify brain regions and interconnections that have a significant effect on creativity. Brain networks are often expressed in terms of symmetric adjacency matrices, with row and column indices of the matrix representing the regions of interest (ROI), and a cell entry signifying the estimated number of fiber bundles connecting the corresponding row and column ROIs. Current statistical practices for regression analysis with the brain network as the predictor and the measure of creativity as the response typically vectorize the network predictor matrices prior to any analysis, thus failing to account for the important structural information in the network. This results in poor inferential and predictive performance in presence of small sample sizes. To answer the scientific questions discussed above, we develop a flexible Bayesian framework that avoids reshaping the network predictor matrix, draws inference on brain ROIs and interconnections significantly related to creativity, and enables accurate prediction of creativity from a brain network. A novel class of *network shrinkage priors* for the coefficient corresponding to the network predictor is proposed to achieve these goals simultaneously. The Bayesian framework allows characterization of uncertainty in the findings. Empirical results in simulation studies illustrate substantial inferential and predictive gains of the proposed framework in comparison with the ordinary high dimensional Bayesian shrinkage priors and penalized optimization schemes. Our framework yields new insights into the relationship of brain regions with creativity, also providing the uncertainty associated with the scientific findings.

Keywords: Brain connectome; High dimensional regression; Influential nodes; Influential edges; Network predictors; Network shrinkage prior.

1 Introduction

In recent years, there has been a growing interest in the study of functional and structural interconnection networks in the human brain (Fornito *et al.*, 2013; Park and Friston, 2013; Jbabdi *et al.*, 2015), also referred to as *brain connectomics*. This article focuses on structural connections in the human brain, coinciding with fiber bundles that are estimated using diffusion weighted magnetic resonance imaging (DWI). The complex structural organization of the white matter of the brain can be depicted in vivo in great detail with advanced DWI (Hagmann *et al.*, 2006). Using data from DWI, the brain can be segmented into different regions of interest (ROI), with the fiber bundles connecting the different regions estimated.

Information on fiber connections in the human brain can be used to generate network valued object data, with the network nodes corresponding to regions of interest (ROI) and the estimated number of fiber bundles connecting any two ROIs signifying the strength of interconnection between them. There are several compelling possibilities in terms of relating these objects to brain related phenotypic traits of individuals (Zhang *et al.*, 2019). For instance, the number of fibers connecting pairs of ROIs varies across individuals, and perhaps features of these fiber connections as well as certain ROIs in the brain relate to traits of individuals, such as their creativity. In this article, we provide necessary Bayesian statistical tools to draw such inferences. In particular, we consider the problem of predicting the creativity of individuals (measured using the *composite creativity index* or CCI) based on neuroimaging data signifying the connectivity network among different brain regions. The objective of these studies is twofold. First, neuroscientists are often interested in identifying regions of the brain and interconnections between them which are involved in creative thinking. Secondly, it is important to determine how the strength of connection among these influential regions affects the level of creativity of the individual.

From a statistical point of view, our inferential problem can be best cast as a network regression problem. In a typical network regression framework, one is interested in the relationship between the structure of the network and one or more global attributes of the experimental unit on which the network data is collected. However, an overwhelming literature on network objects mainly relies on modeling the network as a random variable, rather than treating it as a predictor in a typical network regression. A number of classic models have emerged in the literature viewing the network as a random variable with a goal of either predicting an unobserved link in a network or investigating *homophily*, i.e., the process of formation of social ties due to matching individual traits. Examples include exponential random graph models (Frank and Strauss, 1986), social space models (Hoff *et al.*, 2002; Hoff, 2005) including random dot product graph (RDPG) models (Young and Scheinerman, 2007) and stochastic block models (Nowicki and Snijders, 2001). Alternatively, models which explain covariates related to network nodes as a function of the network structure (e.g., Fowler and Christakis, 2008; Shoham *et al.*, 2015) and joint models for the co-evolution of the network structure and nodal attributes (e.g., Fosdick and Hoff, 2015; Durante *et al.*, 2017; De la Haye *et al.*, 2010; Guhaniyogi and Rodriguez, 2018) have also found prominence in the statistical literature of networks.

However, the inferential goals in this article in the context of the brain connectome data mentioned earlier do not quite fit in to the aforementioned literature. To address these, we employ a Bayesian model for network regression with the brain network as the predictor and the composite creativity index as the response. In particular, we construct a Bayesian *network shrinkage prior* that combines ideas from spectral decomposition methods and spike-and-slab priors to generate a model that takes into account the topological structure of the network predictor. The model produces accurate predictions, allows us to identify both ROIs and interconnections among them that have influence on creativity, and yields well-calibrated interval estimates for the model parameters. For instance, a notable finding from the brain connectome dataset using our model is that a substantial number of ROIs influencing creativity are located in the *temporal* and *frontal* lobes of the human brain.

Earlier literature involving a network predictor and a scalar response has not amply exploited the relational nature of the network predictor. A common approach to network regression is to use a few summary measures from the network in the context of a flexible regression or classification approach (e.g., see Bullmore and Sporns, 2009 and references therein). Clearly, the success of this approach is highly dependent on selecting the right summaries to include. Furthermore, such an approach cannot identify the impact of specific nodes on the response, which is of clear interest in our setting. Alternatively, a number of authors have proceeded to vectorize the network predictor (originally obtained in the form of a symmetric *adjacency* matrix). Subsequently, the continuous response would be regressed on the high dimensional collection of edge weights (e.g., see Richiardi *et al.*, 2011 and Craddock *et al.*, 2009). This approach can take advantage of the recent developments in high dimensional regression, consisting of both penalized optimization (Tibshirani, 1996) and Bayesian shrinkage (Park and Casella, 2008; Carvalho *et al.*, 2010). However, this approach treats the links of the network as fully exchangeable, ignoring the fact that coefficients involving common nodes can be expected to be correlated a priori.

Recently, Arroyo Relión *et al.*, 2019 proposed a penalized optimization scheme that not only enables classification using network predictors, but also identifies important nodes and interconnections/edges between nodes. Although this model seems to perform well for prediction problems, uncertainty quantification is difficult because standard bootstrap methods are not consistent for Lasso-type methods (e.g., see Kyung *et al.*, 2010). Modifications of the bootstrap that produce well-calibrated confidence intervals in the context of standard Lasso regression have been proposed (e.g., see Chatterjee and

Lahiri, 2011), but it is not clear whether they extend to the kind of group Lasso penalties discussed in Arroyo Relión *et al.*, 2019. Viewing the network predictor matrix as a two dimensional symmetric tensor, recent developments in tensor regression (e.g., see Guhaniyogi *et al.*, 2017; Raskutti *et al.*, 2019; Fan *et al.*, 2019) are also relevant to our work. However, these approaches do not generally take into account the symmetric constraint in the network predictor matrix, tend to focus mainly on prediction and identification of important network edges, and are not specifically designed to detect important nodes impacting the response. Moreover, their specification leads to a low-rank solution of the tensor predictor coefficient, whereas the estimated network predictor coefficient from our approach does not bear any guarantee of being low-rank.

The rest of the article evolves as follows. Section 2 provides a description of the brain connectome dataset we analyze in this article. Section 3 proposes the novel network shrinkage prior and discusses posterior computation for the proposed model. Empirical investigations with various simulation studies are presented in Section 4. Section 5 analyzes the brain connectome dataset, presenting scientific findings on influential *regions of interest* (ROIs) and *edges*. Finally, Section 6 concludes the article with an eye towards future work.

2 Human Brain Network Data Description

Human creativity has been at the crux of the evolution of the human civilization, and has been the topic of research in several disciplines, including neuroscience. Though creativity can be defined in numerous ways, one could envision a creative idea as one that is unusual as well as effective in a given social context (Flaherty, 2005). Neuroscientists generally concur that a coalescence of several cognitive processes determines the creative process, which often involves a *divergence of ideas* to conceivable solutions for a given problem. To measure the creativity of an individual, Jung *et al.*, 2010 propose the *composite creativity index* (CCI), which is formulated by linking measures of divergent thinking and creative achievement to cortical thickness of young (23.7 ± 4.2 years), healthy subjects. Three independent judges grade the creative products of a subject from which the CCI is derived.

Along with CCI measurements, brain network information for n = 79 subjects is gathered using diffusion weighted magnetic resonance imaging (DWI). DWI is an imaging technique that enables measurement of the restricted diffusion of water in tissue in order to produce neural tract images. The brain imaging data we use has been pre-processed using the NDMG pre-processing pipeline (Kiar *et al.*, 2016). In the context of DWI, the human brain is divided according to the Desikan atlas (Desikan *et al.*, 2006) that identifies 34 cortical regions of interest (ROIs) in each of the left and right hemispheres of the human brain, implying 68 cortical ROIs in all. These 68 ROIs are contained in 6 *lobes* each in the left and the right hemispheres, namely the *temporal, frontal, occipital, parietal, insula* and *cingulate* lobes. A 'brain network' for each subject is represented by a 68 × 68 symmetric adjacency matrix whose rows and columns correspond to different ROIs and entries correspond to estimates of the number of ' fibers' connecting pairs of brain regions. Figure 1 shows maps of the brain network for two representative individuals in the sample.

Our goal is to predict the CCI of a subject from his/her brain network, and to identify brain regions (nodes in the brain network) that are involved with creativity, as well as influential connections between different brain regions. Along with identifying a brain region as being influential, we also aim to provide the probability of the same, which serves as a measure of uncertainty. This necessitates the development of a flexible but parsimonious Bayesian model with a network predictor and a scalar response in a network regression context. The model and prior constructions follow in the next section.

3 Model Formulation

Let $y_i \in \mathbb{R}$ and $\boldsymbol{A}_i \in \mathbb{R}^{V \times V}$ represent the observed continuous scalar response and the weighted network predictor for the i-th sample, $i = 1,\ldots,n$, respectively. It is assumed that all network predictors are defined on a common set of nodes. For example, in our brain connectome application, y_i corresponds to a phenotype (CCI), while $\boldsymbol{A}_i$ encodes the strength of the network connections between different regions of the brain for the i-

th individual. Mathematically, this amounts to $\boldsymbol{A}_i$ being a $V \times V$ matrix, with the (k, l)-th entry of $\boldsymbol{A}_i$ denoted by $a_{i,k,l} \in \mathbb{R}$. In this paper, we focus on networks that contain no self relationship, i.e., $a_{i,k,k} \equiv 0$, and are undirected ($a_{i,k,l} = a_{i,l,k}$). The brain connectome application that motivates our work naturally justifies these assumptions. However, as will become evident later, extensions to directed networks with self-relations are straightforward.

3.1 Bayesian Network Regression Model

We propose the high dimensional regression model of the response y_i for the i-th individual on the undirected network predictor $\boldsymbol{A}_i = ((a_{i,k,l}))_{k,l=1}^{V}$ as

$$y_i = \mu + \langle \boldsymbol{A}_i, \boldsymbol{B} \rangle_F + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \tau^2), \qquad (1)$$

where $\boldsymbol{B}$ is the symmetric network coefficient matrix of dimension $V \times V$ whose (k, l)-th element is given by $\gamma_{k,l}/2$ and $\langle \boldsymbol{A}_i, \boldsymbol{B} \rangle_F = Trace(\boldsymbol{B}'\boldsymbol{A}_i)$ denotes the Frobenius inner product between $\boldsymbol{A}_i$ and $\boldsymbol{B}$. The Frobenius inner product is the natural inner product in the space of matrices and is a generalization of the dot product from vector to matrix spaces. Similar to the network predictor, the network coefficient matrix $\boldsymbol{B}$ is assumed to be symmetric with zero diagonal entries. The parameter τ^2 is the variance of the observational error.

Since self relationship is absent and both $\boldsymbol{A}_i$ and $\boldsymbol{B}$ are symmetric, $\langle \boldsymbol{A}_i, \boldsymbol{B} \rangle_F = \sum_{1 \le k < l \le V} a_{i,k,l}\gamma_{k,l}$, and (1) can be rewritten as

$$y_i = \mu + \sum_{1 \le k < l \le V} a_{i,k,l}\gamma_{k,l} + \epsilon_i, \qquad \epsilon_i \sim N(0, \tau^2). \qquad (2)$$

Equation (2) connects the network regression model with the linear regression framework with $a_{i,k,l}$'s as predictors and $\gamma_{k,l}$'s as the corresponding coefficients. However, while in ordinary linear regression the predictor coefficients are indexed by the natural numbers $\mathbb{N}$, model (2) indexes the predictor coefficients by their positions in the matrix ***B***. This is done in order to keep tabs not only on the edge itself but also on the nodes connecting the edges.

3.2 Developing the Network Shrinkage Prior

3.2.1 Vector Shrinkage Prior

High dimensional regression with vector predictors has recently been of interest in Bayesian statistics. Continuous shrinkage priors, which strongly shrink coefficients corresponding to unimportant variables to zero while minimizing the shrinkage of coefficients corresponding to influential variables, have become particularly popular. Many of these priors can be expressed as a scale mixture of normal distributions, commonly referred to as *global-local scale mixtures* (Polson and Scott, 2010), that enable fast computation employing simple conjugate Gibbs sampling. More precisely, in the context of model (2), a global-local scale mixture prior would take the form $\gamma_{k,l} \sim N(0, s_{k,l}\tau^2), s_{k,l} \sim g_1, \tau^2 \sim g_2, 1 \le k < l \le V$.

Note that $s_{1,2},\ldots,s_{V-1,V}$ are local scale parameters controlling the shrinkage of the coefficients, while τ^2 is the global scale parameter. Different choices of g_1 and g_2 lead to different classes of Bayesian shrinkage priors. The Bayesian Lasso (Park and Casella, 2008) shrinkage prior takes g_1 as exponential and g_2 as the Jeffreys prior.

The direct application of this global-local prior in the context of (2) is unappealing. In practice, we expect the matrix of coefficients ***B*** (which itself can be regarded as describing a weighted network) to exhibit transitivity effects, i.e., we expect that if the interactions between regions *k* and *l* and between regions *k* and l' both influence the response, the interaction between regions *l* and l' will likely be influential (e.g., see Li *et al.*, 2013). Ordinary global-local shrinkage priors do not necessarily conform to such an important restriction.

3.2.2 Network Shrinkage Prior

We propose a shrinkage prior on the coefficients $\gamma_{k,l}$ and refer to it as the *Bayesian Network Shrinkage prior* (BNSP). The prior borrows ideas from low-order spectral representations of matrices, and aims to capture transitivity effects in the matrix of regression coefficients. Let $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_V \in \mathbb{R}^R$ be a collection of *R*-dimensional latent variables, one for each node, such that $\boldsymbol{u}_k$ corresponds to node *k*. We draw each $\gamma_{k,l}$ conditionally independent from a density that can be represented as a location and scale mixture of normals. More precisely,

$$\gamma_{k,l} \mid s_{k,l}, \boldsymbol{u}_k, \boldsymbol{u}_l, \tau^2 \sim N(\boldsymbol{u}_k{}'\boldsymbol{\Lambda}\boldsymbol{u}_l, \tau^2 s_{k,l}), \qquad s_{k,l} \sim Exp(\theta/2), \quad \theta \sim Gamma(\zeta, \iota), \tag{3}$$

where $s_{k,l}$ is the scale parameter corresponding to each $\gamma_{k,l}$, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_R)$ is an $R \times R$ diagonal matrix with $\lambda_r \in \{0,-1,1\}$.

Conditional on the latent variables $\boldsymbol{u}_k, \boldsymbol{u}_l$ and $\boldsymbol{\Lambda}$, if $s_{k,l} = 0$ then (3) implies a reduced rank-decomposition $\boldsymbol{\Gamma} = 2\boldsymbol{B} = \boldsymbol{U}'\boldsymbol{\Lambda}\boldsymbol{U}$, where $\boldsymbol{U}$ is an $R \times V$ matrix whose *k*-th column corresponds to $\boldsymbol{u}_k$ and $\boldsymbol{\Gamma} = ((\gamma_{k,l}))_{k,l=1}^{V}$. Drawing intuition from the random dot product graph (RDPG) models (Young and Scheinerman, 2007), we can interpret the latent vectors $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_V$ as the positions of the nodes in a latent space, with the strength of the edge effect being controlled by the inner product or the angular distance between the vectors. In this interpretation, $\sum_{r=1}^{R} \delta_{\lambda_r \in \{-1,1\}} = R_{eff} \le R$, represents the *effective dimensionality* of the latent space, where $\delta_{\lambda_r \in \{-1,1\}}$ takes the value 1 if $\lambda_r \in \{-1,1\}$, and is 0 otherwise. The effect of the interaction between the *k*-th and *l*-th nodes has a positive, negative or neutral impact on the response depending on whether $\boldsymbol{u}_k{}'\boldsymbol{\Lambda}\boldsymbol{u}_l > 0, \boldsymbol{u}_k{}'\boldsymbol{\Lambda}\boldsymbol{u}_l < 0$ or $\boldsymbol{u}_k{}'\boldsymbol{\Lambda}\boldsymbol{u}_l = 0$, respectively. This kind of bilinear structure is also commonly used to model social and biological networks because of its ability to capture the kind of transitive effects we discussed before (Hoff, 2005).

In order to learn which components of $\boldsymbol{u}_k$ are informative for (3), we assign a hierarchical prior

$$\lambda_r \sim \begin{cases} 0, & \text{w.p. } \pi_{1,r}, \\ 1, & \text{w.p. } \pi_{2,r}, \\ -1, & \text{w.p. } \pi_{3,r}, \end{cases} \qquad (\pi_{1,r}, \pi_{2,r}, \pi_{3,r}) \sim Dirichlet(r^{\eta}, 1, 1),\ \eta > 1.$$

The choice of hyper-parameters of the beta distribution is crucial. In particular, note that $E[\delta_{\lambda_r \in \{-1,1\}}] = 2/(2+r^{\eta}) \to 0$ as $r \to \infty$ and that $\sum_{r=1}^{R} var(\delta_{\lambda_r \in \{-1,1\}}) = \sum_{r=1}^{R} [\frac{2(r^{\eta}+1)}{(r^{\eta}+2)^2(r^{\eta}+3)} + \frac{2(r^{\eta}+1)}{(r^{\eta}+3)(r^{\eta}+4)}] < \infty$ as $R \to \infty$. The first property provides (weak) identifiability of the different latent dimensions, while the second ensures that $\lim_{R \to \infty} var(R_{eff}) < \infty$.

In fact, we can think of our model as a level-*R* truncation of an infinite dimensional model, similar in spirit to the stick-breaking construction of the Indian Buffet process (Teh *et al.*, 2007). Therefore, as long as *R* is chosen to be "large enough", the inferences will be roughly invariant to this choice. In our illustrations, we perform sensitivity analyses to determine an optimal value of *R* that maintains computational efficiency, and at the same time ensures the robustness of the results.

In order to determine which nodes are most influential in explaining the response, we assign a *spike-and-slab* mixture prior (Ishwaran and Rao, 2005) to the latent factor $\boldsymbol{u}_k$,

$$\boldsymbol{u}_k \sim \begin{cases} N(\mathbf{0}, \boldsymbol{M}), & \text{if } \xi_k = 1, \\ \delta_{\mathbf{0}}, & \text{if } \xi_k = 0, \end{cases} \qquad \xi_k \sim Ber(\Delta), \qquad (4)$$

where $\delta_{\mathbf{0}}$ is the Dirac-delta function at $\mathbf{0}$ and $\boldsymbol{M}$ is a covariance matrix of order $R \times R$. The parameter Δ corresponds to the probability of the nonzero mixture component. Note that if the *k*-th node of the network predictor is not influential in predicting the response then, a-posteriori, ξ_k should provide high probability to 0. Thus, based on the posterior probability of ξ_k, it will be possible to identify unimportant nodes, which we loosely refer to as "uninfluential nodes", in the network regression.

The rest of the hierarchy is accomplished by assigning prior distributions on $\Delta \sim Beta(a_\Delta, b_\Delta)$ and $\boldsymbol{M} \sim IW(\nu, \mathbf{I})$, where $IW(\nu, \mathbf{I})$ denotes an Inverse-Wishart distribution with identity scale matrix **I** and degrees of freedom ν. Finally, we choose a non-informative prior on (μ, τ^2) such that $p(\mu, \tau^2) \propto \frac{1}{\tau^2}$. Appendix A in the supplementary material shows the propriety of the posterior distribution under this prior. Note that integrating over the $s_{k,l}$'s alone leads to double exponential priors that are reminiscent of the Bayesian Lasso. On the other hand, while no closed form expression exists for the marginal prior after integrating over $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_V$, it is easy to see that, marginally, the edge coefficients have mean zero and are not independent. Hence, from this point of view, the latent positions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_V$ simply provide a mechanism to sparsely model the prior dependence among coefficients.

Before concluding this section, we note that the BNSP prior in (3) can be alternatively written as $\gamma_{k,l} = \boldsymbol{u}_k' \boldsymbol{\Lambda} \boldsymbol{u}_l + \tilde{\gamma}_{k,l}, \tilde{\gamma}_{k,l} \sim N(0, \tau^2 s_{k,l}), s_{k,l} \sim Exp(\theta^2/2).$ Integrating over $s_{k,l}$, the distribution of $\tilde{\gamma}_{k,l}$ becomes a double exponential distribution. Notably, the distributions of $\tilde{\gamma}_{k,l}$'s are independent over all (k, l) pairs ($k < l$) and are independent of the distribution of $\boldsymbol{u}_k$ and $\boldsymbol{u}_l$, and $\tilde{\gamma}_{k,l} \in (-\infty, \infty)$. Since there is no restriction imposed on $\boldsymbol{\Gamma} = ((\tilde{\gamma}_{k,l}/2))_{k,l=1}^V, \boldsymbol{B} = \boldsymbol{U}' \boldsymbol{\Lambda} \boldsymbol{U}/2 + \boldsymbol{\Gamma}$ *does not assume* low-rankness. Therefore, the proposed prior should ideally be able to accurately estimate the true predictor coefficient matrix even when it does not assume any specific low-rank structure.

3.3 Posterior Computation

Although summaries of the posterior distribution cannot be computed in closed form, full conditional distributions for all the parameters are available and correspond, in most cases, to standard families. Thus, posterior computation can proceed through a Markov chain Monte Carlo algorithm. We note, however, that a naive implementation of such an algorithm to update $\boldsymbol{\gamma}$, the vector composed of the upper triangular elements of $\boldsymbol{\Gamma}$, would have complexity $O(q^3)$, where $q = V(V-1)/2$. The resulting algorithm would therefore be computationally too expensive for situations such as our real data application, where V = 68 and q = 2278. To address this issue, we develop a procedure that proposes the use of the Woodbury matrix identity (Harville, 1998) to instead compute the inverse of an $n \times n$ matrix. Since in the type of applications with which this paper is concerned, n is typically much smaller than q, this approach leads to substantial computational savings that make real-life applications viable. Details of the Markov chain Monte Carlo algorithm and the efficient sampling procedure for γ are presented in Appendix B of the online supplementary material.

While inferences on the latent positions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_V$ is not our main focus, being able to visualize these positions can be helpful in terms of interpreting the model results. However, note that vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_V$ are not identifiable because the model is invariant to rotations of the latent space. Hence, before we can use the posterior samples generated by our algorithm to conduct inferences on these latent positions we must first rotate them to a common orientation. This is done using a "Procrustean" transformation (Sibson, 1978; Hoff *et al.*, 2002; Hoff, 2005). For each posterior sample $\boldsymbol{U}^{(\ell)}$ we find the rotation $\boldsymbol{U}^{(\ell)}$ that has the smallest sum of squared deviations from an arbitrary fixed reference matrix $\boldsymbol{U}_0$. This rotation is given by $\boldsymbol{U}^{(\ell)} = \boldsymbol{U}_0 \left(\boldsymbol{U}^{(\ell)}\right)' \left\{\boldsymbol{U}^{(\ell)} \boldsymbol{U}_0' \boldsymbol{U}_0 \left(\boldsymbol{U}^{(\ell)}\right)'\right\}^{-1/2} \boldsymbol{U}^{(\ell)}$. In our analysis, we use the first iterate after burn-in, $\boldsymbol{U}^{(1)}$, as the reference matrix $\boldsymbol{U}_0$. Appendix E of the online supplementary material provides an illustration of the inference on latent positions using $\boldsymbol{U}^{(\ell)}$ in one of the simulation cases.

In order to identify whether the k-th node is important in terms of predicting the response, we rely on the post burn-in L samples $\xi_k^{(1)}, \ldots, \xi_k^{(L)}$ of ξ_k. Node k is said to be influential if $\frac{1}{L}\sum_{l=1}^{L} \xi_k^{(l)} > 0.5$. To identify influential edges we utilize a modification of the algorithm proposed in Li and Pati, 2017 that allows us to estimate the false discovery rate of the procedure as a function of the number of discoveries. Details are provided in Appendix C of the online supplementary material. Finally, an estimate of $P(R_{eff} = r \mid Data)$

is given by $\frac{1}{L}\sum_{l=1}^{L} I(\sum_{m=1}^{R} \lambda_m^{(l)} = r)$, where $I(A)$ for an event A is 1 if the event A happens and 0 otherwise, and $\lambda_m^{(1)},\ldots,\lambda_m^{(L)}$ are the L post burn-in MCMC samples of λ_m.

4 Simulation Studies

This section comprehensively contrasts both the inferential and predictive performances of our proposed approach with a number of competitors in various simulation settings. As competitors, we consider both penalized likelihood methods as well as Bayesian shrinkage priors for high-dimensional regression. Our first set of competitors use generic variable selection and shrinkage methods that treat edges between nodes as " bags of predictors" and rely on high dimensional regression, thereby ignoring the relational nature of the predictor. More specifically, we use Lasso (Tibshirani, 1996), which is a popular penalized optimization scheme, and the Bayesian Lasso (Park and Casella, 2008) and Horseshoe priors (Carvalho *et al.*, 2010), which are popular Bayesian shrinkage regression methods. The Horseshoe in particular is considered to be a state-of-the-art Bayesian shrinkage prior and is known to perform well, both in sparse and not-so-sparse regression settings. We use the `glmnet` package in `R` (Friedman *et al.*, 2010) to implement Lasso regression, and the `monomvn` package in `R` (Gramacy, 2013) to implement the Bayesian Lasso (BLasso for short) and the Horseshoe. A thorough comparison with these methods will indicate the relative advantage of exploiting the structure of the network predictor.

Additionally, we compare our method to a frequentist approach that develops network regression in the presence of a network predictor and scalar response (Arroyo Relión *et al.*, 2019). To be precise, we adapt Arroyo Relión *et al.*, 2019 to a *continuous response* context and propose to estimate the network regression coefficient matrix $\boldsymbol{B}$ by solving

$$\begin{aligned}\boldsymbol{B} = \arg \min_{\boldsymbol{B}\in\ ,\boldsymbol{B}=\boldsymbol{B}',diag(\boldsymbol{B})=\boldsymbol{0}} \Big\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu - \langle \boldsymbol{A}_i, \boldsymbol{B}\rangle_F)^2 + \\ \frac{\varphi}{2}\|\boldsymbol{B}\|_F^2 + \varsigma\left(\sum_{k=1}^{V}\|\boldsymbol{B}_{(k)}\|_2 + \rho\|\boldsymbol{B}\|_1\right)\Big\},\end{aligned} \tag{5}$$

where $\|\boldsymbol{B}\|_F = \sqrt{\langle \boldsymbol{B}, \boldsymbol{B} \rangle_F}$ denotes the Frobenius norm, $\|\boldsymbol{B}\|_1$ is the sum of the absolute values of all the elements of matrix $\boldsymbol{B}$, $\|\cdot\|_2$ is the l_2 norm of a vector, $\boldsymbol{B}_{(k)}$ is the k-th row of $\boldsymbol{B}$ and φ, ρ, ς are tuning parameters. The best possible choice of the tuning parameter triplet $(\varphi, \rho, \varsigma)$ is made using cross validation over a grid of possible values. Comparison with (5) will highlight the advantages of a carefully structured Bayesian network shrinkage prior over the penalized optimization scheme incorporating network information. In the absence of open source code, we implemented the algorithm in Arroyo Relión *et al.*, 2019 ourselves. While a better and more efficient implementation of (5) is perhaps possible, our implementation provides an approximate understanding of its performance. All Bayesian competitors are allowed to draw 50, 000 MCMC samples, with posterior inference carried out on the last 20, 000 MCMC samples after suitable thinning.

4.1 Predictor and Response Data Generation

In all simulation studies, y_i is generated according to the network regression model

$$y_i = \mu_0 + \langle \boldsymbol{A}_i, \boldsymbol{B}_0 \rangle_F + \epsilon_i, \qquad \epsilon_i \sim N(0, \tau_0^2), \quad (6)$$

with τ_0^2 as the true noise variance. All simulations use V = 20 nodes and n = 70 samples.

4.1.1 Simulation 1

In this group of simulations, the (k, l)-th entry of $\boldsymbol{B}_0$ is given by $\frac{\boldsymbol{w}_k' \boldsymbol{w}_l}{2}$, where the vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_V$, each of dimension R_{gen}, are generated from a mixture

$$\boldsymbol{w}_k \sim \pi_w N_{R_{gen}}(\boldsymbol{w}_{mean}, \boldsymbol{w}_{sd}^2) + (1 - \pi_w)\boldsymbol{\delta_0}, \qquad k \in \{1, \ldots, V\}, \qquad (7)$$

where $\boldsymbol{\delta_0}$ is the Dirac-delta function and π_W is the probability of any $\boldsymbol{w}_k$ being nonzero. Since $(1 - \pi_w)$ is the probability of a node not being influential, it is referred to as the *node sparsity parameter*. This data generation mechanism is quite similar (although not identical) to our hierarchical prior. Hence, the goal of this first simulation is to evaluate the ability of the model to recover the true data-generation mechanism and, in particular, its ability to identify the true dimension of the latent space, as well as the sensitivity of the results to the choice of the maximum latent dimension *R*. The ability of the model to identify the true dimension of the latent space is assessed by looking at the posterior mode of the effective dimensionality R_{eff} (discussed in Sections 3.2.2 and 3.3). As we illustrate in Appendix D of the online supplementary material, the posterior mode of R_{eff} coincides with the true dimension R_{gen} in all cases under *Simulation 1*.

For a comprehensive picture of *Simulation 1*, we consider 10 different cases as summarized in Table 1. In each of these cases, the network predictor coefficient and the response are generated by changing the sparsity $(1-\pi_w)$ and the true dimension R_{gen} of the latent variables $\boldsymbol{w}_k$'s. The table also presents the maximum dimension *R*, used to fit the model, of the latent variables $\boldsymbol{u}_k$'s for the network regression model (2). Note that we include various cases of model mis-specification in which $R > R_{gen}$. For all simulations, $\boldsymbol{w}_{mean}$ and $\boldsymbol{w}_{sd}^2$ are set as $0.8 \times \mathbf{1}_{Rgen}$ and $\boldsymbol{I}_{R_{gen} \times R_{gen}}$, respectively, and the variance τ_0^2 is fixed at 1. In Cases 1-8, the entries of the network predictor $\boldsymbol{A}_i$ for the *i*-th sample are simulated from a standard normal distribution. In Cases 9 and 10 the network predictor $\boldsymbol{A}_i$ for the *i*-th sample follows a stochastic blockmodel. In Case 9, we assume that each brain network has 3 local clusters with high within-cluster and low between-cluster connectivity. More specifically, the matrices $\boldsymbol{A}_i$'s consist of 3 symmetric block diagonal matrices of dimensions 6 × 6, 7 × 7 and 7 × 7, respectively. Elements in these matrices are drawn from $N(j, j^2)$ 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 10, $\boldsymbol{A}_i$'s consist of 3 block diagonal matrices of dimensions 5 × 5, 8 × 8 and 7 × 7. As before, the elements in these matrices have been drawn from $N(j, j^2)$ 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 *N*(4, 1), *N*(5, 1) and *N*(6, 1).

4.1.2 Simulation 2

In this case, B_0 is constructed by first generating *V* binary indicators $\xi_1^0, \ldots, \xi_V^0$ independently from a $Ber(\pi_{2,w})$, one for each node in the network. If both $\xi_k^0 = \mathbf{1}$ and $\xi_l^0 = \mathbf{1}$, the edge coefficient connecting the *k*-th and the *l*-th nodes (*k* < *l*) is simulated from $N(0.8, 1)$. Otherwise, we set the (*k*, *l*)-th edge coefficient to be 0. Similar to *Simulation 1*, we refer to ($1 - \pi_{2,w}$) as the *node sparsity parameter*. While this simulation scenario has some similarities to our proposed model, the mean effect for influential edges is constant. Therefore, the goal of this simulation is to evaluate the performance of the model in situations where there are weak network effects in the matrix of coefficients. The network predictor A_i for the *i*-th sample is simulated by drawing $a_{i,k,l}$ independently from a *N*(0, 1) distribution for *k* < *l* and setting $a_{i,k,l} = a_{i,l,k}$ and $a_{i,k,k} = 0$ for all $k, l \in \{1, \ldots, V\}$. Finally, the variance τ_0^2 is fixed at 1 as in *Simulation 1*. Table 2 presents the two cases we consider for *Simulation 2*, which are obtained by varying the node sparsity parameter. Unlike in *Simulation 1*, the true network predictor coefficient matrix B_0 is not guaranteed to be low-rank in *Simulation 2*.

4.1.3 Simulation 3

In this case, we draw *V* indicator variables $\xi_1^0, \ldots, \xi_V^0$ from a $Ber(\pi_{2,w})$ corresponding to the *V* nodes of the network. If both $\xi_k^0 = \mathbf{1}$ and $\xi_l^0 = \mathbf{1}$, then the edge coefficient connecting the *k*-th and the *l*-th nodes ($1 \le k < l \le V$) is simulated from a mixture distribution given by $\pi_{3,w} N(0.8, 1) + (1 - \pi_{3,w})\delta_0$. Otherwise, if $\xi_k^0 = 0$ for any *k*, we set the (*k*, *l*)-th edge coefficient to be 0 for all *l*. Contrary to *Simulation 2*, *Simulation 3* allows the possibility of an edge between the *k*-th and the *l*-th nodes having no impact on the response even when both ξ_k^0 and ξ_l^0 are nonzero. In the context of *Simulation 3*, $(1 - \pi_{2,w})$ and $(1 - \pi_{3,w})$ are referred to as the *node sparsity* and the *edge sparsity* parameters, respectively. Hence, the goal of this simulation is to evaluate the impact of edge sparsity and its interaction with node sparsity on model performance in situations where there are weak network effects in the matrix of coefficients. Network predictors are randomly generated using the same mechanism as in *Simulation 2* and the true variance τ_0^2 is again fixed at 1 for all cases. Table 3 presents the four cases we consider in this evaluation, which are generated by changing the *node sparsity* and *edge sparsity*. Similar to *Simulation 2*, the true network predictor coefficient matrix $\boldsymbol{B}_0$ is not guaranteed to be low-rank in *Simulation 3*.

4.2 Results

In all simulation results shown in this section, our BNSP model is fitted with the choices of the hyper-parameters given by $\nu = 10$, $a_\Delta = 1, b_\Delta = 1$, $\zeta = 1$ and $\iota = 1$. Our extensive simulation studies reveal that both inference and prediction are robust to various choices of the hyper-parameters.

4.2.1 Identification of Influential Nodes

Figures 2 and 3 show the posterior probability of the k-th node being detected as influential, i.e., $P(\xi_k = 1|Data)$, for each node and each case within *Simulation 1*, *Simulation 2* and *Simulation 3*. In the case of *Simulation 1*, the model is able to accurately identify nodes influencing the response for any reasonable cutoff threshold. Indeed, the receiver operating characteristic (ROC) curves associated with all these simulations have areas under the curve (AUC) very close to 1. For both cases in *Simulation 2*, using our default threshold of 0.5, the model identifies all nodes to be inactive, though the performance is a bit better in Case 1 with a higher node sparsity. In particular, in Case 1, the model tends to assign lower posterior probabilities to truly non-influential nodes, and hence, the associated AUC is quite high (1.00). In *Simulation 3*, the model performs well when the node sparsity is high (AUC 1.00 for both Cases 1 and 3), and somewhat unsatisfactorily when the node sparsity is low. Furthermore, under lower node sparsity, the AUC is higher with lower edge sparsity (AUC for Case 2 and Case 4 are 0.81 and 0.57, respectively). Among the competitors, only Arroyo Relión *et al.*, 2019 allow identification of influential nodes, and when applied to these simulations, selects all nodes as significant in every case.

4.2.2 Parameter estimation

Tables 4 and 5 present the mean squared error (MSE) of all the competitors in *Simulations 1*, *2* and *3*, respectively. Given that both the fitted network regression coefficient $\boldsymbol{B}$ and the true coefficient $\boldsymbol{B}_0$ are symmetric, the MSE is calculated as $\frac{2}{V(V-1)}\sum_{k<l}(\hat{\gamma}_{k,l}-\gamma_{k,l,0})^2$, where $\hat{\gamma}_{k,l}$ is the point estimate of $\gamma_{k,l}$. For Bayesian models (including our proposed model), $\hat{\gamma}_{k,l}$ is taken to be the posterior mean of $\gamma_{k,l}$.

Table 4 shows that BNSP outperforms all its competitors in all cases of *Simulation 1*. In Cases 1-7, where the sparsity parameter is low to moderate, we perform overwhelmingly better than all the competitors. When the sparsity parameter in *Simulation 1* is high (Case 8), our simulation scheme sets a very large proportion of $\gamma_{k,l,0}$'s to zero. As a result, BNSP only slightly outperforms BLasso and Horseshoe. BNSP also shows superior performance when the network predictor has modular structure (Cases 9-10). While BNSP is expected to perform much better than BLasso, Horseshoe and Lasso due to incorporation of network information, it is important to note that the carefully chosen global-local shrinkage prior with a well formulated hierarchical mean structure seems also to outperform Arroyo Relión *et al.*, 2019, which is explicitly designed to account for the network structure.

For *Simulations 2* and *3*, Table 5 demonstrates that, when node or edge sparsity are high, BNSP is marginally outperformed by Horseshoe. This might be due to the fact that a high degree of sparsity in the edge coefficients in the truth favors ordinary high dimensional regression. As node sparsity decreases, so that more edge coefficients are nonzero in the truth and the network structure in the predictors dominates, BNSP tends to show increasing advantage in terms of estimating the network coefficient $\boldsymbol{B}$.

4.2.3 Identifying influential edges

Tables 6 and 7 show the true positive rates (TPR) and false positive rates (FPR) associated with the detection of important edges for *Simulation 1* and *Simulation 3* using BNSP, Lasso and Arroyo Relión *et al.*, 2019. Since *Simulation 3* is a more general case of *Simulation 2*, a corresponding table for *Simulation 2* is omitted in the interest of space. The results for our method are based on controlling the FDR at 0.05 using the algorithm described in Appendix C of the Supplementary material.

In *Simulation 1*, BNSP outperforms Lasso and Arroyo Relión *et al.* (2019). Lasso is competitive with BNSP in *Simulation 3*, although in this case all models tend to perform unsatisfactorily when node sparsity is low. Arroyo Relión *et al.*, 2019 identify all edges as important in all the simulation scenarios, resulting in high FPRs, although their performance could perhaps be improved with a more appropriate choice of tuning parameters rather than the choice made by a 10-fold cross-validation scheme in this article (Leng *et al.*, 2006; Meinshausen and Bühlmann, 2010). However, in the absence of any available package for our context, we refrain from a more detailed investigation of this issue as it does not constitute the main Bayesian focus of this article.

4.2.4 Predictive Inference

We compare the out-of-sample predictive ability of the different models based on the point prediction and characterization of predictive uncertainties using test samples of size $n_{pred} = 30$. To assess point prediction, we employ the mean squared prediction error (MSPE). As measures of predictive uncertainty, we provide coverage and length of 95% predictive intervals. For frequentist competitors, 95% predictive intervals are obtained by using predictive point estimates plus and minus 1.96 times standard errors.

Tables 8 and 9 show results for *Simulation 1*, *Simulation 2* and *Simulation 3*, respectively. For *Simulation 1*, BNSP clearly outperforms other competitors in terms of point prediction. Horseshoe becomes competitive in cases with a higher degree of sparsity (Cases 2 and 8). Lasso and BLasso are competitive only in Case 8, while our approach seems to dominate the method of Arroyo Relión *et al.* (2019) in all cases. In terms of prediction uncertainty, BNSP tends to generate the shortest intervals among the competitors, showing slight under-coverage only in Cases 2 and 8. As in the case of point prediction, Horseshoe seems to yield results very similar to our model in Cases 2 and 8.

In the case of *Simulations 2* and *3*, BNSP seems to outperform all other methods in terms of point prediction in situations where the node sparsity is low. Similar observations can be made with respect to the coverage of the intervals. BNSP seems to have shorter intervals and good coverage among all Bayesian competitors in Case 2 of *Simulation 2* and in Cases 2 and 4 of *Simulation 3*, making it the top performer among the Bayesian competitors. In Case 2 of *Simulation 2*, BNSP and Horseshoe are the two best performers in terms of characterizing uncertainties. For the remaining cases in *Simulations 2* and *3*, Horseshoe and BLasso also are competitive with our method. Notably under *Simulations 2* and *3*, the true coefficient matrix does not necessarily assume any low-rank structure. BNSP is a competitive performer in terms of parametric and predictive inference even under such settings.

4.2.5 Sensitivity to the choice of *R*

In order to examine the behavior of the model with increasing R, we rerun our model for each simulation scenario with R = 10, 15 and 20 (in addition to our original choice of R). For the sake of brevity, we only provide results for the data corresponding to Case 5 in *Simulation 1* (see Table 10). The behavior of all metrics is quite stable. The only values that seem to be slightly affected are the posterior means of R_{eff} and the length of the 95% credible intervals, which show little increase when we go from R = 5 to R = 20.

4.2.6 Scalability and Computation Time

Computation times (in hours) for competing methods are provided in Table 11. It may be noted that computation times for frequentist methods correspond to the time required to compute the method for each combination of tuning parameters and each fold in the cross-validation procedure. For BNSP, we provide the total time taken to run 50, 000 MCMC iterations. Since frequentist methods yield results just after a few iterations, they require substantially smaller computation times compared to BNSP.

5 Findings from Brain Connectome Data using BNSP

This section presents the analysis of the brain connectome data using BNSP and highlights the important findings. Note that the interest lies in predicting the CCI of a subject from his/her brain network, and in identifying brain regions (nodes in the brain network) that are involved with creativity, as well as influential connections between different brain regions. Before carrying out our analysis, each cell of the adjacency matrix is standardized by subtracting the mean and dividing by the standard deviation with respect to all n = 79 samples. CCI is also standardized in a similar fashion. The MCMC chain for our model is run for 50, 000 iterations, with the first 30, 000 iterations discarded as burn-in. Convergence is assessed by comparing different simulated sequences of representative parameters started at different initial values (Gelman *et al.*, 2014). We monitor the auto-correlation plots and effective sample sizes. Prior specification is identical as in the simulation studies. For the purpose of this data analysis, BNSP is fitted with R = 9. Later, we show that the results are robust to moderate increases in the value of R.

We first focus on identifying influential ROIs in the brain network. To add robustness to this process in the absence of any ground truth, BNSP has been fitted 10 times on the real data. In each fit, the kth node is identified as *influential* if $P(\xi_k = 1|Data)$ exceeds 0.5. Finally, we report those nodes to be *influential* that are found as so in more than 50% of the 10 runs. This criteria identifies 19 out of 68 ROIs as *influential*.

Of the influential ROIs, 7 belong to the left hemisphere and 12 belong to the right hemisphere. (see Table 13). Our findings coincide with results that have been previously presented in the literature. A large number of the 19 influential nodes detected by our method are part of the *frontal* (6) and *temporal* (6) cortices in both hemispheres. The frontal cortex has been scientifically associated with divergent thinking and problem solving ability, in addition to motor function, spontaneity, memory, language, initiation, judgement, impulse control and social behavior (Stuss *et al.*, 1985). Some of the other functions directly related to the frontal cortex seem to be behavioral spontaneity, interpreting environmental feedback and risk taking (Razumnikova, 2007; Miller and Milner, 1985; Kolb and Milner, 1981). Similarly, Finkelstein *et al.*, 1991 report *de novo* artistic expression to be associated with the frontal and temporal regions. Our results also show substantial overlap with those of Jung *et al.*, 2010, in which a regression model is used to understand the relationship between CCI and ROI-specific measures to account for the relationship between creativity and different brain regions. In particular, both approaches identify the *left superior temporal sulcus*, the *pars triangularis*, the *right superiorparietal*, the *orbitofrontal* and the *right posterior cingulate* regions as influencing CCI. However, although there is significant intersection between the findings of Jung *et al.*, 2010 and our method, there are a few regions that we detect as influential and they do not, and vice versa. For example, our model detects the *right paracentral* and the *precentral* regions in both the hemispheres to be significantly related to CCI, while Jung *et al.*, 2010 do not. On the other hand, they identify the *fusiform* region to be significant while we do not. Applying the method of Arroyo Relión *et al.*, 2019 to our dataset leads to the identification of 65 out of 68 ROIs as influential. The three regions that are found to be uninfluential are the *frontalpole*, *temporalpole* and the *transversetemporal* regions in the right hemisphere.

Along with influential ROIs, we are interested in identifying the statistically significant edges or connections between the 68 ROIs. Figure 4 plots the 98 interconnections that appear to be influential, controlling for a 0.05 FDR. A posteriori, the average of the mean effective dimensionality over 10 runs is 3.37.

Our interest turns now to the predictive ability of the Bayesian network regression model. Table 12 reports the average mean squared prediction error (MSPE) between observed and predicted responses, length and coverage of 95% predictive intervals for a ten-fold cross-validation exercise, over 10 independent model fits. As reference, we also present MSPE, length and coverage values for Lasso, BLasso and Arroyo Relión *et al.*, 2019. BNSP outperforms all other methods in terms of point prediction. In terms of prediction intervals, all methods perform similarly. The coverage of BNSP and Horseshoe are slightly under our target, while those of the other methods are slightly above target.

Finally, we assess the sensitivity of the model to the choice of R. Table 14 shows nearly identical results by choosing R = 9, 10, 12 and 15, suggesting that our original choice of R is sufficiently large for this application.

6 Conclusion and Future Work

This article finds regions of interest (ROI) and interconnections among ROIs in the human brain predictive of the composite creativity index (CCI) using a novel Bayesian regression framework. This framework involves CCI as the response and the brain network as the predictor. To account for the correlation in the regression coefficients that is expected from the relational nature of the brain network predictor, we carefully construct a novel class of network shrinkage priors. A notable finding of this article is that the majority of brain regions significantly related to CCI belong to the *frontal* and *temporal* lobes in both hemispheres. Further, the Bayesian framework allows careful quantification of predictive uncertainty, as well as the uncertainty related to the identification of each ROI to be influential. Empirical results from simulation studies show that our method is superior to popular alternatives in situations where the level of network node sparsity is at least moderate, and mostly competitive in other circumstances.

A number of future directions emerge from this work. There exist brain connectome applications where there is a binary response phenotype (say Intelligence Quotient or IQ), and the goal is to simultaneously classify a sample of brain networks into "high" and "low" IQ groups and identify influential ROIs and the interconnections between ROIs predictive of IQ. The methodological problem in such applications requires extending our current approach to the context of a binary response. We will explore embedding our proposed hierarchical prior into a generalized linear model and extend the computational algorithms to deal with binary regression schemes using standard latent data augmentation (Albert and Chib, 1993).

The efficient MCMC algorithm proposed here provides substantial benefits if N is small to moderate, but faces longer computation time when N is large. We plan to suitably adapt existing Bayesian methods for high dimensional regression in the presence of a large sample size and a large number of predictors in the context of our proposed network regression, with a large sample size N and a larger number of nodes V. Finally, we could replace the global-local priors we use in this paper with non-local priors (Johnson and Rossell, 2012).

Supplementary Material

Appendix A: This section shows posterior propriety of parameters in the BNSP model.

Appendix B: This section provides details of posterior computation for all the parameters.

Appendix C: This section describes the procedure for edge selection in BNSP.

Appendix D: This section presents inference on the *effective dimensionality* in BNSP.

Appendix E: This section presents inference on the *latent positions* in the BNSP model.

Acknowledgement

Sharmistha Guha was partially supported by funds from the Trinity College of Arts & Sciences at Duke University.

Abel Rodriguez was partially supported by award NSF-DMS 1738053 and 1740850.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association,* **88**(422), 669–679.

Arroyo Relión, J. D., Kessler, D., Levina, E., Taylor, S. F., *et al.* (2019). Network classification with applications to brain connectomics. *The Annals of Applied Statistics,* **13**(3), 1648–1677.

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

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

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association,* **106**(494), 608–625.

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.

De la Haye, K., Robins, G., Mohr, P., and Wilson, C. (2010). Obesity-related behaviors in adolescent friendship networks. *Social Networks,* **32**(3), 161–167.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., *et al.* (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage,* **31**(3), 968–980.

Durante, D., Dunson, D. B., *et al.* (2017). Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis.*

Fan, J., Gong, W., and Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of Econometrics.*

Finkelstein, Y., Vardi, J., and Hod, I. (1991). Impulsive artistic creativity as a presentation of transient cognitive alterations. *Behavioral Medicine,* **17**(2), 91–94.

Flaherty, A. W. (2005). Frontotemporal and dopaminergic control of idea generation and creative drive. *Journal of Comparative Neurology,* **493**(1), 147–153.

Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage,* **80**, 426–444.

Fosdick, B. K. and Hoff, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association,* **110**(511), 1047–1056.

Fowler, J. H. and Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal,* **337**.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association,* **81**(395), 832–842.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software,* **33**(1), 1–22.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis,* volume 2. CRC Press Boca Raton, FL.

Gramacy, R. B. (2013). `R` package `monomvn`.

Guhaniyogi, R. and Rodriguez, A. (2018). Joint modeling of longitudinal relational data and exogenous variables. *Bayesian Analysis.*

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *Journal of Machine Learning Research,* **18**(79), 1–31.

Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.-P., Wedeen, V. J., and Meuli, R. (2006). Understanding diffusion mr imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *Radiographics,* **26**(suppl_1), S205–S223.

Harville, D. A. (1998). Matrix Algebra from a Statistician's Perspective.

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

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association,* **97**(460), 1090–1098.

Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics,* **33**(2), 730–773.

Jbabdi, S., Sotiropoulos, S. N., Haber, S. N., Van Essen, D. C., and Behrens, T. E. (2015). Measuring macroscopic brain connections in vivo. *Nature neuroscience,* **18**(11), 1546.

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association,* **107**(498), 649–660.

Jung, R. E., Segall, J. M., Jeremy Bockholt, H., Flores, R. A., Smith, S. M., Chavez, R. S., and Haier, R. J. (2010). Neuroanatomy of creativity. *Human Brain Mapping,* **31**(3), 398–409.

Kiar, G., Gray Roncal, W., Mhembere, D., Bridgeford, E., Burns, R., and Vogelstein, J. (2016). ndmg: Neurodata's MRI graphs pipeline.

Kolb, B. and Milner, B. (1981). Performance of complex arm and facial movements after focal brain lesions. *Neuropsychologia,* **19**(4), 491–503.

Kyung, M., Gill, J., Ghosh, M., Casella, G., *et al.* (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis,* **5**(2), 369–411.

Leng, C., Lin, Y., and Wahba, G. (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica,* **16**(4), 1273.

Li, H. and Pati, D. (2017). Variable selection using shrinkage priors. *Computational Statistics & Data Analysis,* **107**, 107–119.

Li, Y., Qin, Y., Chen, X., and Li, W. (2013). Exploring the functional brain network of alzheimer?s disease: based on the computational experiment. *PloS one,* **8**(9), e73186.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **72**(4), 417–473.

Miller, L. and Milner, B. (1985). Cognitive risk-taking after frontal or temporal lobectomy-II. The synthesis of phonemic and semantic information. *Neuropsychologia,* **23**(3), 371–379.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic block structures. *Journal of the American Statistical Association,* **96**(455), 1077–1087.

Park, H.-J. and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science,* **342**(6158), 1238411.

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

Polson, N. G. and Scott, J. G. (2010). Shrink globally, Act locally: Sparse Bayesian Regularization and Prediction. *Bayesian Statistics,* **9**, 501–538.

Raskutti, G., Yuan, M., Chen, H., *et al.* (2019). Convex regularization for high-dimensional multiresponse tensor regression. *The Annals of Statistics,* **47**(3), 1554–1584.

Razumnikova, O. M. (2007). Creativity related cortex activity in the remote associates task. *Brain Research Bulletin,* **73**(1), 96–102.

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.

Shoham, D. A., Hammond, R., Rahmandad, H., Wang, Y., and Hovmand, P. (2015). Modeling social norms and social influence in obesity. *Current Epidemiology Reports,* **2**(1), 71–79.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal of the Royal Statistical Society. Series B (Methodological),* pages 234–238.

Stuss, D., Ely, P., Hugenholtz, H., Richard, M., LaRochelle, S., Poirier, C., and Bell, I. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery,* **17**(1), 41–47.

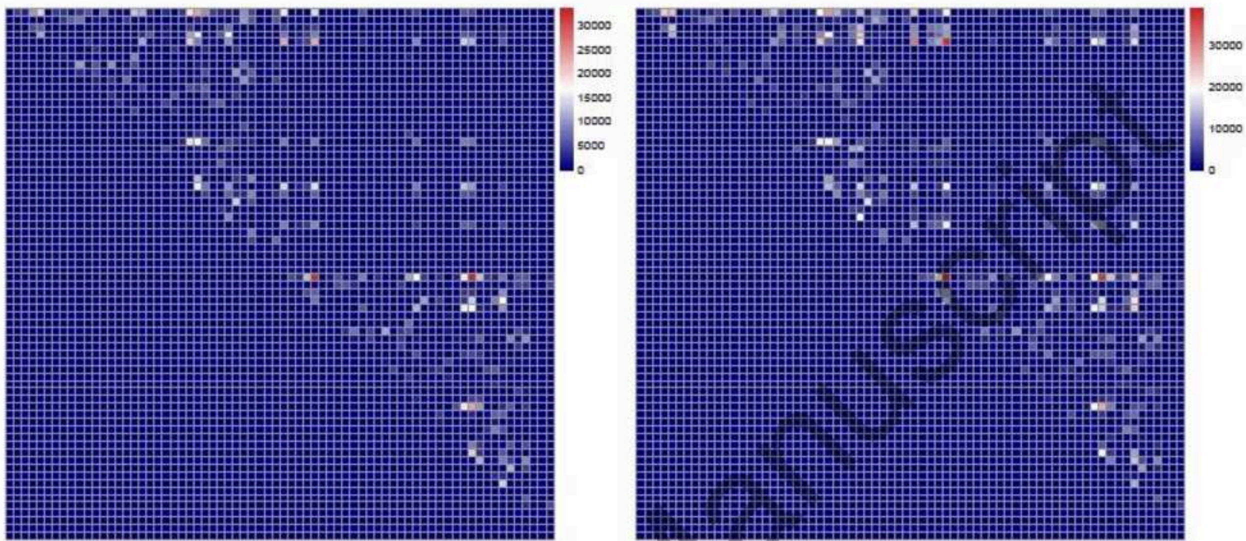
Teh, Y. W., Grür, D., and Ghahramani, Z. (2007). Stick-breaking construction for the indian buffet process. In *Artificial Intelligence and Statistics,* pages 556–563.

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

Young, S. J. and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph,* pages 138–149. Springer.

Zhang, Z., Descoteaux, M., and Dunson, D. B. (2019). Nonparametric bayes models of fiber curves connecting brain regions. *Journal of the American Statistical Association,* pages 1–23.

Fig. 1 Maps of the brain network (weighted adjacency matrices) for two representative individuals in the sample. Since the (*k*, *l*)-th off-diagonal entry in any adjacency matrix corresponds to the number of *fibers* connecting the *k*-th and the *l*-th ROIs, the adjacency matrices are symmetric. Hence the figure only shows the upper triangular portion.

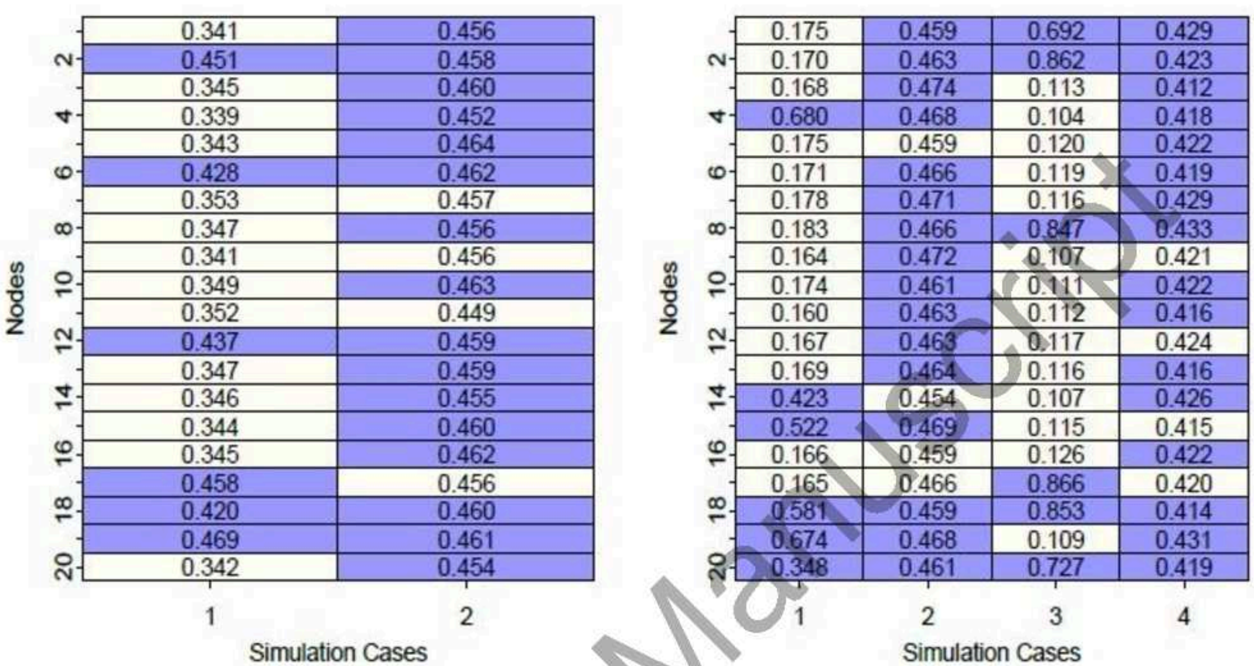


(a) Representative Network Adjacency Matrix 1 (b) Representative Network Adjacency Matrix 2

Fig. 2 Posterior probability that a node is influential, $P(\xi_k = 1|Data)$, for each node and each of the 10 cases associated with *Simulation 1*. Dark cells correspond to the truly influential nodes.

Nodes \ Simulation Cases	1	2	3	4	5	6	7	8	9	10
1	1	1	0.210	1	0.978	1	0.115	0.097	0.363	1
	1	0.031	0.175	1	1	0.019	0.117	0.112	0.374	1
3	1	1	0.175	1	0.967	0.999	0.968	0.084	0.358	0.005
	1	0.026	0.871	1	0.998	0.013	0.114	0.080	0.774	1
5	0.049	0.035	0.896	0.034	0.028	0.019	0.966	0.090	0.364	0.004
	0.787	1	0.898	0.845	0.031	0.009	0.930	0.092	0.341	0.006
7	0.036	0.089	0.174	0.012	1	1	0.967	0.922	0.755	0.211
	0.033	1	0.899	0.019	0.999	0.011	0.123	0.086	0.374	0.004
9	0.027	0.056	0.892	0.012	1	0.018	0.968	0.919	0.748	0.007
	1	0.028	0.899	1	0.036	1	0.911	0.791	0.771	0.009
11	1	0.027	0.900	1	0.024	1	0.966	0.093	0.767	1
	1	0.023	0.901	1	0.061	1	0.115	0.089	0.454	0.018
	0.867	1	0.896	0.986	0.021	1	0.741	0.087	0.381	0.022
14	0.035	1	0.892	0.018	0.047	0.021	0.184	0.087	0.769	0.008
	0.077	0.026	0.900	0.047	1	0.020	0.143	0.822	0.691	0.021
	1	1	0.186	1	1	1	0.968	0.099	0.451	1
17	0.031	0.036	0.896	0.011	0.042	1	0.966	0.917	0.767	0.020
	0.037	1	0.199	0.015	0.999	1	0.110	0.915	0.772	0.009
	0.032	0.028	0.900	0.014	0.997	1	0.107	0.085	0.374	1
20	0.045	0.031	0.895	0.022	0.034	1	0.137	0.091	0.756	0.007

Fig. 3 Posterior probability that a node is influential, $P(\xi_k = 1|Data)$, for each node and all cases associated with *Simulation 2* and *Simulation 3*. Dark cells correspond to the truly influential nodes.



(a) Simulation 2

(b) Simulation 3

Fig. 4 Significant inter-connections detected among influential brain regions of interest (ROIs) in the Desikan atlas. White cells show influential nodal associations among ROIs. Prefix 'lh-' and 'rh-' in the ROI names denote their positions in the left and right hemispheres of the brain respectively.

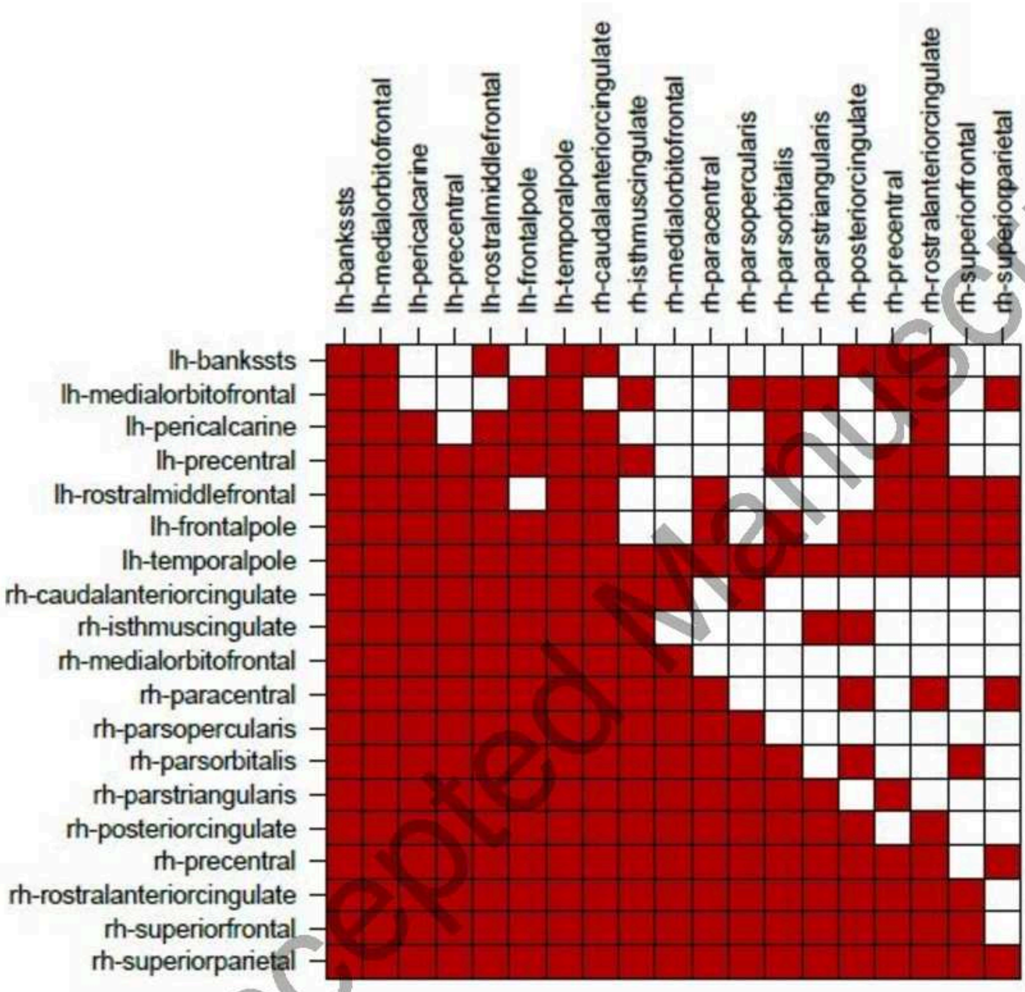


Table 1 Table presents different cases for *Simulation 1*. The true dimension R_{gen} is the dimension of vector object $\boldsymbol{w}_k$ using which data has been generated. The maximum dimension R is the dimension of vector object $\boldsymbol{u}_k$ using which the model has been fit. *Sparsity* refers to the fraction of generated $\boldsymbol{w}_k = \boldsymbol{0}$, i.e., $(1-\pi_w)$.

	Cases									
Quantity	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
R_{gen}	2	2	2	2	2	2	3	2	3	3
R	2	3	5	4	5	3	4	4	5	5
Sparsity	0.5	0.6	0.3	0.4	0.5	0.4	0.5	0.7	0.5	0.6

Table 2 Table presents different cases for *Simulation 2*. The maximum dimension R is the dimension of vector object $\boldsymbol{u}_k$ using which the model has been fit. *Simulation 2* only has one sparsity parameter $\pi_{2,w}$.

	Cases	
Quantity	Case 1	Case 2
R	5	5
Sparsity	0.7	0.2

Table 3 Table presents different cases for *Simulation 3*. The maximum dimension R is the dimension of vector object u_k using which the model has been fit. While *Simulation 2* only has a sparsity parameter, *Simulation 3* has a node sparsity ($\pi_{2,w}$) and an edge sparsity ($\pi_{3,w}$) parameter, respectively.

	Cases			
Quantity	Case 1	Case 2	Case 3	Case 4
R	5	5	5	5
Node Sparsity	0.7	0.2	0.7	0.2
Edge Sparsity	0.5	0.5	0.3	0.7

Table 4 Performance of BNSP vis-a-vis competitors for cases in *Simulation 1*. Point estimation of edge coefficients has been captured through the Mean Squared Error (MSE). The minimum MSE among competitors for any case is made bold.

	MSE									
Method	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
BNSP	**0.007**	**0.008**	**0.058**	**0.006**	**0.007**	**0.005**	**0.050**	**0.006**	**0.396**	**0.004**
Lasso	0.438	0.660	1.295	0.455	0.371	0.412	1.344	0.010	2.231	0.025
Relión(2019)	0.524	0.929	1.117	0.552	0.493	0.655	1.629	0.069	2.207	0.047
BLasso	0.472	0.863	1.060	0.465	0.699	0.464	1.638	0.008	0.751	0.018
Horseshoe	0.395	0.012	1.070	0.393	0.299	0.493	1.381	0.007	0.706	0.012

Table 5 Performance of BNSP vis-a-vis competitors for cases in *Simulation 2* and *Simulation 3*, respectively. Point estimation of edge coefficients has been captured through the Mean Squared Error (MSE). The minimum MSE among competitors for any case is made bold.

	MSE					
	Simulation 2		Simulation 3			
Method	Case 1	Case 2	Case 1	Case 2	Case 3	Case 4
BNSP	0.053	**0.732**	0.006	**0.485**	0.010	**0.128**
Lasso	0.012	0.843	0.006	0.636	**0.004**	0.178
Relión(2019)	0.036	0.859	0.017	0.617	0.036	0.145
BLasso	0.008	0.836	0.004	0.669	0.005	0.182
Horseshoe	**0.006**	0.948	**0.002**	0.629	**0.004**	0.145

Table 6 True Positive Rates (TPR) and False Positive Rates (FPR) for edges for cases in *Simulation 1*.

Method		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
BNSP	TPR	0.98	1	0.90	1	0.82	0.98	0.93	0.87	0.83	0.85
	FPR	0.05	0.02	0.02	0.08	0.06	0.08	0.06	0.01	0.02	0.06
Lasso	TPR	0.60	0.86	0.14	0.53	0.47	0.59	0.60	0.73	0.58	0.61
	FPR	0.29	0.25	0.05	0.23	0.27	0.29	0.27	0.22	0.18	0.17
Relión	TPR	1	1	1	1	1	1	1	1	1	1
	FPR	1	1	1	1	1	1	1	1	1	1

Table 7 True Positive Rates (TPR) and False Positive Rates (FPR) for edges for cases in *Simulation 3*.

Method		Case 1	Case 2	Case 3	Case 4
BNSP	TPR	0.72	0.82	1	0.91
	FPR	0.14	0.62	0.51	0.84
Lasso	TPR	0.86	0.36	0.91	0.23
	FPR	0.20	0.21	0.15	0.07
Relión	TPR	1	1	1	1
	FPR	1	1	1	1

Table 8 MSPE, coverage and length of 95% predictive intervals (PIs) under the BNSP vis-a-vis competitors for cases in *Simulation 1*. Lowest MSPE for any case is made bold.

	MSPE									
Method	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
BNSP	**0.021**	**0.003**	**0.052**	**0.011**	**0.025**	**0.006**	**0.025**	**0.041**	**0.186**	**0.037**
Lasso	0.425	0.311	1.022	0.294	0.555	0.447	0.533	0.075	0.823	4.887
Relión	0.587	0.461	0.498	0.369	0.561	0.539	0.605	0.365	0.821	0.461
BLasso	0.451	0.414	0.815	0.249	0.759	0.821	0.572	0.060	0.381	0.252
Horseshoe	0.434	0.005	0.772	0.243	0.361	0.745	0.563	0.046	0.383	0.182
	Coverage of 95% PI									
Method	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
BNSP	1.000	0.933	0.990	0.967	0.990	1.000	0.967	0.933	0.990	0.990
Lasso	1.000	0.967	0.900	1.000	0.933	0.967	0.967	1.000	0.933	0.965
Relión	0.667	0.767	0.767	0.900	0.633	0.667	0.633	0.900	0.867	0.833
BLasso	1.000	0.967	1.000	1.000	0.967	0.900	1.000	1.000	0.990	0.967
Horseshoe	0.967	1.000	1.000	1.000	1.000	0.967	0.967	0.967	0.990	0.900
	Length of 95% PI									
Method	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
BNSP	10.923	6.118	24.650	7.894	8.697	8.008	27.491	4.753	34.726	5.638
Lasso	39.732	59.778	38.259	52.177	27.43	56.922	70.655	23.964	75.309	13.216
Relión	18.348	34.138	30.126	26.848	15.115	32.434	31.746	12.122	61.221	10.419
BLasso	40.297	60.158	67.251	39.728	43.027	75.322	83.132	8.578	55.603	11.485
Horseshoe	33.489	9.366	61.534	33.529	30.132	76.089	68.103	5.846	69.886	6.618

Table 9 MSPE, coverage and length of 95% predictive intervals (PIs) under the BNSP vis-a-vis competitors for cases in *Simulation 2* and *Simulation 3*, respectively. Lowest MSPE for any case is made bold.

	MSPE					
	Simulation 2		**Simulation 3**			
Method	Case 1	Case 2	Case 1	Case 2	Case 3	Case 4
BNSP	0.519	**0.576**	0.243	**0.661**	0.148	**0.438**
Lasso	0.252	0.671	0.252	0.815	0.134	0.696
Relión	0.453	0.739	0.452	0.867	0.407	0.576
BLasso	0.153	0.722	0.220	0.875	0.124	0.781
Horseshoe	**0.122**	0.816	**0.210**	0.873	**0.119**	0.595
	Coverage of 95% PI					
	Simulation 2		**Simulation 3**			
Method	Case 1	Case 2	Case 1	Case 2	Case 3	Case 4
BNSP	0.990	0.933	1.000	1.000	0.900	0.967
Lasso	1.000	0.700	1.000	0.900	1.000	0.433
Relión	0.833	0.600	0.967	0.633	0.900	0.633
BLasso	0.933	0.833	0.933	1.000	0.933	0.867
Horseshoe	0.967	0.867	0.967	1.000	0.833	0.967
	Length of 95% PI					
	Simulation 2		**Simulation 3**			
Method	Case 1	Case 2	Case 1	Case 2	Case 3	Case 4
BNSP	14.053	32.219	6.889	33.094	10.069	18.704
Lasso	15.940	34.413	13.080	27.121	22.774	9.397
Relión	9.544	20.117	7.877	17.913	11.760	11.675
BLasso	8.957	45.959	6.508	52.087	5.980	22.049
Horseshoe	6.879	43.834	5.268	50.072	5.654	20.227

Table 10 Model behavior in terms of model performance metrics with changing values of R for data corresponding to *Simulation 1*, Case 5. We report MSE, MSPE, length and coverage of 95% predictive intervals, and the posterior mean of effective dimensionality R_{eff}.

R	MSE	MSPE	Coverage	Length of 95% PI	Posterior Mean of R_{eff}
5	0.0066	0.025	0.990	8.697	2.12
10	0.0063	0.022	0.967	7.167	2.27
15	0.0062	0.021	0.967	7.251	2.86
20	0.0063	0.019	0.990	8.014	2.79

Table 11 Computation times of competing methods for different values of sample size (N) and number of nodes (V). For the Bayesian method BNR, the table records the run time in hours for all 50, 000 MCMC iterations. The last two columns record total run time for frequentist methods with a single combination of tuning parameter values.

V	N	BNR	Lasso	Relión(2019)
20	70	0.312	0.0001	0.0005
20	100	0.363	0.0002	0.0006
20	150	0.484	0.0001	0.0016
40	70	1.595	0.0001	0.0008
40	100	2.069	0.0001	0.0008
40	150	2.876	0.0002	0.0008
60	70	7.051	0.0003	0.0016
60	100	9.412	0.0002	0.0025
60	150	13.458	0.0002	0.0026
80	70	16.541	0.0005	0.0026
80	100	25.605	0.0004	0.0011
80	150	39.68	0.0006	0.0017
100	70	47.23	0.0002	0.0011
100	100	63.41	0.0004	0.0011
100	150	75.90	0.0004	0.0023

Table 12 Predictive performance of competitors in terms of mean squared prediction error (MSPE), coverage and length of 95% predictive intervals, obtained through 10-Fold Cross Validation in the context of real data. Note that since the response has been standardized, an MSPE value greater than or around 1 will denote an inconsequential analysis.

	BNSP	Lasso	Relión(2019)	BLasso	Horseshoe
MSPE	**0.84**	0.98	0.98	1.84	1.78
Coverage of 95% PI	0.91	0.97	0.97	0.97	0.93
Length of 95% PI	3.52	3.88	3.89	3.40	4.99

Table 13 Brain regions (ROIs) detected as influential for CCI by BNSP.

Region of Interest (ROI)	Hemisphere	Lobe
bank of the superior temporal sulcus	Left	Temporal
rostral middle frontal gyrus	Left	Frontal
pericalcarine	Left	Parietal
medial orbitofrontal	Left	Temporal
precentral	Left	Parietal
frontal pole	Left	Temporal
temporal pole	Left	Temporal
medial orbitofrontal	Right	Temporal
superior parietal lobule	Right	Temporal
caudal anterior cingulate	Right	Cingulate
paracentral	Right	Frontal
isthmus cingulate cortex	Right	Occipital
pars opercularis	Right	Frontal
pars orbitalis	Right	Frontal
pars triangularis	Right	Occipital
posterior cingulate cortex	Right	Frontal
precentral	Right	Parietal
superior frontal gyrus	Right	Parietal
rostral anterior cingulate cortex	Right	Frontal

Table 14 Predictive performance of BNSP with $R = 9, 10, 12, 15$ to assess the sensitivity of predictive inference with the choice of *R*.

	BNSP (R = 9)	BNSP (R = 10)	BNSP (R = 12)	BNSP (R = 15)
MSPE	0.84	0.85	0.85	0.85
Coverage of 95% PI	0.91	0.91	0.91	0.91
Length of 95% PI	3.52	3.61	3.65	3.64
Posterior Mean of R_{eff}	3.37	3.57	3.62	3.59