

## BAYESIAN MODELING OF THE STRUCTURAL CONNECTOME FOR STUDYING ALZHEIMER'S DISEASE

BY ARKAPRAVA ROY<sup>\*,1</sup>, SUBHASHIS GHOSAL<sup>†,1</sup>, JEFFREY PRESCOTT<sup>‡</sup>,  
KINGSHUK ROY CHOUDHURY<sup>\*</sup> AND FOR THE ALZHEIMER'S DISEASE  
NEUROIMAGING INITIATIVE<sup>2</sup>

*Duke University\**, *North Carolina State University†* and  
*MetroHealth Medical Center‡*

We study possible relations between Alzheimer's disease progression and the structure of the connectome which is white matter connecting different regions of the brain. Regression models in covariates including age, gender and disease status for the extent of white matter connecting each pair of regions of the brain are proposed. Subject inhomogeneity is also incorporated in the model through random effects with an unknown distribution. As there is a large number of pairs of regions, we also adopt a dimension reduction technique through graphon (*J. Combin. Theory Ser. B* **96** (2006) 933–957) functions which reduces the functions of pairs of regions to functions of regions. The connecting graphon functions are considered unknown but the assumed smoothness allows putting priors of low complexity on these functions. We pursue a nonparametric Bayesian approach by assigning a Dirichlet process scale mixture of zero to mean normal prior on the distributions of the random effects and finite random series of tensor products of B-splines priors on the underlying graphon functions. We develop efficient Markov chain Monte Carlo techniques for drawing samples for the posterior distributions using Hamiltonian Monte Carlo (HMC). The proposed Bayesian method overwhelmingly outperforms a competing method based on ANCOVA models in the simulation setup. The proposed Bayesian approach is applied on a dataset of 100 subjects and 83 brain regions and key regions implicated in the changing connectome are identified.

**1. Introduction.** Alzheimer's disease (AD) is a neurodegenerative disorder that affects approximately five million people in the US and 30 million globally. Current thought is that detecting pathologic changes in the brain before the development of clinical symptoms will allow for a successful treatment. In particular

---

Received September 2018; revised March 2019.

<sup>1</sup>Supported in part by NSF Grant DMS-1510238.

<sup>2</sup>Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

*Key words and phrases.* ADNI, B-spline prior, brain image, connectome, graphical model, Graphon, HMC.

one area of active investigation is how changes in white matter, which constitutes the brain's structural network or "connectome," occur alongside Alzheimer's disease progression (Phillips et al. (2015)). Recent research has found that changes in structural connectivity become more widespread with Alzheimer's disease progression (Tucholka et al. (2018)) and that changes in the connectome appear to preferentially affect less connected areas of the brain (Daianu et al. (2015)). Further, changes in white matter are correlated with amyloid plaque burden which is one of the pathological hallmarks of the AD and has been shown to become elevated years before the onset of clinical symptoms (Prescott et al. (2014)). From a methodological viewpoint all these papers rely on modelling specific scalar metrics, such as connectivity indices which are derived from the connectivity graphs obtained via image processing brain scans (described later in methods). One drawback of this approach is that the conclusions can depend on the choice of metric (Phillips et al. (2015)). In this paper we overcome this dilemma by modelling changes to the entire graph of the network.

Our study is performed using data obtained by Alzheimer's Disease Neuroimaging Initiative (ADNI—[adni.loni.usc.edu](http://adni.loni.usc.edu)). Using T1-weighted magnetic resonance (MR) images from ADNI, the cortex of the brain is divided into several regions using a standard anatomic atlas. These regions are connected by white matter fibers and identified by using diffusion tensor imaging (DTI) MR. The data used in this study were from the ADNI 2 study and included all subjects with diffusion-weighted imaging as of November 2012. Readers can be referred to the "Materials and Methods" section of Prescott et al. (2014) for details of MR image acquisition and processing. The graphical representation of these white matter connections between cortical regions is referred to as the connectome. It is thought that some of these connections between brain regions become weakened over time due to the AD. Some other factors like age or sex might also affect the connectome, as well as subject-specific random effects. We model the connectome using graph-theoretic metrics, accounting for patients' specific effects and implement a Bayesian analysis.

Here, we consider a connectome with 83 cortical regions in the brain. Mathematically, the connectome can be viewed as a graph  $(V, E)$ , where  $V$  denotes the set of nodes or vertices standing for brain regions and  $E$  for the edges between pairs of vertices whenever present. As in a graphical model, edges are marked with certain measurements. In our context the measurements consist of observing the presence or absence of white matter fibers connecting two regions, the number of white matter fiber and the mean length of white matter fibers between them. Our aim is to identify the significant pairs of regions corresponding to different covariates in the connectome. The aspects of the presence of a connection, the number of connections and the mean length are modeled respectively by a binary regression model with a probit link, a Poisson regression model with an exponential link and a normal regression model. Interactions of covariates with pairs of regions are considered. Because the number of region-pairs is prohibitively high,

it leads to a very high dimensional regression model if a naive approach is taken. Hence, we use a dimension reduction technique that introduces a latent variable for each region and expresses functions of region-pairs in terms of a single smooth unknown function of each pair of latent variables as in a graphon model (Lovász and Szegedy (2006)). If  $(a_{jk} : j, k = 1, \dots, J)$  is an array of parameters, then we model  $a_{jk} = g(u_j, u_k)$ , where  $u_1, \dots, u_J$  are latent variables and  $g$  is a function of two arguments. Symmetry of the matrix  $((a_{jk}))$  is respected if  $g$  is symmetric in its arguments. The original motivation for this representation is that if  $a_{jk}$ ,  $j, k = 1, \dots, J$  are random variables such that the matrix  $((a_{jk}))$  is distributionally invariant under permutations of rows and columns, then  $a_{jk} = g(u_j, u_k)$  for some latent variables  $u_1, \dots, u_J$  that can be assumed to be uniformly distributed and for a function  $g$ , to be called a graphon function. In the present context if the parameters  $a_{jk}$ ,  $j, k = 1, \dots, J$  are treated as random, then such row and column wise exchangeability conditions are natural noninformativeness conditions. Given the graphon function, thus the strength of connections is determined by only 83 latent variables linked with each region instead of being  $\binom{83}{2} = 3486$  making huge computational savings. The graphon function is also treated as an unknown without any specific parametric form and is nonparametrically estimated from the data. More specifically, we put a finite random series prior based on tensor products of B-splines, where the coefficients are given appropriate prior distributions. Finite random series priors are widely used in the literature to construct priors on various functions and are systematically studied, for instance in Shen and Ghosal (2015), but it seems to have not been used before for putting prior distributions on graphon functions. The assumed smoothness of the graphon function helps keep the number of basis function required for the basis expansion relatively small. This is because to approximate a function in  $[0, 1]^d$  of smoothness index  $\beta$  by a finite series within accuracy  $\epsilon$ , one needs to use only  $O(\epsilon^{-d/\beta})$  many elements of standard bases like polynomials, B-splines or wavelets, that is, fewer functions are needed for smoother functions. The proposed Bayesian procedure can be shown to lead to consistent posterior in the setting where the number of regions remains fixed, but the number of subjects increases to infinity. The result goes beyond the particular application we are addressing in the present paper. Although the proof uses the standard hypothesis testing and prior positivity of Kullback–Leibler neighborhoods approach developed by Schwartz, our major technical contribution is the construction of exponentially consistent tests for random effects in a Poisson regression model, along with weakening conditions on the predictor variables to include a larger variety of applications. Details on posterior consistency are shown in the Supplementary Material (Roy et al. (2019)).

The rest of the paper is organized as follows. In the next section we describe the details of modeling the connectome dataset from ADNI. In Section 3 we describe the prior construction and develop posterior computing techniques. A simulation study comparing the proposed Bayesian procedure with ANCOVA-based ones is conducted in Section 4. The real-data on connectome from ADNI is analyzed in Section 5. Finally, Section 6 ends with some concluding remarks.

**2. Data description and modeling.** Data used in the preparation of this article were obtained from the ADNI database. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Here, we examine if the connectome can be used as a biomarker. In the ADNI dataset on connectome, the brain is divided into  $J = 83$  regions. For each pair of brain regions, the number of white matter fibers between them and their average lengths is obtained. The data is obtained for  $n = 100$  subjects, for whom information regarding disease prognosis, sex and age are also available. For many edges the mean lengths are not defined where there are no white matter fibers. There are three disease prognosis states—Alzheimer, mild cognitive impairment (MCI) and no cognitive impairment (NC).

Except for age, the other two covariates, sex and disease prognosis, are categorical. Since the disease prognosis has three possible states, we introduce dummy variables  $Z^{\text{MCI}}$  and  $Z^{\text{AD}}$  respectively standing for the onset of MCI and AD, setting NC at the baseline. Similarly, the dummy variable  $Z^{\text{M}}$  indicating male gender is introduced setting females at the baseline. Let  $Z = (Z^{\text{MCI}}, Z^{\text{AD}}, Z^{\text{M}}, \text{Age})'$  stand for the whole vector of covariates and  $Z_i$  stand for its value for the  $i$ th subject. Let  $N_{ijk}$  stand for the number of white matter fiber connecting brain regions  $j$  and  $k$  in the  $i$ th subject, and  $L_{ijk}$  the mean length of such fibers, provided that  $N_{ijk} \geq 1$ . In ADNI dataset there were no missing values in  $N_{ijk}$  or  $L_{ijk}$ . For three individuals gender and age were missing. They were imputed using R package MICE (van Buuren and Groothuis-Oudshoorn (2010)). Under the assumption that missing data are Missing At Random (MAR), the dataset is imputed based on conditional distributions of each variable. MAR means that the probability that a value is missing depends only on the observed values. In this technique multiple imputed datasets are generated in parallel to reduce uncertainty in imputation.

It seems natural to consider a Poisson model for the counts of fibers connecting two regions in a subject. However, as shown in Figure 1, the proportions of connected edges among individuals are between 10% to 30%. Thus, the abundance of zero connections makes the Poisson model somewhat inappropriate. We overcome the problem by considering a zero-inflated Poisson model, by boosting the probability of zero through a binary latent variable  $\Xi_{ijk}$  with parameter  $\Phi(\pi_{ijk})$ , where  $\Phi$  stands for the standard normal distribution function and  $\pi_{ijk}$  is a real-valued parameter. If  $\Xi_{ijk} = 0$ , then  $N_{ijk}$  is set at zero, while if  $\Xi_{ijk} = 1$ , the number of connections  $N_{ijk}$  is assumed to be Poisson distributed with some positive mean  $e^{\lambda_{ijk}}$ . Note that in our formulation  $\Xi_{ijk}$  is not completely identifiable since the value  $N_{ijk} = 0$  is compatible with both possible values of  $\Xi_{ijk}$ . If  $N_{ijk} \geq 1$ , we assume that the mean fiber length, in the logarithmic scale, is normally distributed

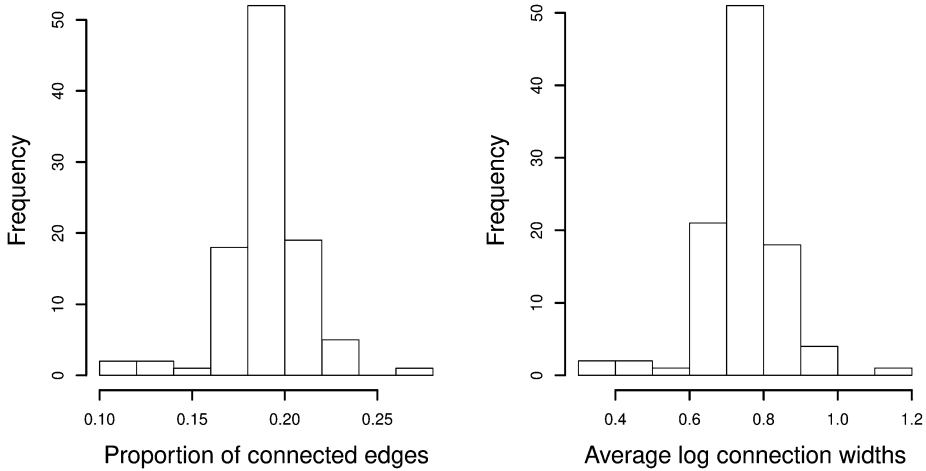


FIG. 1. Histogram of proportion of connected edges and average log connection widths across individual. *dip.test* of R package *diptest* (Maechler (2013)) suggests that these are unimodal with kurtosis ( $\mu_4/\mu_2^2 - 3$ ) of 3.74 and 3.79 respectively, in comparison to kurtosis of normal, being zero.

with some mean  $\mu_{ijk}$ , and variance  $\sigma^2/N_{ijk}$  for some unknown  $\sigma > 0$ . The heuristic justification of the choice of the variance  $\sigma^2/N_{ijk}$  stems from the fact that  $L_{ijk}$  is an average of  $N_{ijk}$  (independent) variables and should be approximately normal with variance inversely proportional to the number  $N_{ijk}$  of averaging variables. Since fiber lengths are positive, the model seems to fit the data better in the logarithmic scale, and the heuristics for the choice of the variance extends to the logarithmic scale by the delta method, at least when  $N_{ijk}$  is large. Thus we can represent the data generating process as

$$\begin{aligned}
 \Xi_{ijk} &\sim \text{Bin}(1, \Phi(\pi_{ijk})), \\
 N_{ijk}|\{\Xi_{ijk} = 1\} &\sim \text{Poisson}(e^{\lambda_{ijk}}), \\
 \log L_{ijk} &= \mu_{ijk} + \epsilon_{ijk}, \\
 \epsilon_{ijk}|\{N_{ijk} \geq 1\} &\sim N(0, \sigma^2/N_{ijk}).
 \end{aligned}
 \tag{1}$$

A simple analysis of covariance (ANCOVA)-type model can be formulated to describe linear effects of the covariates  $Z_i$  on each unrestricted parameter  $\pi_{ijk}$ ,  $\lambda_{ijk}$  and  $\mu_{ijk}$  for each pair of brain regions  $(j, k)$ :

$$\begin{aligned}
 \mu_{ijk} &= ((\mu_0))_{j,k} + Z'_i \chi_{jk} + \eta_{1i}, \\
 \pi_{ijk} &= ((\pi_0))_{j,k} + Z'_i \beta_{jk} + \eta_{2i}, \\
 \lambda_{ijk} &= ((\lambda_0))_{j,k} + Z'_i v_{jk} + \eta_{3i},
 \end{aligned}
 \tag{2}$$

where  $((\mu_0))_{j,k}$ ,  $((\pi_0))_{j,k}$  and  $((\lambda_0))_{j,k}$  are baseline values of  $\mu_{ijk}$ ,  $\pi_{ijk}$  and  $\lambda_{ijk}$  respectively, for covariate value  $Z_i = 0$ . Let

$$(3) \quad \begin{aligned} \chi_{jk} &= (((\chi_{MCI}))_{j,k}, ((\chi_{AD}))_{j,k}, ((\chi_M))_{j,k}, ((\chi_{Age}))_{j,k})', \\ \beta_{jk} &= (((\beta_{MCI}))_{j,k}, ((\beta_{AD}))_{j,k}, ((\beta_M))_{j,k}, ((\beta_{Age}))_{j,k})', \\ \nu_{jk} &= (((\nu_{MCI}))_{j,k}, ((\nu_{AD}))_{j,k}, ((\nu_M))_{j,k}, ((\nu_{Age}))_{j,k})' \end{aligned}$$

be regression coefficients for the average length, connection probability and number of connections respectively, and  $((\cdot))_{j,k}$  denote  $(j, k)$ th element of a matrix, and  $\eta_i = (\eta_{1i}, \eta_{2i}, \eta_{3i})'$ ,  $i = 1, \dots, n$ , be independent random effects of the  $i$ th subject distributed according to an unknown common distribution. It may be noted that the normal distribution function  $\Phi$  and the exponential function are used respectively as link functions for binary and Poisson regression. For the latter, the exponential link is almost a universal choice, while for binary regression both logistic and probit (i.e.,  $\Phi$ ) links are commonly used and usually give similar results. Our preference for the probit link is due to its computational advantage in a Gibbs sampling scheme for Bayesian computation, through a data-augmentation technique (see [Albert and Chib \(1993\)](#)).

For a preliminary analysis we fit the model using a generalized heteroscedastic ANCOVA, ignoring the zero-inflation aspect and the random effects in the model. The model thus has  $3 \times \binom{83}{2} = 10,458$  parameters and 34,860 observations of mean length and number of white matter fibers corresponding to 100 subjects and 3486 potential edges between different brain regions. We observed that for several edges  $(j, k)$ , the maximum likelihood method failed to give estimates of either  $\mu_{0,jk}$  or  $\chi_{jk}$ . For the Poisson regression the `glm` function in R did not converge for several pairs  $(j, k)$ . This is due to an insufficient number of observations. Thus, it suggests using a dimension reduction of the parameter space through further modeling if we want to conduct an edge-wise analysis. The dimension reduction also helps with computation and gives easy interpretability of the results.

Since the parameters are indexed by edges, a substantial reduction of dimension will be possible if these can be viewed as arising from some latent characteristics of nodes through some fixed but unknown function. This can be motivated from exchangeability considerations. In the absence of initial information about connections between regions, exchangeability seems to be an appealing assumption. By a well known representation theorem of exchangeable random graphs (cf. [Aldous \(1981\)](#), [Hoover \(1979\)](#)), a function of edge  $(j, k)$  can then be represented as  $f(\xi_i, \xi_j)$ , where  $\xi_i$ , for each node  $i$  is a latent variable independently and identically distributed and  $f$  is a fixed function, called a graphon, irrespective of the size of the network. Assuming that the function  $f$  is sufficiently smooth, a basis expansion can approximate it using only a few terms. Thus the graphon technique in our context will be able to reduce a parameter array of size  $\binom{83}{2} = 3486$  to only a parameter vector of size  $83 + K$ , where  $K$  is the number of parameters used to

approximate the unknown smooth graphon function. Typically, a modest number of terms suffices for well-behaved functions using standard bases such as B-splines or polynomials. As a result a substantial dimension reduction is possible through the graphon technique. This leads to modeling the arrays of baseline values and regression coefficient as

$$\begin{aligned}
 ((\mu_0))_{j,k} &= \mu(\xi_j, \xi_k), & ((\pi_0))_{j,k} &= \pi(\xi_j, \xi_k), & ((\lambda_0))_{j,k} &= \lambda(\xi_j, \xi_k), \\
 ((\chi_l))_{j,k} &= \chi_l(\delta_j, \delta_k), & l &= \text{MCI, AD, M, Age}, \\
 ((\beta_l))_{j,k} &= \beta_l(\delta_j, \delta_k), & l &= \text{MCI, AD, M, Age}, \\
 ((\nu_l))_{j,k} &= \nu_l(\delta_j, \delta_k), & l &= \text{MCI, AD, M, Age},
 \end{aligned}
 \tag{4}$$

where, with an abuse of notations,  $\mu, \pi, \lambda, \chi_{\text{MCI}}, \chi_{\text{AD}}, \chi_{\text{M}}, \chi_{\text{Age}}, \beta_{\text{MCI}}, \beta_{\text{AD}}, \beta_{\text{M}}, \beta_{\text{Age}}, \lambda_{\text{MCI}}, \lambda_{\text{AD}}, \lambda_{\text{M}}$  and  $\lambda_{\text{Age}}$  are smooth functions on the unit square  $[0, 1]^2$  and symmetric in their arguments, and  $\xi_1, \dots, \xi_J$  and  $\delta_1, \dots, \delta_J$  are latent variables taking values in the unit interval. The reason for choosing two separate sets of latent variables  $\xi$  and  $\delta$  is to distinguish between fixed and main effects.

### 3. Prior specification and posterior computation.

3.1. *Prior specification.* To proceed with a nonparametric Bayesian analysis, we put prior distributions on the smooth functions appearing in the graphon representation through basis expansion in tensor products of B-splines and on the coefficients of the basis expansion. The coefficients can be arranged in the form of a square matrix. The symmetry of the matrices of coefficients ensures symmetry of the resulting functions in its arguments as required by graphon functions. Given other sets of parameters and values of the random effects, (independent) normal priors on the coefficients of the tensor products of B-splines will lead to conjugacy in the normal regression model for the length, allowing a simple and fast posterior updating rule. In the binary regression model for the connection probability, normal prior still leads to conjugacy using the data augmentation technique of Albert and Chib (1993). Since no conjugacy is possible for the Poisson regression for the number of connections, gradient-based Hamiltonian Monte Carlo (HMC) sampling algorithm is applied. Alternatively, adaptive rejection sampling can be applied to obtain posterior updates. On the distribution  $G$  of the random effects, we put a Dirichlet process scale mixture of zero mean normal prior (see Chapter 5 of Ghosal and van der Vaart (2017) and West (1987)) for scale mixture of normal). The histogram plot of Figure 1 as well as Figure 1 from the Supplementary Material motivate us that the distributions of random effects are symmetric but nonnormal with higher kurtosis.

More specifically, the prior can be completely described by the following set of relations:

(i) Graphon functions:

$$\mu(\xi_j, \xi_k) = \sum_{m=1}^K \sum_{m'=1}^K \theta_{1,mm'} B_m(\xi_j) B_{m'}(\xi_k),$$

$$\pi(\xi_j, \xi_k) = \sum_{m=1}^K \sum_{m'=1}^K \theta_{2,mm'} B_m(\xi_j) B_{m'}(\xi_k),$$

$$\lambda(\xi_j, \xi_k) = \sum_{m=1}^K \sum_{m'=1}^K \theta_{3,mm'} B_m(\xi_j) B_{m'}(\xi_k),$$

and for  $l = \text{MCI, AD, M}$ ,

$$\chi_l(\delta_j, \delta_k) = \sum_{m=1}^K \sum_{m'=1}^K \gamma_{1l,mm'} B_m(\delta_j) B_{m'}(\delta_k),$$

$$\beta_l(\delta_j, \delta_k) = \sum_{m=1}^K \sum_{m'=1}^K \gamma_{2l,mm'} B_m(\delta_j) B_{m'}(\delta_k),$$

$$\nu_l(\delta_j, \delta_k) = \sum_{m=1}^K \sum_{m'=1}^K \gamma_{3l,mm'} B_m(\delta_j) B_{m'}(\delta_k),$$

where  $\theta_{t,mm'} = \theta_{t,m'm}$  for all  $t = 1, 2, 3$ , and  $\gamma_{tl,mm'} = \gamma_{t,m'm}$  for all  $t = 1, 2, 3, l = \text{MCI, AD, M}$ , and that

(a) graphon coefficients: For some chosen  $a > 0$ ,

$$\theta_{t,mm'} \stackrel{\text{ind}}{\sim} N(0, a^2), \quad \gamma_{tl,mm'} \stackrel{\text{ind}}{\sim} N(0, a^2), \quad 1 \leq m \leq m' \leq K,$$

for  $t = 1, 2, 3, l = \text{MCI, AD, M}$ ;

(b) latent variables:

$$\text{logit}(\xi_1), \dots, \text{logit}(\xi_J) \stackrel{\text{ind}}{\sim} N(0, a^2), \quad \text{logit}(\delta_1), \dots, \text{logit}(\delta_J) \stackrel{\text{ind}}{\sim} N(0, a^2),$$

here,  $\text{logit}$  stands for the  $\text{logit}(x) = \log(x/(1 - x))$ .

(ii) Random effects distribution: For  $t = 1, 2, 3$  and  $i = 1, \dots, n$ ,

$$\eta_{ti} | \tau_{ti} \stackrel{\text{ind}}{\sim} N(0, \tau_{ti}^2), \quad \tau_{ti}^2 \stackrel{\text{ind}}{\sim} G_t, \quad t = 1, 2, 3, \quad G_t \stackrel{\text{ind}}{\sim} \text{DP}(\alpha_t \text{IG}(b_1, b_2)),$$

where  $\text{DP}$  stands for the Dirichlet process,  $\text{IG}$  for the inverse-gamma distribution and the precision parameter  $\alpha_t$  of the Dirichlet process is given a gamma prior  $\alpha_t \sim \text{Ga}(c_1, c_2)$ .

(iii) Error variance:  $\sigma^{-2} \sim \text{Ga}(d_1, d_2)$ .



3.2. *Posterior updating.* Introduce a latent variable  $I_j$  the indicator of the  $\text{Un}(0, 1)$  component of the distribution of  $\delta_j$ ,  $j = 1, \dots, J$ . Now, the conditional log-likelihood is given by

$$\begin{aligned}
 C - \sum_{i,j,k} \exp \left\{ \sum_{m,m'} \left[ \theta_{3,mm'} B_m(\xi_j) B_{m'}(\xi_k) \right. \right. \\
 + \sum_l \gamma_{3l,mm'} B_m(\delta_j) B_{m'}(\delta_k) Z_{il} \left. \right] + \eta_{3i} \left. \right\} \\
 + \sum_{i,j,k} N_{ijk} \left[ \sum_{m,m'} \left( \theta_{3,mm'} B_m(\xi_j) B_{m'}(\xi_k) \right. \right. \\
 + \sum_l \gamma_{3l,mm'} B_m(\delta_j) B_{m'}(\delta_k) Z_{il} \left. \right) + \eta_{3i} \left. \right] \\
 - \frac{1}{2\sigma^2} \sum_{i,j,k} N_{ijk} \left[ \log L_{ijk} - \sum_{m=1}^K \sum_{m'=1}^K \left( \theta_{1,mm'} B_m(\xi_j) B_{m'}(\xi_k) \right. \right. \\
 + \sum_l \gamma_{1l,mm'} B_m(\delta_j) B_{m'}(\delta_k) Z_{il} \left. \right) + \eta_{1i} \left. \right]^2 \\
 + \sum_{i,j,k} I(N_{ijk} = 0) \log \Phi \left( \sum_{m,m'} \left( \theta_{2,mm'} B_m(\xi_j) B_{m'}(\xi_k) \right. \right. \\
 + \sum_l \gamma_{2l,mm'} B_m(\delta_j) B_{m'}(\delta_k) Z_{il} \left. \right) + \eta_{2i} \left. \right) \\
 + \sum_{i,j,k} (1 - I(N_{ijk} = 0)) \log \left( 1 - \Phi \left( \sum_{m,m'} \left( \theta_{2,mm'} B_m(\xi_j) B_{m'}(\xi_k) \right. \right. \right. \\
 + \sum_l \gamma_{2l,mm'} B_m(\delta_j) B_{m'}(\delta_k) Z_{il} \left. \right) + \eta_{2i} \left. \right) \left. \right) \\
 - \frac{1}{2a^2} \sum_{m \leq m'} (\theta_{1,mm'}^2 + \theta_{2,mm'}^2 + \theta_{3,mm'}^2) \\
 - \frac{1}{2a^2} \sum_{m \leq m'} \sum_l (\gamma_{1,mm'l}^2 + \gamma_{2,mm'l}^2 + \gamma_{3,mm'l}^2) \\
 + \log \left( (1 - I_j) \delta_j^{M-1} (1 - \delta_j)^{M-1} \frac{\Gamma(M)^2}{\Gamma(2M)} + I_j \right) \\
 + I_j \log q + (1 - I_j) \log(1 - q) \\
 - (nJ^2/2 + d_1 - 1) \log \sigma^2 - d_2/\sigma^2,
 \end{aligned}
 \tag{5}$$

where  $C$  involves only hyperparameters  $a, M, K, b_1, b_2, c_1, c_2, d_1, d_2, q$  and the observations, but not the parameters, of the model.

The parameters having conjugacy are updated using Gibbs sampling scheme and most of the other parameters are updated using gradient based HMC. Detailed calculations of all the posterior updates are shown in the Supplementary Material.

**3.3. Tuning.** The leapfrog parameter is kept fixed at 10. The step length is tuned after each 100 iterations to ensure an acceptance rate between 55% to 90%. The number of B-spline basis functions ( $K$ ) is tuned via a grid search over a sequence of values in the range 7–20. For each possible values of  $K$ , we generate 10 sets of latent variables. For each set of latent variables, we can fit a simple linear regression to estimates the B-spine coefficients and calculate average the Akaike Information Criterion (AIC) over all the sets of latent variables. Based on these AIC values, we pick the  $K$  with the lowest AIC value or the smallest value after which there is not much improvement in the AIC.

**4. Simulation.** In this section we study the performance of the proposed Bayesian method in comparison with ANCOVA. For computational simplicity we do not consider the random effects in the data generating process as well as in the model, that is, we consider the following analog of (1):

$$\begin{aligned}
 \Xi_{ijk} &\sim \text{Bin}(1, \Phi(\pi_{ijk})), \\
 N_{ijk}|\{\Xi_{ijk} = 1\} &\sim \text{Poisson}(e^{\lambda_{ijk}}), \\
 \log L_{ijk} &= \mu_{ijk} + \epsilon_{ijk}, \\
 \epsilon_{ijk}|\{N_{ijk} \geq 1\} &\sim \text{N}(0, \sigma^2/N_{ijk}), \\
 \pi_{ijk} &= ((\pi_0))_{j,k} + Z_i' \beta_{jk}, \\
 \mu_{ijk} &= \mu_{0,jk} + Z_i' \chi_{jk}, \\
 \lambda_{ijk} &= \lambda_{0,jk} + Z_i' \nu_{jk}.
 \end{aligned}
 \tag{6}$$

We consider  $n = 50, 100, 200, 500, 1000, 2000$  subjects for data generation with  $J = 20$  and  $n = 100, 250, 500, 1000$  for  $J = 40$  nodes. There is another simulation setup with sample size 100 and  $J = 80$  which is similar to our real-data application.

*Data generation:*

The true matrices are generated as follows with  $\epsilon_{l,jk} = (e_{l,jk} + e_{l,kj})/2$  and  $e_{jk} \sim \text{N}(0, (\sqrt{0.05})^2)$  for  $l = 1, \dots, 15$ ,

$$\mu_0(j, k) = ((-0.5\xi_j - 0.4\xi_k)^3 + (-0.4\xi_j - 0.5\xi_k)^3)/2 + \epsilon_{1,jk},$$

$$\pi_0(j, k) = ((-0.7\xi_j - \xi_k)^3 + (-\xi_j - 0.7\xi_k)^3)/2 + \epsilon_{2,jk},$$

$$\lambda_0(j, k) = ((-0.5\xi_j - 0.4\xi_k)^3 + (-0.4\xi_j - 0.5\xi_k)^3)/2 + \epsilon_{3,jk},$$

$$\begin{aligned}
\chi_{\text{MCI}}(j, k) &= (\exp(-0.5\delta_j - 0.4\delta_k) + \exp(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{4,jk}, \\
\beta_{\text{MCI}}(j, k) &= (\exp(-0.7\delta_j - \delta_k) + \exp(-\delta_j - 0.7\delta_k))/2 + \epsilon_{5,jk}, \\
\nu_{\text{MCI}}(j, k) &= (\exp(-0.5\delta_j - 0.4\delta_k) + \exp(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{6,jk}, \\
\chi_{\text{AD}}(j, k) &= (\sin(-0.5\delta_j - 0.4\delta_k) + \sin(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{7,jk}, \\
\beta_{\text{AD}}(j, k) &= (\sin(-0.7\delta_j - \delta_k) + \sin(-\delta_j - 0.7\delta_k))/2 + \epsilon_{8,jk}, \\
\nu_{\text{AD}}(j, k) &= (\sin(-0.5\delta_j - 0.4\delta_k) + \sin(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{9,jk}, \\
\chi_{\text{M}}(j, k) &= (\cos(-0.5\delta_j - 0.4\delta_k) + \cos(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{10,jk}, \\
\beta_{\text{M}}(j, k) &= (\cos(-0.7\delta_j - \delta_k) + \cos(-\delta_j - 0.7\delta_k))/2 + \epsilon_{11,jk}, \\
\nu_{\text{M}}(j, k) &= (\cos(-0.5\delta_j - 0.4\delta_k) + \cos(-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{12,jk}, \\
\chi_{\text{Age}}(j, k) &= ((-0.5\delta_j - 0.4\delta_k) + (-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{13,jk}, \\
\beta_{\text{Age}}(j, k) &= ((-0.7\delta_j - \delta_k) + (-\delta_j - 0.7\delta_k))/2 + \epsilon_{14,jk}, \\
\nu_{\text{Age}}(j, k) &= ((-0.5\delta_j - 0.4\delta_k) + (-0.4\delta_j - 0.5\delta_k))/2 + \epsilon_{15,jk}.
\end{aligned}$$

If  $\Xi_{ijk}$  is generated as zero, the edge  $(j, k)$  of  $i$ th individual will be missing. The generated data based on these functions have similar missingness compared to the real-data. We add the error component along with the functional values to deviate it a little bit from an exact functional form. We have performed 50 replications for each case. We collect 5000 MCMC samples after burning in 5000 initial samples to draw the inference. The result for  $J = 40$  are based on 30 replications and 3000 post-burn samples after burning in 3000 initial samples.

*Choice of the hyperparameters:* We choose the hyperparameters  $a = 10$ ,  $M = 10$ ,  $b_1 = b_2 = 0.1$  and  $c_1 = c_2 = 10$  in (5). We take seven B-spline basis functions based on the AIC values over a grid of possible number of B-spline basis functions. For all of the simulated datasets, it always produces the optimal number as seven, as the number of nodes is fixed at 20.

For the ANCOVA based estimation we use the weighted least squares technique for the normal model and the generalized linear regression for a Poisson regression model with the exponential link function.

We present a comparative plot of squared bias, variance and MSE of the estimates across different sample sizes in Figure 3 for small sample sizes and Figure 4 for large sample sizes for  $J = 20$ . For  $J = 40$ , we present similar plots for sample sizes 100, 250, 500 and 1000 in Figure 5. We have not included the Bayesian ANCOVA estimates in any of these plots as estimation MSEs since the frequentist ANCOVA and the Bayesian ANCOVA estimates are very similar. For the largest sample size 2000 with  $J = 20$ , Table 1 contains the bias squares and variance values for both of these two methods. These comparisons are also made for the sample size 100 with  $J = 80$  in Table 2. The bias squares and variances of the estimated matrices are calculated after averaging over bias squares and variances

TABLE 1  
*Comparison of estimation accuracy of order  $10^{-2}$  for  $n = 2000$  with  $J = 20$*

	Bayes Graphon			ANCOVA			Bayes ANCOVA		
	Square bias	Variance	MSE	Square bias	Variance	MSE	Square bias	Variance	MSE
$\chi_{MCI}$	1.363	0.018	1.381	45.5	4.1	49.6	52.181	5.905	58.087
$\chi_{AD}$	1.057	0.011	1.068	25	2.63	27.63	27.669	3.845	31.515
$\chi_M$	1.768	0.006	1.775	82.5	2.4	84.9	82.162	3.520	85.682
$\chi_{Age}$	1.298	0.004	1.302	24.3	0.6	24.9	25.806	0.932	26.739
$\beta_{MCI}$	1.041	0.019	1.059	25.5	1.2	26.7	40.037	5.321	45.358
$\beta_{AD}$	1.825	0.012	1.837	61.3	1	62.3	77.604	4.069	81.674
$\beta_M$	1.418	0.013	1.431	45.2	0.6	45.8	54.186	3.214	57.4
$\beta_{Age}$	2.375	0.004	2.378	84.3	0.2	84.5	87.355	0.898	88.254
$\nu_{MCI}$	7.427	0.023	7.449	43.7	0.7	44.4	45.838	0.530	46.368
$\nu_{AD}$	3.549	0.015	3.563	23.2	0.5	23.7	24.249	0.366	24.615
$\nu_M$	9.066	0.007	9.072	81.3	0.4	81.7	83.559	0.369	83.928
$\nu_{Age}$	3.126	0.005	3.132	23.9	0.1	24	24.082	0.106	24.188

of the individual entries. In the case of ANCOVA, there are several missing values in the estimates of these matrices. Thus, the bias square and variances cannot be calculated at those entries. Thus, these are calculated by averaging over only the available ones. In Tables 3, 4 and 5, we present the prediction MSE and the mean predictive likelihood for the logarithm of the connection width and the number of connections, respectively.

TABLE 2  
*Comparison of estimation accuracy of order  $10^{-2}$  for  $n = 200$  with  $J = 80$*

	Bayes Graphon			ANCOVA		
	Square bias	Variance	MSE	Square bias	Variance	MSE
$\chi_{MCI}$	2.58	0.04	2.63	74.16	94.93	169.09
$\chi_{AD}$	3.86	0.02	3.89	35.19	70.73	105.92
$\chi_M$	1.82	0.01	1.83	83.43	82.76	166.19
$\chi_{Age}$	4.16	0.01	4.18	30.10	19.07	49.17
$\beta_{MCI}$	2.63	$3.35 \times 10^{-02}$	2.63	16.66	0.92	17.58
$\beta_{AD}$	8.73	$1.87 \times 10^{-02}$	6.73	38.66	0.66	39.32
$\beta_M$	5.71	$1.98 \times 10^{-02}$	5.71	25.95	0.65	26.60
$\beta_{Age}$	18.05	$6.96 \times 10^{-03}$	18.05	61.84	0.12	61.96
$\nu_{MCI}$	14.26	$1.65 \times 10^{-06}$	14.26	104.99	82.21	187.19
$\nu_{AD}$	8.36	$1.09 \times 10^{-06}$	8.36	40.52	53.97	94.49
$\nu_M$	19.05	$4.48 \times 10^{-07}$	19.05	71.58	64.94	136.52
$\nu_{Age}$	8.86	$5.17 \times 10^{-07}$	8.86	37.90	15.94	53.84

TABLE 3

*Comparing prediction performances for different sample sizes across different models for  $J = 20$ . Prediction MSE is used for mean-width and Predictive mean log-likelihood is used for the number of connections. Half of the recorded sample size is used to estimate model parameters and the rest half is used to evaluate prediction performance*

	Bayes Graphon			ANCOVA			Bayes ANCOVA		
Total sample size	2000	1000	500	2000	1000	500	2000	1000	500
Mean-width	1.30	1.32	1.32	1.61	1.63	1.70	1.61	1.64	1.73
Number of connections	-168.61	-169.21	-210.26	-735.28	-835.21	-811.41	-181.43	-184.18	-288.50

TABLE 4

*Comparing prediction performances for different sample sizes across different models for  $J = 40$ . Prediction MSE is used for mean-width, and Predictive mean log-likelihood is used for the number of connections. Half of the recorded sample size is used to estimate model parameters and the rest half is used to evaluate prediction performance*

	Bayes Graphon			ANCOVA			Bayes ANCOVA		
Total sample size	2000	1000	500	2000	1000	500	2000	1000	500
Mean-width	1.33	1.33	1.33	1.64	1.69	1.86	1.63	1.72	1.90
Number of connections	-644.199	-645.43	-809.2	-260.897	-2737.39	-3262.96	-696.38	-711.95	-832.11

TABLE 5

Comparing prediction performances for total sample size 200 across different models for  $J = 80$ . Prediction MSE is used for mean-width, and Predictive mean log-likelihood is used for the number of connections. Half of the recorded sample size is used to estimate model parameters, and the other half is used to evaluate prediction performance

	Bayes Graphon	ANCOVA	Bayes ANCOVA
Mean-width	1.33	2.44	2.55
Number of connections	-2704.31	-36,791.27	-3802.34

In Figure 2 we see that the proposed Bayesian method identifies the true structure in most of the cases. For some cases it captures the structure, but the color levels are different. Usually, the differences are very small as can be observed in Tables 1 and 2. But for ANCOVA, even for a sample size as large as 2000, it could not capture the true structures. Figure 3 suggests that bias squares and variances of ANCOVA estimates are not decreasing as sample size increases for small sizes due to the varying missing entries for different sample sizes. As the sample size increases the number of missing values in the estimate goes down. Thus, more parameters can be estimated. More parameters become estimable as sample size increases, but this change is not huge. Thus for larger sample sizes, bias squares and variances of the ANCOVA estimates are decreasing. From Figure 4 we can also conclude that the proposed Bayesian method performs much better than ANCOVA for sample sizes as large as 500, 1000 or 2000. This is evident from Table 1 where bias squares values for the Bayesian estimates are around 40 times smaller, and variances are 300 times smaller for the Bayesian estimates than ANCOVA. It may be recalled that estimation by the Bayesian method is consistent. For  $J = 40$  in Figure 5, we can see that the Bayesian method overwhelmingly outperforms ANCOVA. In both Table 3 and Table 4, in terms of prediction accuracy, the proposed Bayesian method performs much better than other alternatives.

**5. Real-data analysis.** We analyze a real-dataset of 100 individuals, collected from ADNI. A demographic summary of the data is provided in Table 6. In disease categories we have fewer females than males. The baseline subject is a female subject of the average age of 73.9 with no cognitive impairment. In Figure 6 we show that the total number of connected edges and the total length of the fibers vary with gender, disease state and age. We can see that these distributions are different for male and female. Also, the individuals with Alzheimer's disease seem to have more short-range connections.

**5.1. Modeling.** As described in Section 3, we first generate 10 sets of latent variables for  $\xi$  and  $\delta$ . Then, we model the graphon functions with seven to 20 B-spline basis functions and calculate AIC values using standard R packages for

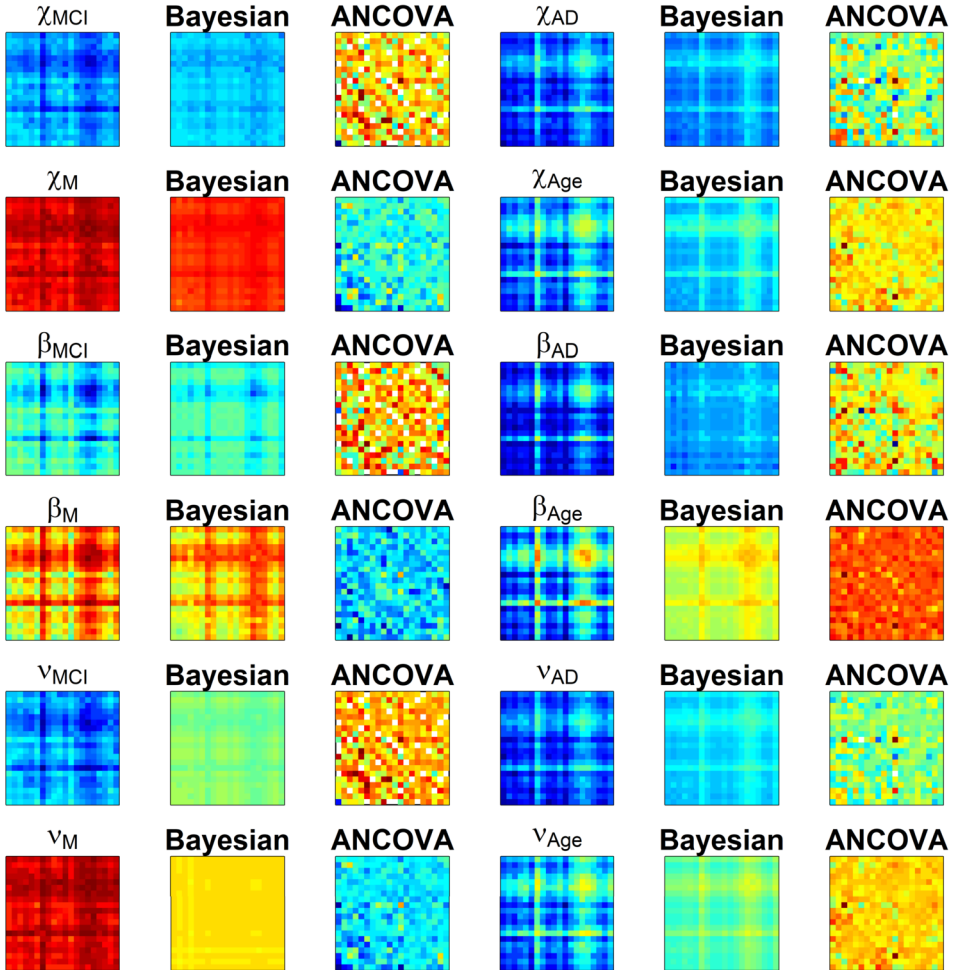


FIG. 2. Heatmap of the truth along with estimates from the Bayesian methods and the ANCOVA for sample size 2000 with 20 nodes. First and fourth columns are the true parameter, and subsequent columns present the corresponding Bayesian and ANCOVA estimates, averaged over 50 replications.

linear and generalized linear regressions, given each set of latent variables. After that we take the average of the AIC values for each case of a number of B-spline basis functions. After comparing these average AIC values, 13 B-spline basis functions are considered for the graphon functions in estimation for the real data. Other hyperparameters are kept the same as the simulation, that is,  $a = 10$ ,  $M = 10$ ,  $b_1 = b_2 = 0.1$  and  $c_1 = c_2 = 10$ . We collect in total 10,000 MCMC samples. Out of those, 5000 are postburn samples, collected after burning-in the first 5000 samples. We perform a test of significance for each edge by checking if zero is included in the 95% credible region, constructed from the postburn samples. After that we rank those by calculating the probability of greater than zero or less



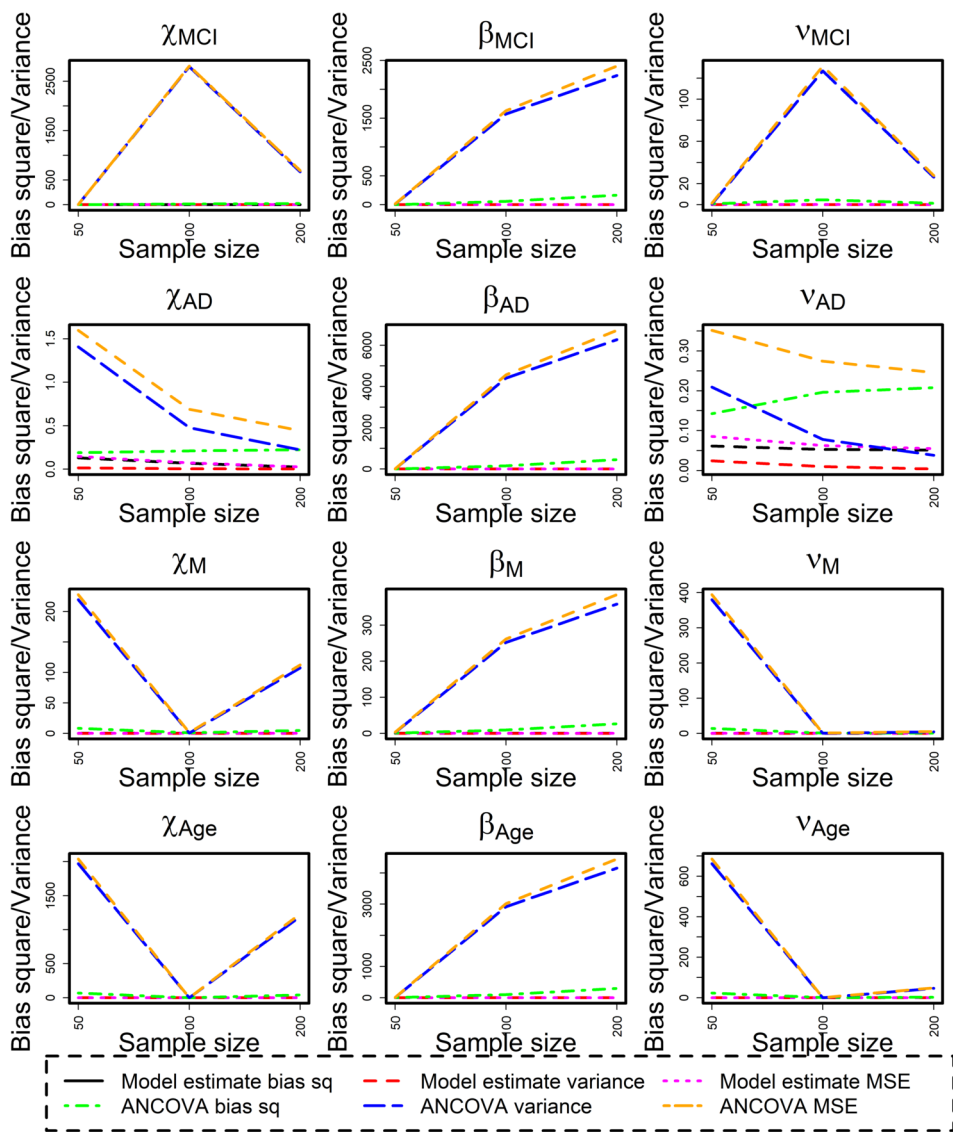


FIG. 3. Plot of squared bias and variance for different parameters against sample sizes for Bayesian method and ANCOVA estimates for small sample sizes 50, 100 and 200 with  $J = 20$ .

than zero depending on whether zero is in the left tail or in the right tail of the empirical distribution, constructed from the postburn samples. This can be used as an alternative for the p-value in the frequentist setup. If this comparison is inconclusive, that is, we get zero as the probability, we then compare the lengths of the credible sets. The shorter the length, the more significant that edge is. This is because a shorter length would suggest more concentration of the posterior distri-

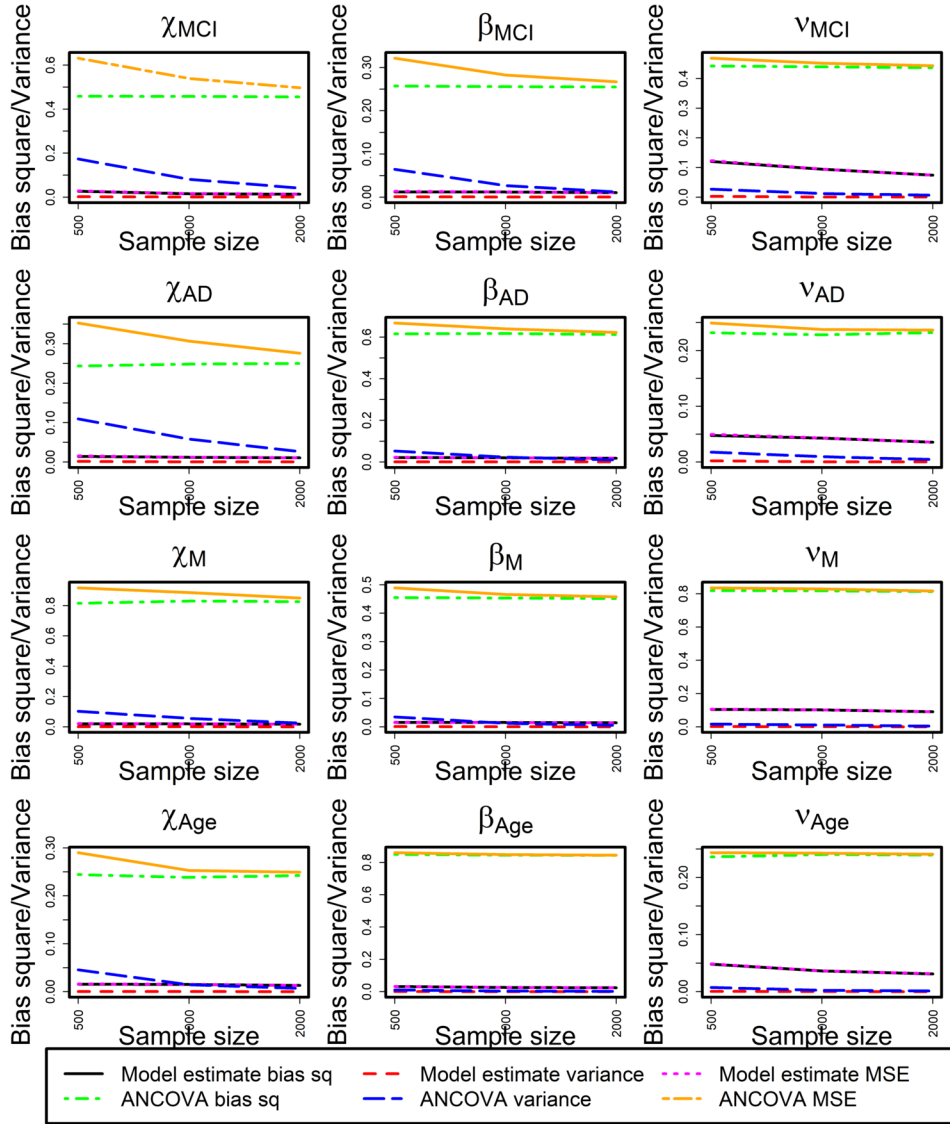


FIG. 4. Plot of squared bias and variance for different parameters against sample sizes for Bayesian method and ANCOVA estimates for large sample sizes 500, 1000 and 2000 with  $J = 20$ .

bution around the posterior mean. All of the estimated effects of the covariates for the top 10 significant edges are negative for the connection probability, that is, the probit regression part of the model as in Tables 5 to 8 in the Supplementary Material. For the number of connection, the estimated effects corresponding to AD, gender and age are mostly negative in their corresponding top ten edges. Some estimated effects corresponding to MCI are also negative. These results can be found

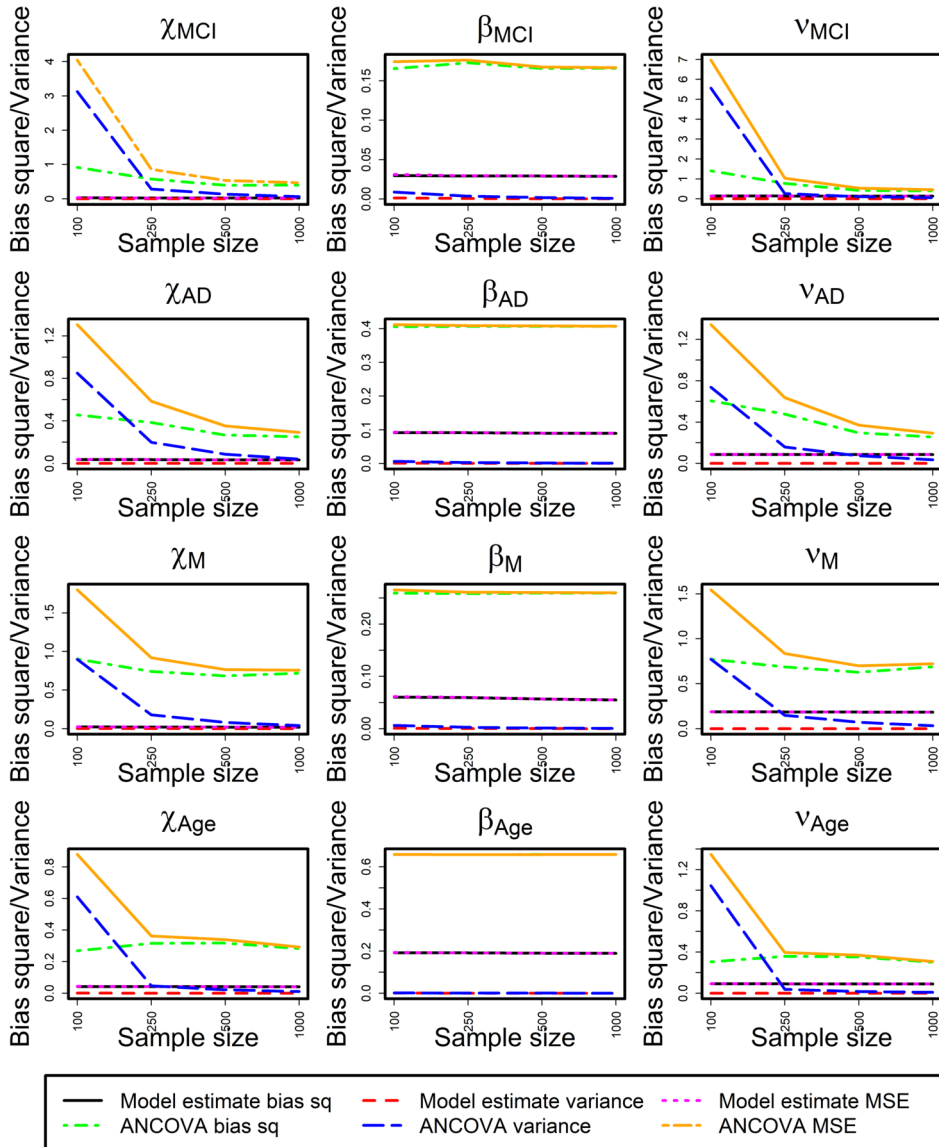


FIG. 5. Plot of squared bias and variance for different parameters against sample sizes for Bayesian method and ANCOVA estimates for small sample sizes 100, 250, 500 and 1000 with  $J = 40$ .

in Tables 9 to 12 in the Supplementary Material. For edge-length all the estimated effects corresponding to gender and age are negative among the top 10 significant edges in Tables 3 to 4 in the Supplementary Material. Some estimated effects of MCI and AD are negative in their corresponding top 10 significant edges as in the first two tables of the Supplementary Material (Roy et al. (2019)).

TABLE 6  
Demographic table

Covariates	Female	Male
No cognitive impairment (NC)	17	22
Alzheimer's disease (AD)	6	15
Mild cognitive impairment (MCI)	14	26
Average Age	71.80 (6.74)	75.15 (7.04)

Significant edges are plotted for each parameter of interest. The circles are the different regions, and their names are mentioned in the legend. These plots are in Figures 7, 8 and 9. We find that for each part of the model, that is, mean connection length, a number of connections and the binary variable signifying the presence or absence of connection, there are separate sets of regions that have the most number of significant edges, connected with them. In the Supplementary Material, there are additional tables containing top 10 significant edges for each covariate.

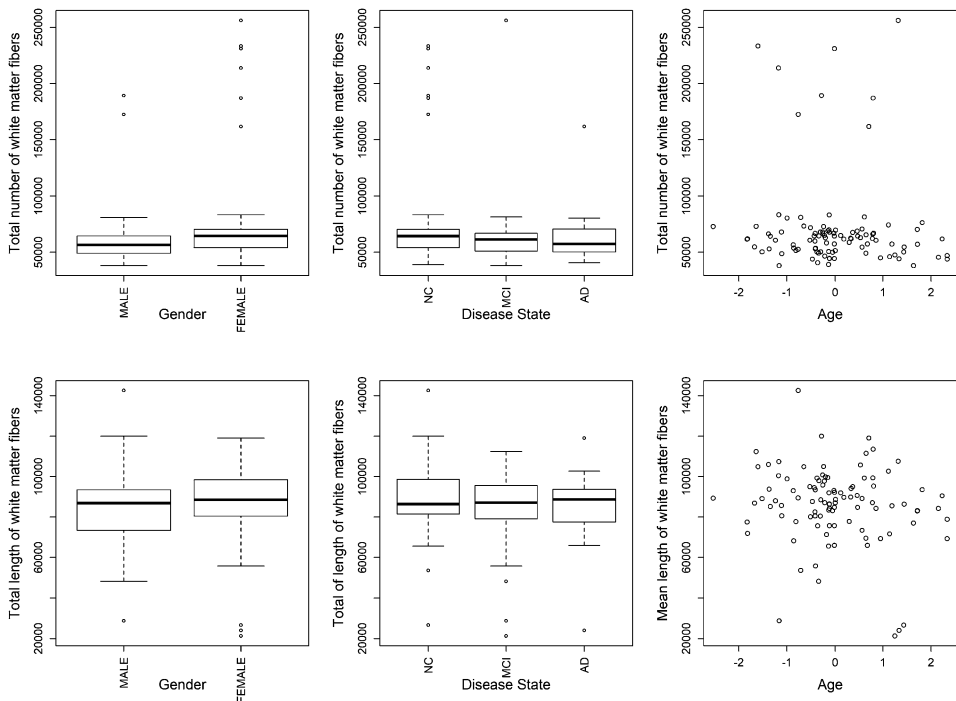


FIG. 6. Comparison of the distribution of total number of connected white matter fibers and their total length across gender, disease states and age.

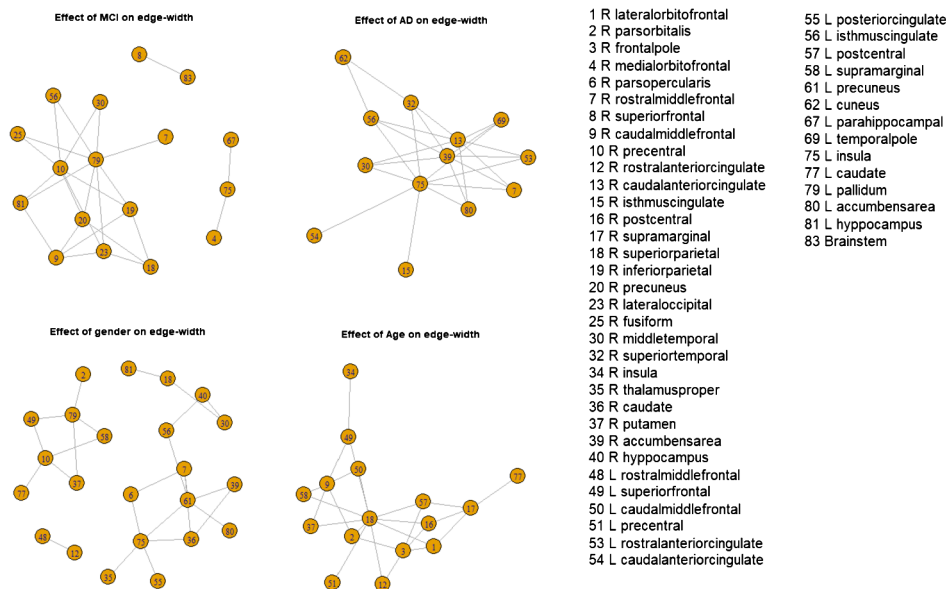


FIG. 7. Effect of different covariates on average connection length, each circle denotes different cortical brain regions.

There are some regions that have many connected edges in these plots like insula, pallidum, inferior temporal, parsorbital, precentral, posterior cingulate, superior temporal, superior parietal, middle temporal, paracentral, caudal middle frontal etc. In some of the previous studies on Alzheimer’s disease, they were mentioned. Some of those are mentioned in Section 6.

**6. Conclusions and discussion.** We study the effects of some common measurable covariates on the human brain connectome. Our work extends statistical inference on graph structure from the two sample problem (Tang et al. (2017)) to a more general regression modeling framework. We propose regression models to explain the extent of connections between different cortical brain regions. In this setup traditional techniques of separately regressing connections, between each pair of regions and the covariates, are not appropriate due to missingness at several edges of the connectome. We solve this problem by using graphon functions to reduce the dimension of the parameter space through a fewer number of fundamental parameters and develop a Bayesian method to estimate those. Subject inhomogeneity is incorporated through random effects, and the distributions of the random effects are estimated by using Dirichlet process scale mixture of normal (DPSMN) prior. Figure 2 suggests that our Bayesian method identifies the true structure for all the cases in the simulation setting. ANCOVA could not capture the true structures even when the sample size is as large as 2000. Figure 3 suggests that the small sample performance of ANCOVA is poor. The variance of the

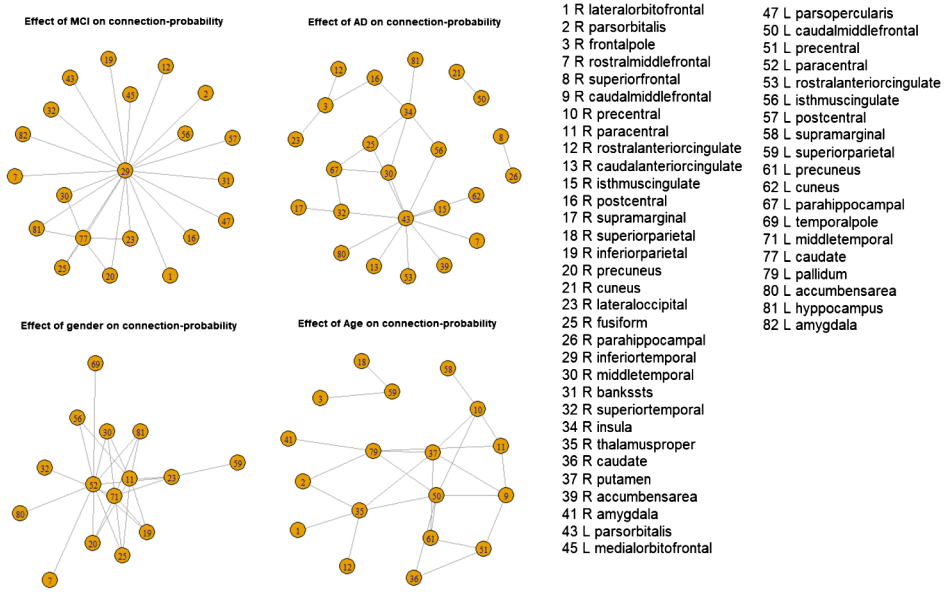


FIG. 8. *Effect of covariates on connection-probability, each circle denotes different cortical brain regions.*

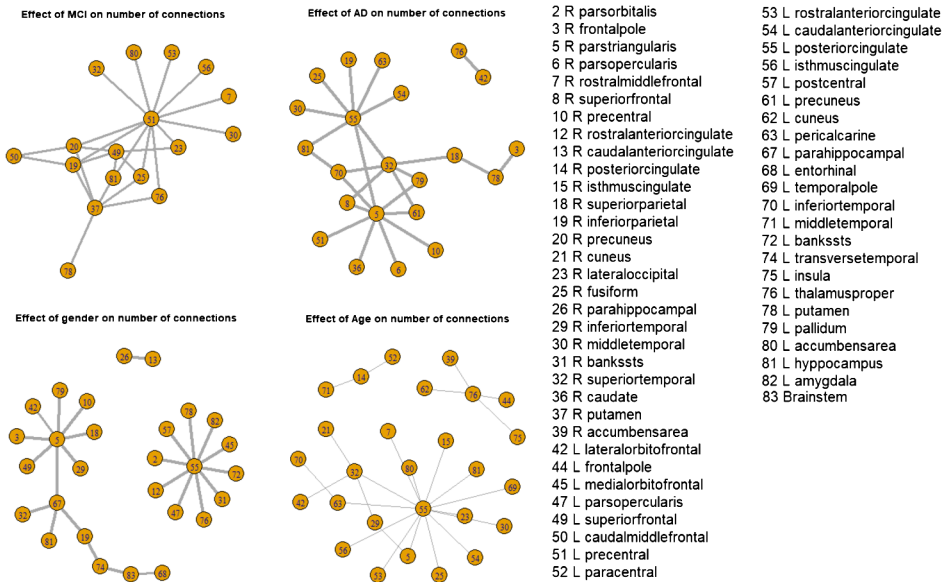


FIG. 9. *Effect of covariates on number of connections, each circle denotes different cortical brain regions.*

ANCOVA estimate is the key issue here. Even for large sample sizes, we conclude that the proposed Bayesian method performs better than ANCOVA as shown in Figure 4.

The regions that have many connected edges are summarized here. We also cite next to each region's name previous literature that studied those regions in the context of Alzheimer's disease or dementia and found significant. These regions are insula (Bonthius, Solodkin and Van Hoesen (2005)), precuneus (Karas et al. (2007), Klaassens et al. (2017)), pallidum (Lehéricy et al. (1991)), inferior temporal (Scheff et al. (2011)), parsorbital (McLimans and Willette (2016)), precentral (Canu et al. (2011)), posteriorcingulate (Leech and Sharp (2014)), superior temporal (Gao et al. (2018)), superior parietal (Migliaccio et al. (2015)), middle temporal (Jack et al. (1998)), paracentral (Karavasilis et al. (2017)) and caudal middle frontal (Bakkour et al. (2013)).

The greater magnitude of the negative effect on a number of white matter fiber connections between regions for AD subjects than for MCI subjects suggests that white matter connectivity is progressively degraded by lost connections across the clinical spectrum of dementia. In the current analysis the most significant relationships between a number of white matter connections and MCI or AD status are among widely distributed regions of the brain. In particular they include connections between regions in the right hemisphere and left hemisphere. They also predominantly involve the frontal, temporal and parietal lobes and the cingulate cortex. These regions are generally involved with the widespread damage associated with the AD. The preferential loss of "long-range" fibers (i.e., fibers between the cerebral hemispheres) in these widely distributed regions corresponds with prior work which has demonstrated that long-range connections become degraded with progression along the AD spectrum, leaving highly connected and short-range "hub" networks relatively intact until late in the disease course (Gao et al. (2014), Sanz-Arigita et al. (2010)).

**7. Acknowledgments.** Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research;

Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

We are grateful to two anonymous reviewers, the Associate Editor and the Editor for their valuable comments that have greatly helped to improve the manuscript.

### SUPPLEMENTARY MATERIAL

**Contents** (DOI: [10.1214/19-AOAS1257SUPP](https://doi.org/10.1214/19-AOAS1257SUPP); .pdf). There are four sections in this. Section S-1 describes posterior computation. Section S-2 contains tables that are mentioned in Section 5. Section S-3 contains more supporting plots and Section S-4 states and proves posterior consistency results of our proposed methodology.

### REFERENCES

- ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. [MR1224394](#)
- ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937](#)
- BAKKOUR, A., MORRIS, J. C., WOLK, D. A. and DICKERSON, B. C. (2013). The effects of aging and Alzheimer's disease on cerebral cortical anatomy: Specificity and differential relationships with cognition. *NeuroImage* **76** 332–344.
- BONTHIUS, D. J., SOLODKIN, A. and VAN HOESEN, G. W. (2005). Pathology of the insular cortex in Alzheimer disease depends on cortical architecture. *J. Neuropathol. Exp. Neurol.* **64** 910–922.
- CANU, E., MCLAREN, D. G., FITZGERALD, M. E., BENDLIN, B. B., ZOCCATELLI, G., ALESSANDRINI, F., PIZZINI, F. B., RICCIARDI, G. K., BELTRAMELLO, A. et al. (2011). Mapping the structural brain changes in Alzheimer's disease: The independent contribution of two imaging modalities. *J. Alzheimer's Dis.* **26** 263–274.
- DAIANU, M., JAHANSHAD, N., NIR, T. M., JACK JR, C. R., WEINER, M. W., BERNSTEIN, M. A., THOMPSON, P. M. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2015). Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network. *Hum. Brain Mapp.* **36** 3087–3103.
- GAO, J., CHEUNG, R. T., CHAN, Y.-S., CHU, L.-W., MAK, H. K. and LEE, T. M. (2014). The relevance of short-range fibers to cognitive efficiency and brain activation in aging and dementia. *PLoS ONE* **9** e90307.
- GAO, Z., FU, H.-J., ZHAO, L.-B., SUN, Z.-Y., YANG, Y.-F. and ZHU, H.-Y. (2018). Aberrant DNA methylation associated with Alzheimer's disease in the superior temporal gyrus. *Exp. Ther. Med.* **15** 103–108.
- GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics **44**. Cambridge Univ. Press, Cambridge. [MR3587782](#)



- HOOVER, D. (1979). Relations on Probability Spaces and Arrays of Random Variables. Preprint, Institute for Advanced Study, Princeton, NJ.
- JACK, C. R., PETERSEN, R. C., XU, Y., O'BRIEN, P. C., SMITH, G. E., IVNIK, R. J., TANGALOS, E. G. and KOKMEN, E. (1998). Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* **51** 993–999.
- KARAS, G., SCHELTENS, P., ROMBOUTS, S., VAN SCHIJNDEL, R., KLEIN, M., JONES, B., VAN DER FLIER, W., VRENKEN, H. and BARKHOF, F. (2007). Precuneus atrophy in early-onset Alzheimer's disease: A morphometric structural MRI study. *Neuroradiology* **49** 967–976.
- KARAVASILIS, E., PARTHIMOS, T. P., PAPATRIANTAFYLLOU, J. D., PAPAGEORGIOU, S. G., KAPASAS, G., PAPANICOLAOU, A. C. and SEIMENIS, I. (2017). A specific pattern of gray matter atrophy in Alzheimer's disease with depression. *J. Neurolinguist.* **264** 2101–2109.
- KLAASSENS, B. L., VAN GERVEN, J. M. A., VAN DER GROND, J., DE VOS, F., MÖLLER, C. and ROMBOUTS, S. A. R. B. (2017). Diminished posterior precuneus connectivity with the default mode network differentiates normal aging from Alzheimer's disease. *Front. Aging Neurosci.* **9** 97.
- LEECH, R. and SHARP, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain* **137** 12–32.
- LEHÉRICY, S., HIRSCH, E. C., HERSH, L. B. and AGID, Y. (1991). Cholinergic neuronal loss in the globus pallidus of Alzheimer disease patients. *Neurosci. Lett.* **123** 152–155.
- LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. [MR2274085](#)
- MAEHLER, M. (2013). Package 'diptest'. In *R Package Version* 0.75–5.
- MCLIMANS, K. E. and WILLETTE, A. A. (2016). Novel CSF biomarker of metabolic dysfunction predicts AD-like associations across the Alzheimer's spectrum. *FASEB J.* **30** lb304–lb304.
- MIGLIACCIO, R., AGOSTA, F., POSSIN, K. L., CANU, E., FILIPPI, M., RABINOVICI, G. D., ROSEN, H. J., MILLER, B. L. and GORNO-TEMPINI, M. L. (2015). Mapping the progression of atrophy in early- and late-onset Alzheimer's disease. *J. Alzheimer's Dis.* **46** 351–364.
- PHILLIPS, D. J., MCGLAUGHLIN, A., RUTH, D., JAGER, L. R., SOLDAN, A., ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE et al. (2015). Graph theoretic analysis of structural connectivity across the spectrum of Alzheimer's disease: The importance of graph creation methods. *NeuroImage: Clinical* **7** 377–390.
- PRESCOTT, J. W., GUIDON, A., DORAISWAMY, P. M., ROY CHOUDHURY, K., LIU, C., PETRELLA, J. R. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2014). The Alzheimer structural connectome: Changes in cortical network topology with increased amyloid plaque burden. *Radiology* **273** 175–184.
- ROY, A., GHOSAL, S., PRESCOTT, J., CHOUDHURY, K. R. and FOR THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2019). Supplement to “Bayesian Modeling of the Structural Connectome for Studying Alzheimer's Disease.” DOI:[10.1214/19-AOAS1257SUPP](#).
- SANZ-ARIGITA, E. J., SCHOONHEIM, M. M., DAMOISEAUX, J. S., ROMBOUTS, S. A. R. B., MARIS, E., BARKHOF, F., SCHELTENS, P. and STAM, C. J. (2010). Loss of 'small-world' networks in Alzheimer's disease: Graph analysis of FMRI resting-state functional connectivity. *PLoS ONE* **5** e13788.
- SCHEFF, S. W., PRICE, D. A., SCHMITT, F. A., SCHEFF, M. A. and MUFSON, E. J. (2011). Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease. *J. Alzheimer's Dis.* **24** 547–557.
- SHEN, W. and GHOSAL, S. (2015). Adaptive Bayesian procedures using random series priors. *Scand. J. Stat.* **42** 1194–1213. [MR3426318](#)
- TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *J. Comput. Graph. Statist.* **26** 344–354. [MR3640191](#)

- TUCHOLKA, A., GRAU-RIVERA, O., FALCON, C., RAMI, L., SANCHEZ-VALLE, R., LLADÓ, A., GISPERT, J. D., MOLINUEVO, J. L., INITIATIVE, A. D. N. et al. (2018). Structural connectivity alterations along the Alzheimer's disease continuum: Reproducibility across two independent samples and correlation with cerebrospinal fluid amyloid- $\beta$  and tau. *J. Alzheimer's Dis.* **61** 1575–1587.
- VAN BUUREN, S. and GROOTHUIS-OUUDSHOORN, K. (2010). Mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* 1–68.
- WEST, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74** 646–648. [MR0909372](#)

A. ROY  
DEPARTMENT OF STATISTICS  
DUKE UNIVERSITY  
DURHAM, NORTH CAROLINA  
USA  
E-MAIL: [arkaprava.roy@duke.edu](mailto:arkaprava.roy@duke.edu)

J. PRESCOTT  
DEPT. OF RADIOLOGY  
METROHEALTH MEDICAL CENTER  
CLEVELAND, OHIO  
USA  
E-MAIL: [jwprescott@gmail.com](mailto:jwprescott@gmail.com)

S. GHOSAL  
DEPARTMENT OF STATISTICS  
NORTH CAROLINA STATE UNIVERSITY  
RALEIGH, NORTH CAROLINA  
USA  
E-MAIL: [sghoshal@ncsu.edu](mailto:sghoshal@ncsu.edu)

K. ROY CHOUDHURY  
DEPT. OF BIostatISTICS AND BIOINFORMATICS  
DUKE UNIVERSITY MEDICAL CENTER  
DURHAM, NORTH CAROLINA  
USA  
E-MAIL: [kingshuk.roy.choudhury@duke.edu](mailto:kingshuk.roy.choudhury@duke.edu)