A Bayesian Multiplex Graph Classifier for Investigating Functional Connectivity Across Diverse Tasks of Cognitive Control

Sharmistha Guha, Jose Rodriguez-Acosta and Ivo D. Dinov

April 22, 2024

Abstract

This article proposes a novel statistical network methodology to examine the relationship between brain stimulation and regional brain activation. Using a generalized linear modeling framework, we can predict binary outcomes, such as whether a region of interest, a node in a brain network, is activated during an external stimulus. Our proposed *Bayesian Multiplex Graph Classifier (BMGC)* model is based on multiplex graphs where each layer of the graph captures interactions among the brain network nodes. Popular regression methods utilizing multiplex graph predictors often face limitations in effectively harnessing information within and across graph layers, leading to potentially less accurate inference and predictive accuracy, especially for smaller sample sizes. To overcome this challenge, our method models edge coefficients at each network layer using bilinear interactions between the latent effects associated with the two nodes connected by the edge. This approach also employs a variable selection framework on node-specific latent effects from all networks layers to identify influential nodes linked to observed outcomes. Crucially, the proposed framework is computationally efficient and quantifies the uncertainty in node identification, coefficient estimation, and binary outcome prediction. Simulation studies demonstrate the superior inferential and BMGC performance compared against alternative methods. An additional BMGC validation was completed using an fMRI study of brain networks in adults. Specifically,

the BMGC technique identified that sensory motor brain network obeys certain lateral symmetries, whereas the default mode network exhibits significant brain asymmetries associated with early aging.

Keywords:*Aging; Bayesian Statistics; Functional Connectivity; Multiplex Graph Classification; Variable Selection*

1 Introduction

In recent years, the study of cognitive control among aging adults has gained significance as a prominent and essential research focus. With the ongoing aging of the population, there is a growing need to comprehend the nuances of cognitive functions in older individuals. This understanding is crucial not just because of the healthcare costs involved but also due to the significant economic and social consequences linked with an aging populace. The field of cognitive neuroscience stands to acquire valuable insights into the aging brain by examining functional connectivity in the human brain during cognitive control tasks. This paper utilizes brain imaging data, representing functional connectivity graphs derived from functional magnetic resonance imaging (fMRI) Richiardi et al. [2011], Dinov [2023], to investigate the interplay between cognitive decline and brain regions. The fMRI study focused on two distinct cognitive task conditions: inhibition and initiation Rieck et al. [2021a]. Inhibition is broadly conceptualized as the ability to suppress or countermand a thought, action, or feeling. Initiation, on the other hand, refers to engaging in purposeful goal-directed activities. For each participant, there are two graphs with nodes as brain regions of interest (ROIs), each corresponding to one of these task conditions. This dual perspective enables us to uncover how aging affects functional connectivity across various cognitive control scenarios, shedding light on the nuanced differences in brain network interactions between inhibition and initiation tasks in aging adults. Using the Schaefer 200 network of fMRI connectivity Yang et al. [2023] for two task conditions (Z-transformed correlation matrices *input*), we examined the left and right hemisphere network differences in the brains of normal and aging adults, as quantified by the Mini-Mental State Examination score (*outcome*).

From a statistical standpoint, the main challenge lies in devising a regression framework that incorporates a binary outcome alongside a two-layer multiplex graph as the predictor.

Our primary objective is to achieve classification of the binary outcome, identify the graph nodes significantly associated with the outcome, and estimate the regression impact of each layer of the multiplex graph on the outcome. The modeling framework proposed in this article enables us to tackle these challenges simultaneously. Although literature on generalized linear models involving a multiplex graph predictor is limited, some previous studies have addressed unsupervised modeling of multiplex graphs. However, these existing approaches do not adequately meet all our inferential objectives, as outlined below.

In regression involving a single-layer graph predictor, commonly used methods often entail transforming the graph into a high-dimensional set of edge weights [Craddock et al., 2009, Richiardi et al., 2011]. Subsequent inference is drawn utilizing developments in generic high-dimensional regression architectures [Tibshirani, 1996, Park and Casella, 2008, Carvalho et al., 2010], or neural network (NN) models [Polson and Ročková, 2018, Dinh and Ho, 2020, Dinov and Velev, 2021]. While these approaches can be extended to handle multiplex graphs straightforwardly, restructuring the multiplex graph using these methods might inadequately capture the effects of intricate interconnections among nodes within different layers on the outcome. This potential limitation could compromise the precision and interpretability of the regression model.

Recent advancements in the realm of object regressions, featuring a scalar outcome and a single-layer graph predictor, have demonstrated promising capabilities in leveraging the graph's structural information. Specifically, Guha and Rodriguez [2021, 2023] have introduced innovative Bayesian shrinkage priors designed to estimate the effects of nodes and edges within the graph on the scalar outcome. By exploiting the topology of the graph predictor, these proposed methods exhibit accurate outcome prediction, robust parameter inference, and precise uncertainty estimates, especially in identifying influential graph nodes and edges. More recently, there has been a growing body of work on interpretable graph neural network approaches that provide theoretically guaranteed predictions for scalar outcomes based on graph inputs. These approaches differ from earlier literature on graph neural networks (GNN) [Zhou et al., 2020, Liu and Zhou, 2022] in their ability to offer statistical interpretability for the constructed GNN [Zhou et al., 2023]. Despite their successes, these approaches are tailored for single-layer graph predictors. Extending them to handle interac-

tions between nodes within different layers for a multiplex graph predictor to offer inference on influential nodes presents nontrivial modeling challenges.

An alternative strategy involves the aggregation of graphs from different layers to form a tensor, which is then employed to construct a regression framework with a scalar outcome and the tensor predictor. This approach can leverage recent advancements in tensor regression, encompassing penalized optimization [Zhou et al., 2013, Fan et al., 2019], low-rank methods [He et al., 2018, Ahmed et al., 2020], and the Bayesian multiway shrinkage literature [Guhaniyogi et al., 2017, Spencer et al., 2022]. However, these approaches do not explicitly account for the symmetry constraint in individual layers of the multiplex graph. Moreover, their primary emphasis is on prediction and identifying significant edges or interconnections, rather than specifically identifying crucial nodes in each layer that influence the outcome.

Within the current body of literature, multiplex graphs are often regarded as random variables rather than being utilized as predictors in a regression framework. In this literature, researchers predominantly concentrate on establishing appropriate relationships among edges and various types of connections defining the multiple layers [Gollini and Murphy, 2016, Han et al., 2015, Heaney, 2014]. These advancements have expanded the use of exponential random graph models [Holland and Leinhardt, 1981, Frank and Strauss, 1986] and latent variable models [Nowicki and Snijders, 2001, Hoff et al., 2002, Airoldi et al., 2008], initially designed for a single graph, to facilitate inference in multiplex graphs. Furthermore, these frameworks have been extended to handle time-varying or dynamic multiplex graphs. In such scenarios, stochastic processes are employed to leverage information across edges and over time, as noted in previous research [Durante et al., 2017, Snijders et al., 2013, Hoff, 2015]. Additionally, these frameworks have also been adapted to accommodate multiplex graphs with covariates associated with the graph nodes, as demonstrated in prior studies [Contisciani et al., 2020, Zhang et al., 2022, Xu et al., 2023]. Nevertheless, these methods for multiplex graphs are fundamentally unsupervised and do not align with our specific inferential goals.

This article presents Bayesian Multiplex Graph Classifier (BMGC), which is a novel generalized linear modeling framework tailored for a binary outcome and a multiplex graph predictor. Within our framework, an edge coefficient within each layer is expressed as a func-

tion of latent effects associated with the nodes linked by that specific edge. Specifically, these edge coefficients are modeled as a bilinear interaction between node-specific latent effects, resulting in a low-rank structure for the coefficients corresponding to the multiplex graph at each layer. This low-rank structure achieves parsimony, particularly beneficial when dealing with multiplex graphs defined on a large number of nodes. To facilitate the identification of network nodes significantly associated with the outcome, we propose a variable selection framework that operates jointly on latent effects across all layers corresponding to a particular node. This structured approach not only ensures simplicity but also enables precise classification of the binary outcome, facilitates inference on network nodes and edges relevant to the outcome, and provides well-calibrated uncertainties for both inference and classification. We choose the Bayesian framework for inference due to its natural ability to provide uncertainty in inference, especially in identifying influential nodes. This is particularly essential when dealing with a moderate sample size, with the number of graph edges exceeding the sample size by a significant margin. While our research originates from a neuro-scientific study, its versatility extends beyond this domain, making it broadly applicable in various scenarios involving multiplex graphs, such as genomics data.

The rest of the article proceeds as follows. Section 2 discusses functional connectivity data during different tasks on cognitive control. Section 3 provides a detailed description of the regression framework with a multiplex graph predictor and prior formulation on model coefficients. Section 4 discusses posterior computation. Simulation studies in Section 5.1 and the neuroimaging data in Section 5.2 demonstrate superior performance of the proposed approach compared to its competitors. Finally, Section 6 offers concluding remarks and possible future extensions. Appendix A describes full conditional posterior distributions for all the model parameters. They are employed to construct Markov Chain Monte Carlo for model computation.

2 Functional Connectivity Data during Cognitive Control for an Adult Lifespan Sample

In this section, we describe the functional connectivity data, derived from functional magnetic resonance imaging (fMRI) of healthy aging adults, recorded under different cognitive tasks [Rieck et al., 2021b]. In this study, 144 individuals (ages 20–86) were recruited from the greater Toronto area to participate in a study focusing on functional activity during cognitive control. The participants underwent scanning using a Siemens 3T MRI scanner as they performed cognitive control tasks designed to gauge functional activity during inhibition and initiation. Throughout the scanning process, participants reclined on the scanner bed with a mirror positioned in front of their face, reflecting a computer screen displaying the experimental stimuli. Participants used earplugs and wore headphones to mitigate scanner noise, and communication with the experimenter between scans was facilitated through a microphone. Holding a response box in their right hand, participants used their index and middle fingers to respond to stimuli presented on the screen (specific finger responses were counterbalanced among participants). In all tasks, letters were displayed in the center of the computer screen using the Courier New font against a dark gray background. Inhibition and initiation were measured employing a go/no-go paradigm, wherein participants were exposed to a sequence of uppercase letters. Their instruction was to respond (termed "go") when encountering the letter "X" and refrain from responding (termed "no-go") for all other letters. Non-X stimuli (i.e., "no-go" stimuli) were selected randomly from a pool of 20 other letters: A, B, C, D, E, F, G, H, I, J, L, M, N, O, P, Q, R, S, T, and U. The task was divided into an "inhibition" block, characterized by a higher number of go trials than no-go trials (120 go, 40 no-go), and a shorter "initiation" block, where there were more no-go trials than go trials (20 go, 60 no-go). The sequence of these two blocks was randomized among participants. Letter stimuli were displayed for a duration of 400 ms, accompanied by an average inter-stimulus interval (marked by a fixation cross) of 1200 ms, with random jittering ranging between 900 and 1500 ms. The overall duration of the go/no-go task amounted to 6 minutes and 24 seconds. Blood-oxygen-level dependent (BOLD) fMRI data were collected with a 12-channel head coil using an echo-planar imaging sequence with 40 axial slices acquired

parallel to the anterior-posterior commissure. A total of 216 volumes were collected for the go/no-go task. High resolution anatomical scans used for warping the BOLD images to MNI space were acquired with a T1-weighted MP-RAGE sequence in which 160 axial slices were collected. Functional data for each task were preprocessed with a mix of AFNI functions as well as Octave and MATLAB scripts using the Optimizing of Preprocessing Pipelines for NeuroImaging software package (OPPNI) [Churchill et al., 2017]. For the current dataset, a fixed pipeline for all participants was conducted with the following steps: (1) rigid-body alignment of the timeseries to correct for movement; (2) removal and interpolation of outlier volumes; (3) correction for physiological (i.e., cardiac and respiratory) noise; (4) slice timing correction; (5) spatial smoothing with a 6 mm smoothing kernel; (6) temporal detrending; (7) regression of six motion parameter estimates (X, Y, and Z translation and rotation) on the timeseries; (8) regression of signal in tissue of no interest (white matter, vessels and cerebrospinal fluid) on the time-series; and finally (9) warping to MNI space and resampling to 4 mm 3 isotropic voxel.

For each participant, there are two correlation matrices corresponding to the two distinct task conditions: inhibition and initiation. Assessment of functional connectivity (quantified with time-series correlations between different brain regions) were computed using the Schaefer 200 parcel 17 network atlas [Schaefer et al., 2018, Yeo et al., 2011]. Functional connectivity was computed for the fMRI task using the CONN toolbox in MATLAB [Whitfield-Gabrieli and Nieto-Castanon, 2012]. Functional time courses for each task condition were extracted from the 200 parcel nodes that comprised 17 networks, resulting in 200×200 connectivity matrices. We used the bilateral 100 brain regions of interest (ROIs) collectively representing 200 nodes covering 17 different known brain-networks. Z-transformed functional connectivity correlation matrices were computed for these Schaefer-100 nodes to examine the underlying neurobiology of aging.

The Mini-Mental State Examination (MMSE) score [Mega et al., 1996, Moon et al., 2023] represents the *binary cognitive outcome measure* of interest we want to predict using the our novel Bayesian multiplex graph modelling framework. In this supervised prediction study, the outcome variable represents the integer-valued MMSE measure binarized into *normal* ($0 \leq \text{MMSE} \leq 29$) and *aging* ($30 = \text{MMSE}$) cohorts.

3 Model and Prior Formulation

3.1 Bayesian Multiplex Graph Classifier

For each subject, we observe a binary outcome $y \in \{0, 1\}$, a p-dimensional predictor $\boldsymbol{x}$, and a L-layer multiplex graph predictor $\{\mathcal{G}_\alpha\}_{\alpha=1}^L$. A L-layer multiplex graph is defined as a sequence of L graphs: $\{\mathcal{G}_\alpha\}_{\alpha=1}^L = \{(\mathcal{N}_\alpha, \mathcal{E}_\alpha)\}_{\alpha=1}^L$, with the αth graph consisting of nodes $\mathcal{N}_\alpha$ and edges $\mathcal{E}_\alpha \subset \mathcal{N}_\alpha \times \mathcal{N}_\alpha$. We assume that the node sets are same across different layers, i.e., $\mathcal{N}_\alpha = \mathcal{N}_{\alpha'} = \mathcal{N}$, for any $1 \leq \alpha \neq \alpha' \leq V$. This common set of nodes is denoted by $\mathcal{N} = \{\mathcal{N}^{(1)}, ..., \mathcal{N}^{(V)}\}$, where $|\mathcal{N}| = V$ represents the number of nodes. The graph predictor at the α-th layer, $\mathcal{G}_\alpha$, is represented by a $V \times V$ adjacency matrix $\boldsymbol{W}^{(\alpha)} \in \mathbb{R}^{V \times V}$, where the entry at position (v, v') signifies the strength of association between nodes $\mathcal{N}^{(v)}$ and $\mathcal{N}^{(v')}$, $v, v' = 1, ..., V$. This article concentrates specifically on undirected graphs without self-relationships within each layer, a characteristic considered scientifically meaningful in the adult lifespan data. This implies that the adjacency matrix $\mathbf{W}^{(\alpha)}$ is symmetric, and its diagonal entries are zero. In what follows, we represent the L-layer multiplex graph by the sequence of adjacency matrices $\boldsymbol{W} = \left\{\boldsymbol{W}^{(\alpha)}\right\}_{\alpha=1}^L$.

For each subject, the proposed high-dimensional generalized linear model representing the relationship between the binary outcome y and the multiplex graph predictor is given by,

$$P(y = 1|\boldsymbol{x}, \boldsymbol{W}) = G^{-1}\left(\mu + \boldsymbol{x}^T\boldsymbol{\beta}_x + \frac{1}{2}\sum_{\alpha=1}^{L}\langle \boldsymbol{W}^{(\alpha)}|\boldsymbol{\Gamma}^{(\alpha)}\rangle\right). \tag{1}$$

Here, $G(\cdot)$ represents the link function, with popular choices being logit, probit, or log-log functions. Here, μ is the intercept. The coefficient $\boldsymbol{\beta}_x$ is the p-dimensional vector corresponding to $\boldsymbol{x}$, $\boldsymbol{\Gamma}^{(\alpha)}$ is the coefficient matrix of dimension $V \times V$ corresponding to the α-th layer of the graph adjacency matrix $\boldsymbol{W}^{(\alpha)}$, and $\langle \boldsymbol{W}^{(\alpha)}|\boldsymbol{\Gamma}^{(\alpha)}\rangle = Trace(\boldsymbol{\Gamma}^{(\alpha)T}\boldsymbol{W}^{(\alpha)})$ denotes the Frobenius inner product of the matrices $\boldsymbol{W}^{(\alpha)}$ and $\boldsymbol{\Gamma}^{(\alpha)}$. The Frobenius inner product serves as an extension of the dot product, transitioning from vector spaces to matrix spaces, and it naturally represents the inner product in the space of matrices. Similar to $\boldsymbol{W}^{(\alpha)}$, we presume that the coefficient matrix $\boldsymbol{\Gamma}^{(\alpha)}$ is symmetric and has zero entries along

its diagonal.

To achieve flexible modeling of the effect of the multiplex graph on the outcome, each graph edge coefficient is expressed as an interaction between the nodes connected by the specific edge. More precisely, let $\boldsymbol{\xi}_v^{(\alpha)} \in \mathbb{R}^H$ denote the H-dimensional coordinate in the latent space, representing the latent effect of the node $\mathcal{N}^{(v)}$ specific to the αth layer, $1 \leq v \leq V$. The coefficient corresponding to the (v, v')th edge at the α-th layer is modeled as a weighted inner product of latent effects for nodes $\mathcal{N}^{(v)}$ and $\mathcal{N}^{(v')}$, given by, $\gamma_{v,v'}^{(\alpha)} = \langle \boldsymbol{\xi}_v^{(\alpha)} | \boldsymbol{\xi}_{v'}^{(\alpha)} \rangle_{\boldsymbol{\Theta}^{(\alpha)}} = \boldsymbol{\xi}_v^{(\alpha)T} \boldsymbol{\Theta}^{(\alpha)} \boldsymbol{\xi}_{v'}^{(\alpha)}$, $1 \leq v \neq v' \leq V$, $\alpha = 1, ..., L$. This mechanism allows nodes with coordinates aligned in the same direction to positively impact the outcome, whereas nodes with coordinates in opposite directions exert a negative influence on the outcome. The interaction effect is dependent on the $H \times H$ weighing matrix $\boldsymbol{\Theta}^{(\alpha)}$ which is a diagonal matrix with the h-th diagonal entry $\theta_h^{(\alpha)} \in \{-1, 0, 1\}$. The diagonal elements indicate the effect of the h-th dimension of the latent effect $\boldsymbol{\xi}_v^{(\alpha)}$ on the graph coefficients. A value of 0 for $\theta_h^{(\alpha)}$ signifies that the h-th dimension of the latent effect specific to the α-th layer is not informative of the outcome. The aforementioned modeling of the graph coefficient leads to a low-order spectral decomposition of $\boldsymbol{\Gamma}^{(\alpha)}$, given by, $\boldsymbol{\Gamma}^{(\alpha)} = \Xi^{(\alpha)T} \boldsymbol{\Theta}^{(\alpha)} \Xi^{(\alpha)}$, where $\Xi^{(\alpha)} = [\boldsymbol{\xi}_1^{(\alpha)} : \cdots : \boldsymbol{\xi}_V^{(\alpha)}]$ denotes a $H \times V$ matrix which constitutes node-specific latent effects at the αth layer. The assumed low-order spectral representation massively reduces the count of estimable parameters from $LV(V-1)/2$ to $LVH + HL = HL(V+1)$, with the typical condition that $H \ll V$.

3.2 Prior Distribution on Coefficients

Given that one of the primary focuses of inference is the identification of influential nodes within $\mathcal{N}$ for predicting the outcome, we employ a variable selection framework on the node-specific latent effects. Node $\mathcal{N}^{(v)}$ has no effect from any of the layers on the outcome if all coefficients corresponding to the edges connected to the node $\mathcal{N}^{(v)}$ are uninformative about the outcome, i.e., $\gamma_{v,v'}^{(\alpha)} = 0$ for all $v' \neq v$ and $\alpha = 1, ..., L$. Given the bi-linear representation of $\gamma_{v,v'}^{(\alpha)}$, this is achieved when the vth node specific latent effects at all layers are not informative about the outcome, which leads to $\boldsymbol{\xi}_v = (\boldsymbol{\xi}_v^{(1)T}, ..., \boldsymbol{\xi}_v^{(L)T})^T = \mathbf{0}$. In order to infer on the influential nodes, a spike-and-slab mixture distribution prior [Ishwaran and

Rao, 2005] is assigned on $\boldsymbol{\xi}_v$. More specifically, we set

$$\boldsymbol{\xi}_v \overset{iid}{\sim} \eta_v N(\mathbf{0}, \boldsymbol{K}) + (1-\eta_v)\Delta(\mathbf{0}), \;\; \eta_v \sim Ber(\delta), \tag{2}$$

where $\Delta(\mathbf{0})$ is the Dirac-delta function at $\mathbf{0}$ and $\boldsymbol{K}$ is a $LH \times LH$ covariance matrix. The parameter δ corresponds to the probability of the nonzero mixture component and η_v is a binary indicator set to 0 if $\boldsymbol{\xi}_v = \mathbf{0}$. Thus, the posterior distributions of η_v's enable identification of nodes related to the outcome. To account for multiplicity in multiple variable selection, we assign $\delta \sim Beta(a, b)$, following the popular literature [Scott and Berger, 2010]. The covariance matrix is assigned $\boldsymbol{K} \sim IW(\nu, \boldsymbol{I}_{LH})$ where $IW(\nu, \boldsymbol{I}_{LH})$ denotes an Inverse-Wishart distribution with a $LH \times LH$ identity matrix $\boldsymbol{I}_{LH}$ and degrees of freedom ν.

To learn which components of $\boldsymbol{\xi}_v^{(\alpha)}$ are informative, we assign a hierarchical prior on $\theta_h^{(\alpha)}$ given by,

$$\begin{aligned}\theta_h^{(\alpha)} &= \Delta(1)\pi_{h,1}^{(\alpha)} + \Delta(0)\pi_{h,2}^{(\alpha)} + \Delta(-1)\pi_{h,3}^{(\alpha)}, \;\; \alpha = 1, ..., L, \;\; h = 1, ..., H, \\ &(\pi_{h,1}^{(\alpha)}, \pi_{h,2}^{(\alpha)}, \pi_{h,3}^{(\alpha)}) \sim Dir(1, h^{\zeta}, 1), \;\; \zeta > 1.\end{aligned}$$

The hyper-parameters of the Dirichlet distribution are chosen to introduce increasing shrinkage on $\theta_h^{(\alpha)}$ as h grows to prevent overfitting due to choice of a large H. Specifically, $\hat{H}^{(\alpha)} = \sum_{h=1}^{H} |\theta_h^{(\alpha)}|$ estimate the dimensions of $\boldsymbol{\xi}_v^{(\alpha)}$ needed for effective modeling, also referred to as the *effective dimension* for $\boldsymbol{\xi}_v^{(\alpha)}$. The prior specification is completed by setting $\boldsymbol{\beta}_x \sim N(\mathbf{0}, \boldsymbol{I}_p)$ and setting a non-informative prior on μ.

4 Posterior Inference

Let $\boldsymbol{w}^{(\alpha)}$ and $\boldsymbol{\gamma}^{(\alpha)}$ denote the upper triangular vectors of dimension $V(V-1)/2 \times 1$ extracted from $\boldsymbol{W}^{(\alpha)}$ and $\boldsymbol{\Gamma}^{(\alpha)}$, respectively. Considering that both $\boldsymbol{w}^{(\alpha)}$ and $\boldsymbol{\gamma}^{(\alpha)}$ are symmetric matrices with zero diagonal entries, the Frobenius inner product between $\boldsymbol{W}^{(\alpha)}$ and $\boldsymbol{\Gamma}^{(\alpha)}$ can be simplified as $\langle \boldsymbol{W}^{(\alpha)} | \boldsymbol{\Gamma}^{(\alpha)} \rangle = 2\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}$. This simplification enables us to establish a connection between the modeling framework with a multiplex graph predictor in (1)

and a binary regression framework, as represented by the following equation,

$$P\left(y=1|\{\boldsymbol{w}^{(\alpha)}\}_{\alpha=1}^{L},\boldsymbol{x}\right)=G^{-1}\left(\mu+\boldsymbol{x}^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right). \tag{3}$$

This article concentrates on employing the logit link function for $G(\cdot)$ to implement Markov chain Monte Carlo (MCMC) sampler for posterior inference on parameters. We utilize the result discussed in Theorem 1 of Polson et al. [2013] to obtain

$$\begin{aligned}
p(y|\boldsymbol{x},\{\boldsymbol{w}^{(\alpha)}\}_{\alpha=1}^{L}) &= \frac{\exp\left\{y(\mu+\boldsymbol{x}^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)})\right\}}{1+\exp\left\{\mu+\boldsymbol{x}^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right\}} \\
&= \exp\left\{(y-0.5)\left(\mu+\boldsymbol{x}^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right)\right\}\times \\
&\quad \int\exp\left(-\frac{\omega}{2}\left(\mu+\boldsymbol{x}^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right)^{2}\right)p(\omega)d\omega,
\end{aligned} \tag{4}$$

where $p(\omega)$ is the density of PG(1,0) distribution. With the outcome, ordinary vector predictor and multiplex graph predictor obtained for n subjects, $\{\{y_i,\boldsymbol{x}_i,\boldsymbol{W}_i\}:i=1,...,n\}$, we use (4) and the data augmentation approach outlined in Polson et al. [2013] for effective posterior computation. For subjects $i=1,...,n$, we introduce latent variables $\omega_1,...,\omega_n$ in the likelihood. More precisely, the augmented likelihood function comes in the following form,

$$\begin{aligned}
p(\boldsymbol{y}\,|\,\boldsymbol{\beta}_{x},\boldsymbol{\Gamma}^{(1)},\cdots,\boldsymbol{\Gamma}^{(L)},\boldsymbol{\omega}) &\propto \prod_{i=1}^{n}p(y_i\,|\,\boldsymbol{\beta}_{x},\boldsymbol{\Gamma}^{(1)},\cdots,\boldsymbol{\Gamma}^{(L)},\omega_i) \\
&\propto \prod_{i=1}^{n}\exp\left\{(y_i-0.5)\left(\mu+\boldsymbol{x}_i^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}_i^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right)-\frac{\omega_i}{2}\left(\mu+\boldsymbol{x}_i^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}_i^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right)^{2}\right\} \\
&\propto \prod_{i=1}^{n}\exp\left\{-\frac{\omega_i}{2}\left[\frac{(y_i-0.5)}{\omega_i}-\left(\mu+\boldsymbol{x}_i^{T}\boldsymbol{\beta}_{x}+\sum_{\alpha=1}^{L}\boldsymbol{w}_i^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)}\right)\right]^{2}\right\}.
\end{aligned}$$

While the original conditional posterior distributions for the parameters are not available in closed forms, the augmented full conditional distributions belong to standard families.

Appendix A provides details of the full conditional distributions of the parameters.

Let $\eta_{v,1}, ..., \eta_{v,S}$ be the S post burn-in MCMC samples corresponding to η_v. The posterior probability of node $\mathcal{N}^{(v)}$ being significantly associated with the outcome can be empirically estimated as $\widehat{P(\eta_v = 1|\boldsymbol{y})} = \frac{1}{S}\sum_{s=1}^{S} \eta_{v,s}$. In what follows, we cluster logit-transformed posterior probabilities into three groups using a Bayesian mixture model [Fraley et al., 2012]. The nodes in the cluster with the highest and lowest cluster means are considered as influential and uninfluential with respect to the outcome, respectively. We classify these nodes as "activated" and "inactivated" nodes, respectively. The analysis is less conclusive about the nodes in the third cluster, which are referred to as "inconclusive." These estimated probabilities also provide insight into the uncertainty surrounding node inference, implying that nodes with posterior probabilities of being influential close to 0 or 1 are associated with less uncertainty in inference. Estimating uncertainty is vital in neuroscientific settings, particularly in presence of moderate sample sizes (n) and low signal-to-noise ratios. Moreover, the proposed framework is utilized for the classification of an unobserved binary outcome.

Let $\boldsymbol{\Gamma}_1^{(\alpha)}, ..., \boldsymbol{\Gamma}_S^{(\alpha)}$, $\boldsymbol{\beta}_{x,1}, ..., \boldsymbol{\beta}_{x,S}$ and $\mu_1, ..., \mu_S$ be the S post burn-in MCMC samples for $\boldsymbol{\Gamma}^{(\alpha)}$, $\boldsymbol{\beta}_x$ and μ, respectively, obtained after suitable burn-in and thinning. To classify an unobserved outcome $y_* \in \{0, 1\}$ associated with a newly observed L-layer multiplex graph predictor $\boldsymbol{W}_*$, and with predictor $\boldsymbol{x}_* \in \mathbb{R}^p$, we compute the estimated classification probability given a post burn-in iterate $\{\mu_s, \boldsymbol{\beta}_{x,s}, \boldsymbol{\Gamma}_s^{(1)}, ..., \boldsymbol{\Gamma}_s^{(L)}\}$, denoted by $\widehat{P(y_* = 1|\mu_s, \boldsymbol{\beta}_{x,s}, \boldsymbol{\Gamma}_s^{(\alpha)}, \boldsymbol{W}_*, \boldsymbol{x}_*)}$, from (1) using $G(\cdot)$ as the logit link function, for $s = 1, ..., S$. The unobserved outcome y_* is predicted to be 1 or 0 if $\frac{1}{S}\sum_{s=1}^{S} \widehat{P(y_* = 1|\mu_s, \boldsymbol{\beta}_{x,s}, \boldsymbol{\Gamma}_s^{(\alpha)}, \boldsymbol{W}_*, \boldsymbol{x}_*)}$ is greater than or less than a selected cut-off t_c, respectively. To judge sensitivity to the choice of the cut-off, we compute ROC curves constructed with True Positive Rates (TPR) and False Positive Rates (FPR) of classification corresponding to a range of cut-off values t_c. We present area under the ROC curve (AUC) which ranges between 0.5 to 1. A value of 0.5 signifies random classification, while 1 signifies perfect classification. A higher AUC denotes superior classification performance by a method.

5 Results

In this section, we apply the proposed *Bayesian Multiplex Graph Classifier (BMGC)* method to perform inference on nodes and classification on multiplex graphs using both simulated data and real biomedical imaging data (brain fMRI study).

5.1 Simulation Study

First, we investigate the performance of the BMGC technique through simulation and compare it against other classification methods. Our evaluation involves examining BMGC's capability to identify influential nodes and then comparing its performance in coefficient estimation and predictive inference with competing models. These competitors encompass penalized optimization, Bayesian shrinkage methods and deep neural networks. In each simulation scenario, the number of layers (L) is fixed at 2, mirroring the number of layers in the neuroimaging dataset under evaluation.

5.1.1 Simulated Data Generation

For the ith sample, $i = 1, ..., n$, we generate a multiplex graph $\boldsymbol{W}_i$ by simulating symmetric adjacency matrices $\boldsymbol{W}_i^{(1)}, ..., \boldsymbol{W}_i^{(L)}$ for L layers. The upper triangular vector $\boldsymbol{w}_i^{(\alpha)}$ of $\boldsymbol{W}_i^{(\alpha)}$ is independently drawn from the normal distribution $N(\boldsymbol{0}, \boldsymbol{I})$, and all diagonal entries of $\boldsymbol{W}_i^{(\alpha)}$ are set to 0. Let the true intercept and the true multiplex graph coefficient be given by μ_0 and $\boldsymbol{\Gamma}_0$, respectively. The response y_i for the ith sample is generated according to model (1) with $G(\cdot)$ as the logit link function. No predictor $\boldsymbol{x}_i$ is included in these simulations. In all scenarios, we set $\mu_0 = 0.5$. We use $n = 400$ samples for model fitting and reserve $n_* = 50$ samples for prediction.

Let δ_0 be the probability of a node being influential w.r.t. the outcome. We refer to $(1 - \delta_0)$ as the node sparsity. For the purpose of generating the true activation pattern of nodes in relation to the outcome, we simulate node-specific activation indicators $\eta_{1,0}, ..., \eta_{V,0} \overset{i.i.d.}{\sim} Ber(\delta_0)$. To simulate the αth layer of the multiplex graph coefficient $\boldsymbol{\Gamma}_0^{(\alpha)}$, we first simulate the LH_0-dimensional latent effects corresponding to each of the V nodes, $\boldsymbol{\xi}_{v,0} = (\boldsymbol{\xi}_{v,0}^{(1)}, ..., \boldsymbol{\xi}_{v,0}^{(L)})^T$, from $\boldsymbol{\xi}_{v,0} \overset{ind.}{\sim} \eta_{v,0} N(\boldsymbol{0}, \boldsymbol{I}_{LH_0}) + (1 - \eta_{v,0})\Delta(\boldsymbol{0})$, for $v = 1, ..., V$. Using these latent effects for nodes, the true coefficients corresponding to the (v, v')th edge at the

αth layer is constructed using the low-rank formulation $\gamma^{(\alpha)}_{v,v',0} = \boldsymbol{\xi}^{(\alpha)T}_{v,0}\boldsymbol{\xi}^{(\alpha)}_{v',0}$, for $v < v'$. We set $\gamma^{(\alpha)}_{v,v',0} = \gamma^{(\alpha)}_{v',v,0}$ and $\gamma^{(\alpha)}_{v,v,0} = 0$ to satisfy the symmetry and zero diagonals the graph coefficient at each layer, respectively. In using this setup, we assume that both the fitted and true graph coefficients have a low-rank structure. This assumption allows us to evaluate how well the proposed method captures and represents the underlying relationships between the graph nodes and the outcome variable. With this setup, we simulate data with $V = 20$, at five different different levels of node sparsity, setting $\delta_0 = 0.2, 0.3, 0.4, 0.5, 0.6$, respectively.

5.1.2 Competing Methods and Performance Metrics

We compare the performance of the proposed BMGC model vis-a-vis competitors that treat the edges between nodes in the multiplex graph predictor as a "long vector of predictors" and perform binary regression of the outcome y_i on the vector $\left(\boldsymbol{w}^{(\alpha)T}_i : \alpha = 1, ..., L\right)^T$ of dimension $LV(V-1)/2$. This approach overlooks the relational nature of the multiplex graph predictor, potentially limiting the ability of such models to capture the effect of intricate interconnections between nodes on the outcome. To this end, we employ the horseshoe prior [Carvalho et al., 2010] and the Bayesian Lasso [Park and Casella, 2008] on the regression coefficients due to their state-of-the-art empirical performance in regressions with a high-dimensional predictor vector. We refer to these competitors as BHS and BLasso, respectively, and implement them using our own codes. As a frequentist high-dimensional binary regression competitor, we adopt a penalized optimization framework with the Lasso penalty on the predictor coefficients [Tibshirani, 1996]. The Lasso is implemented using the `glmnet` (Friedman et al. [2010]) package in R, with the penalty parameter chosen through ten-fold cross-validation. Finally, we also implement a two-layer neural net (NN) as a competitor. NN models are fit using the `tensorflow` [Allaire and Chollet, 2023] and `keras` [Allaire and Tang, 2023] packages in R. The inputs to the NN models are composed of the upper-triangular entries of each graph layer, which are concatenated to create an input vector of dimension $LV(V-1)/2$. The models are fit with two hidden layers, each having the Rectified Linear Unit (ReLU) activation function [Nair and Hinton, 2010]. The first of these two hidden layers includes L_1 and L_2 regularization to perform shrinkage on the coefficients from the input layer. The output layer of each model uses the sigmoid activation function

[Narayan, 1997]. The L_1 and L_2 regularization parameters and learning rate for the models are tuned through five-fold cross validation. For training each NN, 90 percent of the training samples are used to fit the model, while 10 percent are used for validation. The batch size is equal to half of the number of samples used to fit the model. We were not able to fit a binary tensor regression competitor due to not having readily available codes.

To infer on the performance of BMGC in terms of identifying nodes significantly associated with the outcome, we present the estimated posterior probability of a node being influential across the layers, i.e., $\widehat{P(\eta_v = 1|Data)}$. The performance of the competitors in estimating multiplex graph coefficient $\mathbf{\Gamma}$ is evaluated using the scaled mean squared error (MSE). The MSE is defined as $\frac{||\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}_0||^2}{||\mathbf{\Gamma}_0||^2}$, where $\widehat{\mathbf{\Gamma}}$ represents a suitable point estimate of $\mathbf{\Gamma}$. For instance, we use the frequentist point estimate for Lasso, while BMGC, BHS and BLasso employ the posterior mean of $\mathbf{\Gamma}$. To assess predictive power of the competitors, we use the held-out test samples to compute the AUC of the predicted probabilities, $\widehat{P(y_* = 1|\boldsymbol{W}_*)}$, of each model compared against the ground truth labels, $y_{*,1}, ...y_{*,n_*}$. For the Lasso, this quantity corresponds to the point estimate from the fitted model, while for the Bayesian competitors, this is computed using the mean of the posterior predictive distribution. All comparisons presented are averaged over five simulation replicates.

5.1.3 Simulation Results

Accuracy of Influential Node Identification

Table 1 displays a matrix corresponding to five different simulation settings under different true node sparsity $(1-\delta_0)$ levels for $V = 20$. In each matrix, the colored and white cells in the lth column represent the truly influential and truly uninfluential nodes for the lth simulation scenario, respectively, for $l = 1, .., 5$. Overlaid on these matrices are the estimated posterior probabilities $\widehat{P(\eta_v = 1|Data)}$ for $v = 1, \cdots, 20$.

Table 1 shows remarkably accurate identification of truly influential nodes, with the estimated posterior probabilities of all influential nodes in most simulations being above 0.5 in all five simulation scenarios. With high-node sparsity ($\delta_0 = 0.2$), the estimated posterior probabilities tend to be very close to 1 or 0, correctly classifying nodes as influential or not, with very minimal uncertainty. However, as sparsity decreases, the uncertainty in node

identification deteriorates. In fact, when $\delta_0 = 0.6$, the estimated posterior probabilities of a few influential nodes becomes close to 0.5 and a few un-influential nodes becomes close to 0.25, displaying more uncertainty in node identification.

Node	$\delta_0 = 0.2$	$\delta_0 = 0.3$	$\delta_0 = 0.4$	$\delta_0 = 0.5$	$\delta_0 = 0.6$
1	0.9800	0.8737	0.8553	0.7837	0.6230
2	0.1003	0.3243	0.1480	0.2700	0.2357
3	0.9807	0.9283	0.8143	0.8997	0.7960
4	0.0343	0.0617	0.1280	0.2030	0.2227
5	0.0380	0.9307	0.5587	0.7590	0.8277
6	0.0250	0.1157	0.0740	0.2440	0.2267
7	0.0653	0.0330	0.0757	0.2343	0.2047
8	0.9770	0.8007	0.9263	0.6573	0.6943
9	0.0083	0.0400	0.3840	0.7753	0.3637
10	0.0090	0.8343	0.9260	0.6807	0.6743
11	0.0087	0.0937	0.8603	0.7623	0.5460
12	0.9697	0.8960	0.8763	0.9057	0.7413
13	0.0107	0.0290	0.0307	0.0950	0.0933
14	0.0047	0.0113	0.0487	0.1620	0.8287
15	0.0060	0.0170	0.0367	0.1313	0.8203
16	0.0110	0.0307	0.0223	0.8750	0.6587
17	0.0143	0.0173	0.1127	0.5250	0.7757
18	0.0077	0.0567	0.0477	0.1003	0.0993
19	0.0073	0.0580	0.1203	0.1797	0.1130
20	0.0073	0.0213	0.0720	0.0683	0.1397

Table 1: Node selection results across five simulation scenarios, each depicted in a separate column, with varying probabilities of node inclusion (δ_0). Every column displays the posterior probability of the $V = 20$ nodes in the simulation setup being active, corresponding with various levels of node-sparsity in the truth. White and colored cells denote uninfluential and influential nodes, respectively. BMGC excels in identifying nodes, as indicated by high probabilities in colored cells and low probabilities in white cells.

It is crucial to underscore that one of BMGC's key strengths in inference lies in its capability to identify significant nodes in multiplex graphs, considering the interactions across layers while quantifying uncertainty. This feature directly aligns with the inferential goals of our scientific study, which revolves around identification of brain regions of interest associated with accelerated aging. In contrast, prevalent frequentist or Bayesian high-dimensional regression techniques or neural networks lack the capability for node identification with uncertainty in the current context. For the assessment of point estimation in multiplex

graph coefficients, Table 2 showcases the scaled Mean Squared Error (MSE) for each of the competitors across the five simulation scenarios. In all instances, the results consistently highlight the superior performance of BMGC. The performance gap between BMGC and its competitors is substantial with high-node density (e.g., smaller values of δ_0). As node sparsity decreases, the performance of BMGC deteriorates and it becomes close to Lasso when $\delta_0 = 0.6$.

Method	$\delta_0 = 0.2$	$\delta_0 = 0.3$	$\delta_0 = 0.4$	$\delta_0 = 0.5$	$\delta_0 = 0.6$
BMGC	**0.9207**	**0.9693**	**0.9859**	**0.9910**	**0.9946**
Lasso	1.2514	1.0781	1.0306	1.0150	1.0082
BHS	2.6233	1.6423	1.5982	1.8199	1.8967
BLasso	1.5864	1.2707	2.5810	1.5385	1.3124

Table 2: Scaled mean squared error (MSE) of estimating multiplex graph coefficients for all competitors under all five simulation scenarios. Deep neural network (NN) only offers predictive inference, hence the results from NN are not shown in the table. The result corresponding to the best performer under all simulation scenarios is boldfaced.

To assess the out-of-sample predictive performance of the different models, Figure 1 displays the area under the receiver operating characteristic curves (AUCs) obtained using various classification thresholds t_c (as discussed in Section 4). With very high-node sparsity (small values of δ_0), the graph structure is weak, and all competitors essentially follow the behavior similar to penalized regression. Thus, with high-node sparsity, all competitors show similar out-of-sample classification performance. As sparsity decreases, classification performance of all competitors deteriorates with BMGC maintaining an edge over its competitors. This suggests that exploiting the multiplex graph structure of the predictor in BMGC can influence the accuracy of classification. Among the methods that disregard the graph structure associated with the predictor, NN exhibits inferior performance.

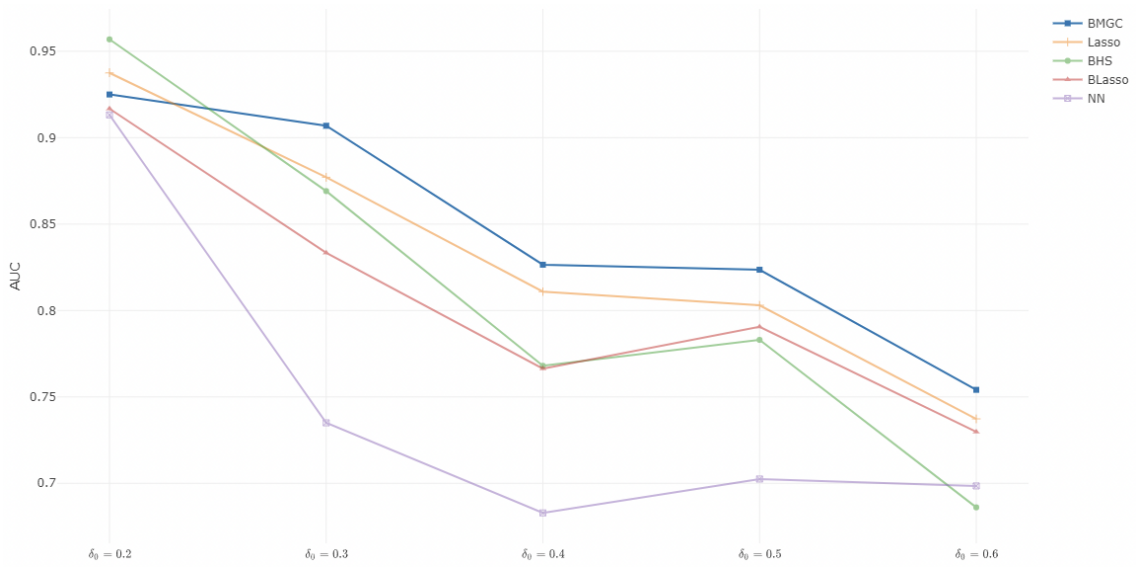

Figure 1: Comparison of the performance of the BMGC method (blue) against alternative classification techniques (different color lines) using the area under the ROC curve (AUC) across different levels of the sparsity parameter δ_0.

5.2 Functional Magnetic Resonance Imaging (fMRI) Study

Next, we will apply the BMGC technique to the functional magnetic resonance imaging (fMRI) brain dataset. Figure 2 shows a heatmap of the initial correlations between the ROI nodes in the left (columns) and right (rows) hemispheres.

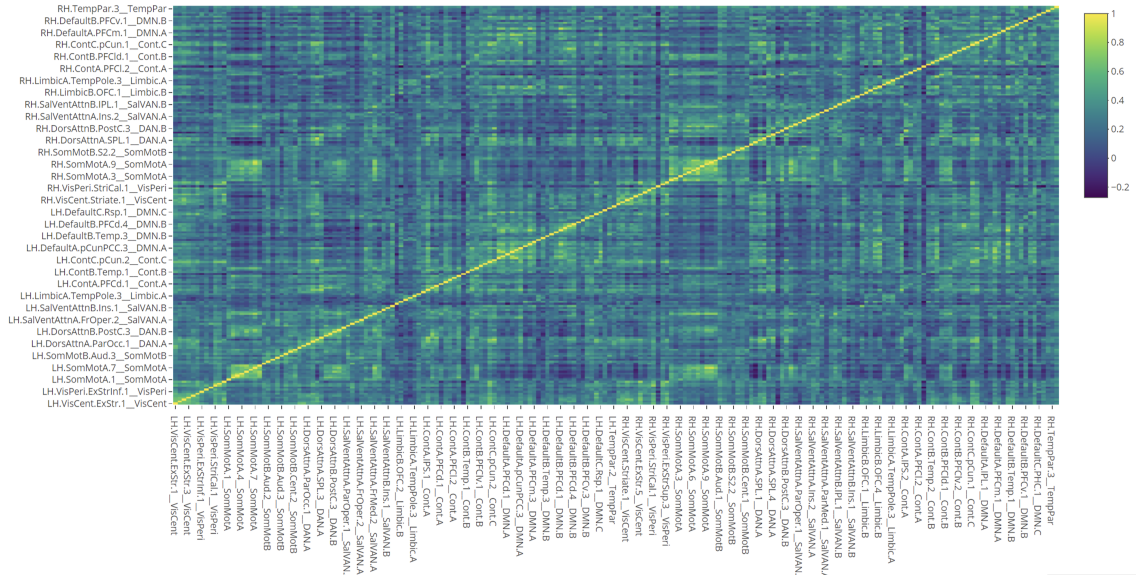

Figure 2: Heatmap of the correlations between the ROI nodes in the left and right hemispheres.

Figure 3 illustrates a partial brain connectogram indicating specific associations between some of the left hemisphere brain areas. Clearly some brain regions are highly connected, whereas other are more isolated. For simplicity, we only show a fragment of connectogram reflecting the following 10 left hemisphere (LH) *central* and *peripheral* (striatal and extrastriatal) visual cortex brain regions, see Table 3.

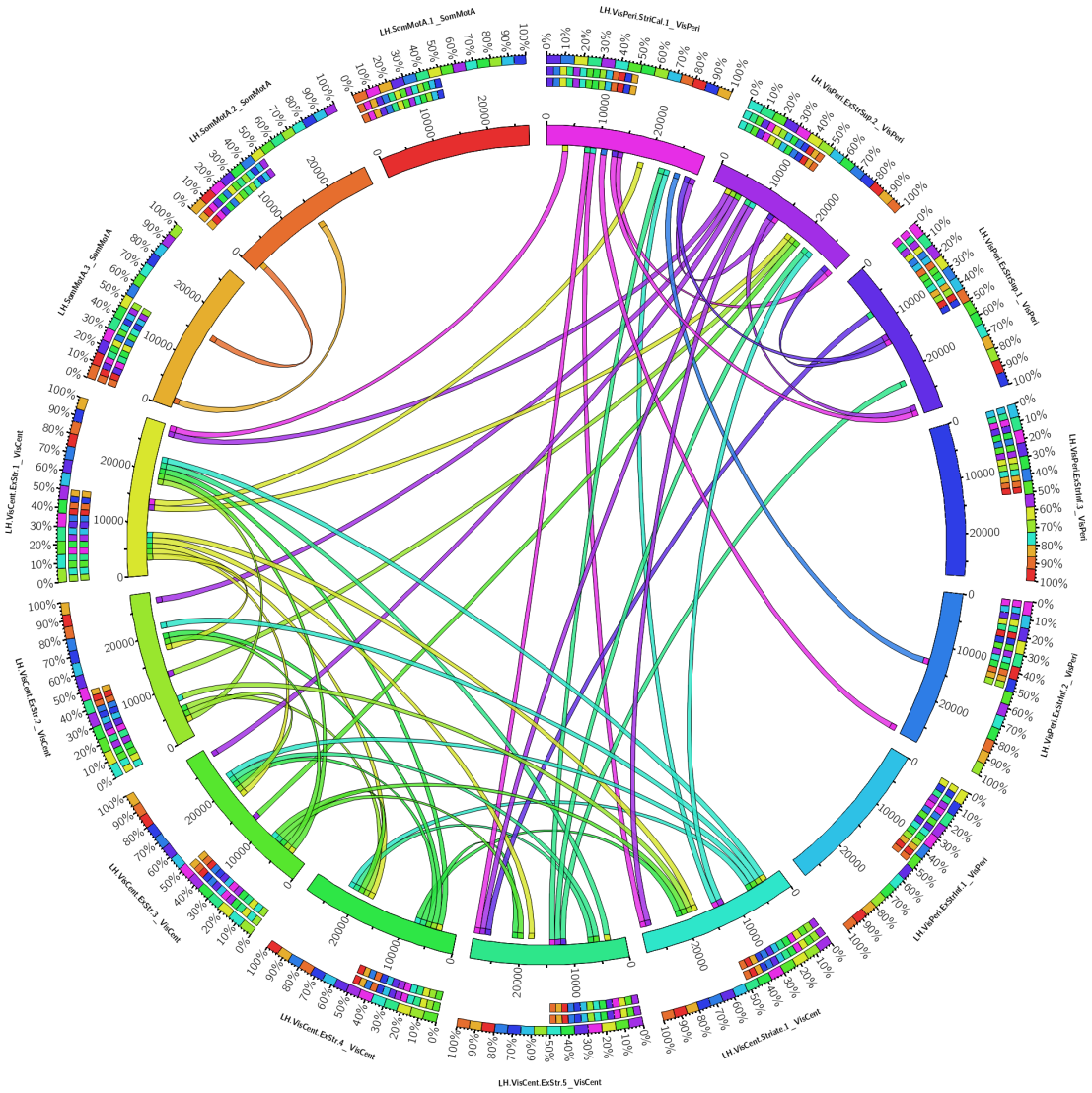

Figure 3: Connectogram showing partial correlations between several different brain regions in the left hemisphere.

Figure 4 shows a scatter plot of the *derived tripartite class labels*, {"inactivated", "inconclusive", or "activated"} computed by our algorithm. The glyph *colors* and *shapes* of

Region Acronym	Region Name
LH.VisCent.ExStr.1.VisCent	Left-Hemisphere Extrastriatal Central Visual Cortex Region 1
LH.VisCent.ExStr.2.VisCent	Left-Hemisphere Extrastriatal Central Visual Cortex Region 2
LH.VisCent.Striate.1.VisCent	Left-Hemisphere Striatal Central Visual Cortex Region 1
LH.VisCent.ExStr.3.VisCent	Left-Hemisphere Extrastriatal Central Visual Cortex Region 3
LH.VisCent.ExStr.4.VisCent	Left-Hemisphere Extrastriatal Central Visual Cortex Region 4
LH.VisCent.ExStr.5.VisCent	Left-Hemisphere Extrastriatal Central Visual Cortex Region 5
LH.VisPeri.ExStrInf.1.VisPeri	Left-Hemisphere Extrastriatal Inferior Peripheral Visual Cortex Region 1
LH.VisPeri.ExStrInf.2.VisPeri	Left-Hemisphere Extrastriatal Inferior Peripheral Visual Cortex Region 2
LH.VisPeri.ExStrInf.3.VisPeri	Left-Hemisphere Extrastriatal Inferior Peripheral Visual Cortex Region 3
LH.VisPeri.StriCal.1.VisPeri	Left-Hemisphere Striatal Caudal Peripheral Visual Cortex Region 1

Table 3: The 10 left-hemisphere (LH) *central* and (striatal and extrastriatal) *peripheral visual cortex* brain regions in Figure 3.

scatter points represent the respective probabilities of the left (x-axis) and right (y-axis) regions of interest (i.e., nodes in the brain graph network). The point 2D Cartesian coordinates represent the corresponding probabilities that the left (x-axis) and right (y-axis) brain hemispherical regions are in one of the three derived class labels. Points along near line bisecting the plane first-quadrant correspond to ROIs with identical derived class labels for the left and right region. Hence, these ROIs correspond to *brain symmetries*, i.e., bilaterally synchronously associations with memory decline (according to the MMSE outcome binary label). Conversely, scatter points off the main diagonal correspond to *brain asymmetries*; left and right brain regions whose derived labels are inconsistent.

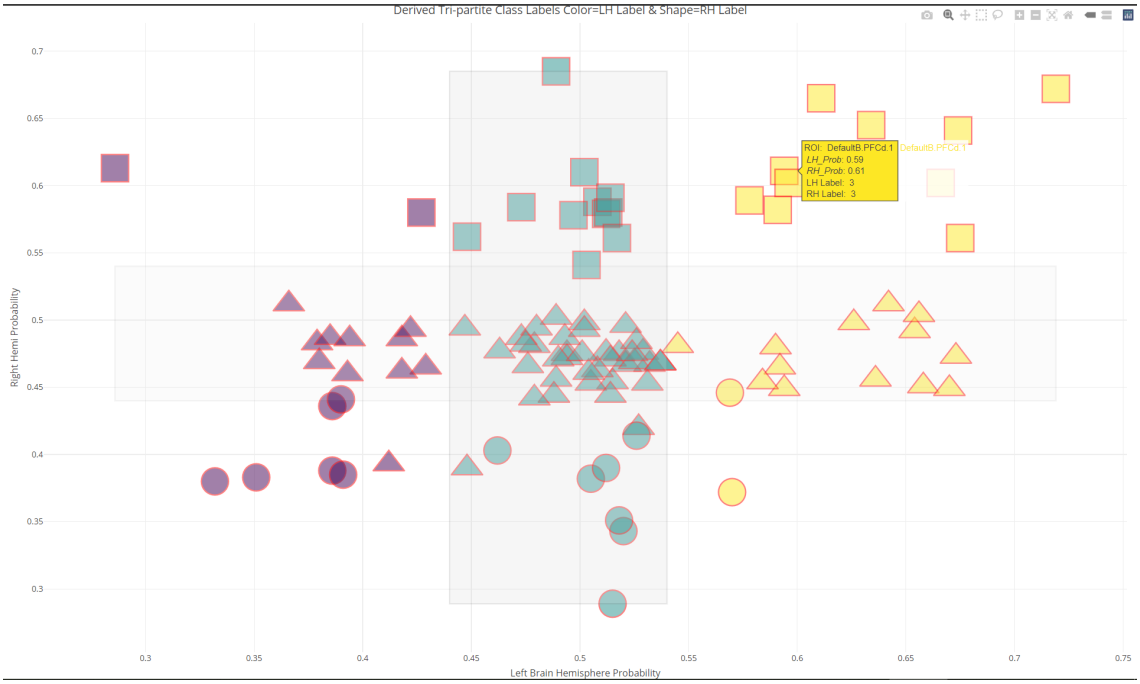

Figure 4: Bivariate plot of the derived tripartite class labels computed by the multilayer network method. The glyph colors and shapes of each scatter point represent the regions of interest (nodes in the brain graph network). The point locations represent the corresponding probabilities that the left (x-axis) and right (y-axis) brain hemispherical regions are in one of the three derived class labels ("inactivated", "inconclusive", or "activated").

Some of these brain asymmetries have been previously reported in normal aging, mild cognitive impairment, and dementia Thompson et al. [1998], Lubben et al. [2021], Mandal et al. [2012]. For instance, we observed brain symmetries involving networks with left and right *somatosensory motor area* (ROI: SomMotA.3), where both left and right probabilities are high. This indicates that in aging, sensory motor brain function obeys certain lateral symmetries. On the other hand, in the *default mode network* (ROI: DefaultC.Rsp.1) the difference in the derived class labels for the left (inactivated) and right (activated) suggest significant brain asymmetries associated with pathological aging.

In this fMRI study, to evaluate the BMGC classification performance against alternative methods, we computed the probability of the outcome being 1 (normal aging) (Jose: The way I encoded it, 1 is indicative of aging) for each sample (inputs are the individual correlation matrices of participants' brain networks). This allowed us to construct ROC curves based on pairs of True Positive Rate (TPR) and False Positive Rate (FPR) values, obtained by varying

the threshold of these estimated probabilities. The AUC estimated for each competitor corresponds to the area under their respective ROC curve. Table 4 presents a comparison of the classification performance between the new BMGC technique and various alternatives (described in Section 5.1.2).

BMGC demonstrates superior performance compared to all other competitors perhaps because it leverages the multiplex graph architecture within the modeling framework. Although Neural Networks (NN) enable capturing complex non-linear relationships between the outcome and the multiplex graph, Bayesian Horseshoe (BHS) and Bayesian Lasso (BLasso) exhibit slightly improved performance over NN, indicating the presence of sparsity in the data structure. Lasso, on the other hand, yields an AUC close to 0.5, suggesting nearly random classification of the outcome. It is worth mentioning that none of the competitors have the capability for node selection, a feature exclusive to BMGC.

Model	BMGC	Lasso	BHS	BLasso	NN
AUC	**0.6623**	0.5315	0.6035	0.5993	0.5708

Table 4: The predictive performance, as measured by the AUC, of BMGC and its competitors on the fMRI data. The best performer is highlighted in bold.

6 Conclusion and Future Work

This article introduces an innovative Bayesian framework designed to address a classification problem characterized by a binary outcome and a multiplex graph predictor, with each layer representing an undirected graph. We model an edge coefficient in a specific layer using a bilinear interaction of latent effects corresponding to nodes connected by that particular edge. A group variable selection framework is employed on node-specific latent effects for all layers to identify nodes influential in the classification of the binary outcome. A Bayesian framework is adopted to allow uncertainty quantification in inference and prediction, especially in the identification of influential nodes. Our empirical results, derived from simulation studies, establish the superiority of our method in scenarios where the regression coefficients indeed exhibit a multiplex graph structure.

Our functional magnetic resonance imaging (fMRI) study suggests that the proposed BMGC technique may be useful to identify brain-network asymmetries involving left and

right cortical regions according to their BMGC posterior predictive class probabilities. For each brain area, these likelihoods yield explicit derived class labels of association with any (binary, categorical, or continuous) clinical outcome of interest, which can be used to identify, confirm, or explore various pathological states.

As an immediate extension of the proposed framework, we aim to accommodate non-linear regression relationships between the outcome variable and the multiplex graph by incorporating a semiparametric regression framework.

Future studies may take a further step in examining the posterior probability distributions, specifically for brain asymmetry related to differences between the left and right hemispheres. Since the brain networks are the inputs, not the outputs of the BMGS model for classifying ROIs as inactive, borderline, or actively participating in the binary classification (output) of participants into normal and accelerated aging. This can facilitate drawing inference about connectivity pattern different cohorts does not appear feasible. For instance, it may be interesting to explore bilaterally the ROIs and brain networks (Schaefer-17) that are most probabilistically distinct between the two cohorts of normal (MMSE=30) and aging (MMSE $\leq$ 30).

As firm supporters of open, reproducible and trustworthy scientific discovery, we have provided data, software tools, protocols, interactive charts, and supplementary materials on these websites: `https://socr.umich.edu/docs/uploads/2024/fMRI_Corr_Pilot.html` and `https://github.com/jeroda7105/Classification-with-Multi-Layer-Graphs`.

References

T. Ahmed, H. Raja, and W. U. Bajwa. Tensor regression using low-rank and sparse tucker decompositions. *SIAM Journal on Mathematics of Data Science*, 2(4):944–966, 2020. doi: 10.1137/19M1299335.

E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(65):1981–2014, 2008.

J. Allaire and F. Chollet. *keras: R Interface to 'Keras'*, 2023. URL `https://tensorflow.rstudio.com/`. R package version 2.11.1.

J. Allaire and Y. Tang. *tensorflow: R Interface to 'TensorFlow'*, 2023. URL `https://github.com/rstudio/tensorflow`. R package version 2.11.0.9000.

C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. ISSN 00063444, 14643510.

N. W. Churchill, P. Raamana, R. Spring, and S. C. Strother. Optimizing fmri preprocessing pipelines for block-design tasks as a function of age. *NeuroImage*, 154:240–254, 2017.

M. Contisciani, E. A. Power, and C. De Bacco. Community detection with node attributes in multilayer networks. *Scientific Reports*, 10(1), 2020. doi: 10.1038/s41598-020-72626-y.

R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, 62(6): 1619–1628, 2009. doi: 10.1002/mrm.22159.

V. C. Dinh and L. S. Ho. Consistent feature selection for analytic deep neural networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 2420–2431, 2020.

I. D. Dinov. Springer International Publishing, Cham, 2023. ISBN 978-3-031-17483-4. doi: 10.1007/978-3-031-17483-4_2. URL `https://doi.org/10.1007/978-3-031-17483-4_2`.

I. D. Dinov and M. V. Velev. Data science: Time complexity, inferential uncertainty, and spacekime analytics. 2021.

D. Durante, N. Mukherjee, and R. C. Steorts. Bayesian learning of dynamic multilayer networks. *Journal of Machine Learning Research*, 18(43):1–29, 2017. URL `http://jmlr.org/papers/v18/16-391.html`.

J. Fan, W. Gong, and Z. Zhu. Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of Econometrics*, 212(1):177–202, 2019. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2019.04.026. Big Data in Dynamic Predictive Econometric Modeling.

C. Fraley, A. E. Raftery, T. B. Murphy, and L. Scrucca. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical report, Citeseer, 2012.

O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986. doi: 10.1080/01621459.1986.10478342.

J. Friedman, R. Tibshirani, and T. Hastie. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. doi: 10.18637/jss.v033.i01.

I. Gollini and T. B. Murphy. Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265, 2016. ISSN 10618600, 15372715.

S. Guha and A. Rodriguez. Bayesian regression with undirected network predictors with an application to brain connectome data. *Journal of the American Statistical Association*, 116(534):581–593, 2021. doi: 10.1080/01621459.2020.1772079.

S. Guha and A. Rodriguez. High-Dimensional Bayesian Network Classification with Network Global-Local Shrinkage Priors. *Bayesian Analysis*, pages 1 – 30, 2023. doi: 10.1214/23-BA1378.

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

Q. Han, K. S. Xu, and E. M. Airoldi. Consistent estimation of dynamic and multi-layer block models. *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, 2015.

L. He, K. Chen, W. Xu, J. Zhou, and F. Wang. Advances in Neural Information Processing Systems. 31, 2018.

M. T. Heaney. Multiplex networks and interest group influence reputation: An exponential random graph model. *Social Networks*, 36:66–81, 2014. ISSN 0378-8733. doi: https://doi.org/10.1016/j.socnet.2012.11.003. URL `https://www.sciencedirect.com/science/article/pii/S0378873312000664`. Special Issue on Political Networks.

P. D. Hoff. Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics*, 9(3), 2015. doi: 10.1214/15-aoas839.

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002. doi: 10.1198/016214502388618906.

P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981. doi: 10.1080/01621459.1981.10477598.

H. Ishwaran and J. S. Rao. Spike and slab variable selection: Frequentist and Bayesian strategies. 2005.

Z. Liu and J. Zhou. Introduction to Graph Neural Networks. *Springer Nature*, 2022.

N. Lubben, E. Ensink, G. A. Coetzee, and V. Labrie. The enigma and implications of brain hemispheric asymmetry in neurodegenerative diseases. *Brain Communications*, 3 (3):fcab211, 2021.

P. K. Mandal, R. Mahajan, and I. D. Dinov. Structural brain atlases: design, rationale, and applications in normal and pathological cohorts. *Journal of Alzheimer's Disease*, 31(s3): S169–S188, 2012.

M. S. Mega, J. L. Cummings, T. Fiorello, and J. Gornbein. The spectrum of behavioral changes in Alzheimer's disease. *Neurology*, 46(1):130–135, 1996.

S. W. Moon, L. Zhao, W. Matloff, S. Hobel, R. Berger, D. Kwon, J. Kim, A. W. Toga, I. D. Dinov, and for the Alzheimer's Disease Neuroimaging Initiative. Brain structure and allelic associations in Alzheimer's disease. *CNS Neuroscience & Therapeutics*, 29(4):1034–1048, 2023. doi: https://doi.org/10.1111/cns.14073. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/cns.14073`.

V. Nair and G. E. Hinton. Rectified linear units improve restricted Boltzmann machines. ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress. ISBN 9781605589077.

S. Narayan. The generalized sigmoid activation function: Competitive supervised learning. *Information Sciences*, 99(1):69–82, 1997. ISSN 0020-0255. doi: https://doi.org/10.1016/S0020-0255(96)00200-9. URL `https://www.sciencedirect.com/science/article/pii/S0020025596002009`.

K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001. doi: 10.1198/016214501753208735.

T. Park and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008. doi: 10.1198/016214508000000337.

N. G. Polson and V. Ročková. Posterior concentration for sparse deep learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 938–949, Red Hook, NY, USA, 2018. Curran Associates Inc.

N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504): 1339–1349, 2013. doi: 10.1080/01621459.2013.829001.

J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville. Decoding brain states from fmri connectivity graphs. *NeuroImage*, 56(2):616–626, 2011. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2010.05.081. Multivariate Decoding and Brain Reading.

J. R. Rieck, G. Baracchini, D. Nichol, H. Abdi, and C. L. Grady. Dataset of functional connectivity during cognitive control for an adult lifespan sample. *Data in Brief*, 39: 107573, 2021a. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2021.107573. URL `https://www.sciencedirect.com/science/article/pii/S2352340921008489`.

J. R. Rieck, G. Baracchini, D. Nichol, H. Abdi, and C. L. Grady. Reconfiguration and dedifferentiation of functional networks during cognitive control across the adult lifespan. *Neurobiology of Aging*, 106:80–94, 2021b.

A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

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

T. A. Snijders, A. Lomi, and V. J. Torló. A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, 35(2):265–276, 2013. doi: 10.1016/j.socnet.2012.05.005.

D. Spencer, R. Guhaniyogi, R. Shinohara, and R. Prado. Bayesian tensor regression using the tucker decomposition for sparse spatial modeling, 2022.

P. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, A. Khan, M. Mega, G. Small, J. Cummings, and A. Toga. Cortical variability and asymmetry in normal aging and alzheimer's disease. *Cerebral Cortex (New York, NY: 1991)*, 8(6):492–509, 1998.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246.

S. Whitfield-Gabrieli and A. Nieto-Castanon. Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141, 2012.

S. Xu, Y. Zhen, and J. Wang. Covariate-assisted community detection in multi-layer networks. *Journal of Business & Economic Statistics*, 41(3):915–926, 2023. doi: 10.1080/07350015.2022.2085726.

Y. Yang, C. Ye, and T. Ma. A deep connectome learning network using graph convolution for connectome-disease association study. *Neural Networks*, 164:91–104, 2023.

B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 2011.

X. Zhang, G. Xu, and J. Zhu. Joint latent space models for network data with high-dimensional node variables. *Biometrika*, 109(3):707–720, 2022. doi: 10.1093/biomet/asab063.

H. Zhou, L. Li, and H. Zhu. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552, 2013. doi: 10.1080/01621459.2013.776499. PMID: 24791032.

J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun. Graph Neural Networks: A Review of Methods and Applications. *AI open*, 1:57–81, 2020.

W. Zhou, A. Qu, K. W. Cooper, N. Fortin, and B. Shahbaba. A model-agnostic graph neural network for integrating local and global information. *arXiv preprint arXiv:2309.13459*, 2023.

Appendix A

Below we present the full conditional posterior distributions for all parameters within the generalized linear model detailed in Section 3. These distributions are utilized for implementing a Markov Chain Monte Carlo (MCMC) algorithm via Gibbs sampling. The samples derived from the MCMC algorithm yield samples from the complete posterior distributions for the model parameters jointly, facilitating posterior inference as elaborated in Section 4. Let $\kappa_i = (y_i - 0.5)/\omega_i$, and $\boldsymbol{\Omega} = diag\left(\frac{1}{\omega_1}, ..., \frac{1}{\omega_n}\right)$. Assuming $\boldsymbol{X}$ to be an $n \times p$ matrix with its ith row as $\boldsymbol{x}_i$, and $\boldsymbol{A}^{(\alpha)}$ to be an $n \times V(V-1)/2$ dimensional matrix with its ith row given by $\boldsymbol{w}_i^{(\alpha)}$. Consequently, the full conditional distributions for the model parameters are expressed as:

- $\mu|- \sim N\left(\frac{\mathbf{1}_n^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\kappa}-\boldsymbol{X}\boldsymbol{\beta}_x-\sum_{\alpha=1}^{L}\boldsymbol{A}^{(\alpha)}\boldsymbol{\gamma}^{(\alpha)})}{\mathbf{1}_n^T\boldsymbol{\Omega}^{-1}\mathbf{1}_n}, \frac{1}{\mathbf{1}_n^T\boldsymbol{\Omega}^{-1}\mathbf{1}_n}\right)$, where $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_n)^T$ is an n-dimensional vector of continuous outcomes over all samples.

- $\boldsymbol{\beta}_x|- \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = (\boldsymbol{X}^T\boldsymbol{\Omega}^{-1}\boldsymbol{X} + \boldsymbol{I}_p)^{-1}, \ \boldsymbol{\mu}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{X}^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\kappa} - \mu\mathbf{1}_n - \sum_{l=1}^{L}\boldsymbol{A}^{(\alpha)}\boldsymbol{\gamma}^{(\alpha)}).$$

- $\boldsymbol{\xi}_v|- \sim \eta_v N(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v) + (1-\eta_v)\Delta(\mathbf{0})$, where

$$\boldsymbol{\Sigma}_v = (\boldsymbol{Z}_v^T\boldsymbol{\Omega}^{-1}\boldsymbol{Z}_v + \boldsymbol{K}^{-1})^{-1}, \ \boldsymbol{\mu}_v = \boldsymbol{\Sigma}_v\boldsymbol{Z}_v^T\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{\kappa}}_v.$$

  Here $\boldsymbol{Z}_v$ is an $n \times LH$ matrix whose ith row is given by

$$\left(\sum_{k<v} w_{i,(k,v)}^{(1)}\boldsymbol{\xi}_k^{(1)T}\boldsymbol{\Theta}^{(1)} + \sum_{v<k} w_{i,(v,k)}^{(1)}\boldsymbol{\xi}_k^{(1)T}\boldsymbol{\Theta}^{(1)}, ..., \sum_{k<v} w_{i,(k,v)}^{(L)}\boldsymbol{\xi}_k^{(L)T}\boldsymbol{\Theta}^{(L)} + \sum_{v<k} w_{i,(v,k)}^{(L)}\boldsymbol{\xi}_k^{(L)T}\boldsymbol{\Theta}^{(L)}\right)$$

  and $\tilde{\boldsymbol{\kappa}}_v$ is an n dimensional vector with its ith entry

$$\tilde{\kappa}_i = \kappa_i - \mu - \boldsymbol{X}\boldsymbol{\beta}_x - \sum_{k<k':k,k'\neq v}\sum_{\alpha=1}^{L} w_{i,(k,k')}^{(\alpha)}\boldsymbol{\xi}_k^{(\alpha)T}\boldsymbol{\Theta}^{(\alpha)}\boldsymbol{\xi}_{k'}^{(\alpha)} \ .$$

- $\eta_v|- \sim Ber(\tilde{\eta}_v)$, where

$$\tilde{\eta}_v = \frac{\delta N(\tilde{\boldsymbol{\kappa}}_v|\mathbf{0}, \boldsymbol{Z}_v\boldsymbol{K}\boldsymbol{Z}_v^T + \boldsymbol{\Omega})}{\delta N(\tilde{\boldsymbol{\kappa}}_v|\mathbf{0}, \boldsymbol{Z}_v\boldsymbol{K}\boldsymbol{Z}_v^T + \boldsymbol{\Omega}) + (1-\delta)N(\tilde{\boldsymbol{\kappa}}_v|\mathbf{0}, \boldsymbol{\Omega})}.$$

- $\delta|- \sim Beta(a + \sum_{v=1}^{V}\eta_v, b + V - \sum_{v=1}^{V}\eta_v)$.

- $\boldsymbol{K}|- \sim IW\left(\nu + \#\{v : \eta_v = 1\}, \boldsymbol{I}_{LH} + \sum_{\{v:\eta_v=1\}}\boldsymbol{u}_v\boldsymbol{u}_v^T\right)$, where $\boldsymbol{u}_v = (\boldsymbol{u}_v^{(1)T}, ..., \boldsymbol{u}_v^{(L)T})^T$.

- $$\theta_h^{(\alpha)} \mid - \sim \begin{cases} 1 & w.p.\ p_{h,1}^{(\alpha)} \\ 0 & w.p.\ p_{h,2}^{(\alpha)} \\ -1 & w.p.\ p_{h,3}^{(\alpha)} \end{cases}$$

  where

  $p_{h,1}^{(\alpha)} = \dfrac{\pi_{h,1}^{(\alpha)} N(\boldsymbol{\kappa}|\mu\mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{A}^{(\alpha)}(\boldsymbol{\gamma}^{(\alpha)})^{(\theta_h^{(\alpha)}=1)} + \sum_{\alpha'\neq\alpha}\boldsymbol{A}^{(\alpha')}(\boldsymbol{\gamma}^{(\alpha')}), \boldsymbol{\Omega})}{S}$,

  $p_{h,2}^{(\alpha)} = \dfrac{\pi_{h,2}^{(\alpha)} N(\boldsymbol{\kappa}|\mu\mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{A}^{(\alpha)}(\boldsymbol{\gamma}^{(\alpha)})^{(\theta_h^{(\alpha)}=0)} + \sum_{\alpha'\neq\alpha}\boldsymbol{A}^{(\alpha')}(\boldsymbol{\gamma}^{(\alpha')}), \boldsymbol{\Omega})}{S}$,

  $p_{h,3}^{(\alpha)} = \dfrac{\pi_{h,3}^{(\alpha)} N(\boldsymbol{\kappa}|\mu\mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{A}^{(\alpha)}(\boldsymbol{\gamma}^{(\alpha)})^{(\theta_h^{(\alpha)}=-1)} + \sum_{\alpha'\neq\alpha}\boldsymbol{A}^{(\alpha')}(\boldsymbol{\gamma}^{(\alpha')}), \boldsymbol{\Omega})}{S}$, where,

  $S = \sum_{s\in\{0,1,-1\}}\pi_{h,s}^{(\alpha)} N(\boldsymbol{\kappa}|\mu\mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta}_x + \boldsymbol{A}^{(\alpha)}(\boldsymbol{\gamma}^{(\alpha)})^{(\theta_h^{(\alpha)}=s)} + \sum_{\alpha'\neq\alpha}\boldsymbol{A}^{(\alpha')}\boldsymbol{\gamma}^{(\alpha')}, \boldsymbol{\Omega})$

- $\omega_i| - PG(1, \mu + \boldsymbol{x}_i^T\boldsymbol{\beta}_x + \sum_{\alpha=1}^{L}\boldsymbol{w}_i^{(\alpha)T}\boldsymbol{\gamma}^{(\alpha)})$, for $i = 1, ..., n$.