



Published in final edited form as:

*Med Image Anal.* 2025 January ; 99: 103309. doi:10.1016/j.media.2024.103309.

## Establishing group-level brain structural connectivity incorporating anatomical knowledge under latent space modeling

Selena Wang<sup>\*1</sup>, Yiting Wang<sup>2</sup>, Frederick H. Xu<sup>3</sup>, Xinyuan Tian<sup>4</sup>, Carolyn A. Fredericks<sup>5</sup>, Li Shen<sup>6</sup>, Yize Zhao<sup>4</sup>, Alzheimer's Disease Neuroimaging Initiative<sup>†</sup>

<sup>1</sup>Department of Biostatistics and Health Data Science, Indiana University School of Medicine, 410 West 10th Street, Suite 3000, Indpls., IN 46202

<sup>2</sup>Department of Statistics, University of Virginia, 148 Amphitheater Way, Charlottesville, VA 22904-4135

<sup>3</sup>Department of Bioengineering, University of Pennsylvania, 240 Skirkanich Hall, 210 S 33rd St, Philadelphia, PA 19104

<sup>4</sup>Department of Neurology, Yale School of Medicine, Yale University, LLCI 912 New Haven, CT 06520

<sup>5</sup>Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, 423 Guardian Dr, Philadelphia, PA 19104

<sup>6</sup>Department of Biostatistics, Yale School of Public Health, 60 College St, New Haven, CT 06510

### Abstract

Brain structural connectivity, capturing the white matter fiber tracts among brain regions inferred by diffusion MRI (dMRI), provides a unique characterization of brain anatomical organization. One fundamental question to address with structural connectivity is how to properly summarize and perform statistical inference for a group-level connectivity architecture, for instance, under different sex groups, or disease cohorts. Existing analyses commonly summarize group-level brain connectivity by a simple entry-wise sample mean or median across individual brain connectivity matrices. However, such a heuristic approach fully ignores the associations among structural connections and the topological properties of brain networks. In this project, we propose a latent space-based generative network model to estimate group-level brain connectivity. Within our

<sup>\*</sup>The correspondence should be addressed to Selena Wang, selewang@iu.edu.

<sup>†</sup>Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.ucla.edu](http://adni.loni.ucla.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).

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

modeling framework, we incorporate the anatomical information of brain regions as the attributes of nodes to enhance the plausibility of our estimation and improve biological interpretation. We name our method the attributes-informed brain connectivity (ABC) model, which compared with existing group-level connectivity estimations, (1) offers an interpretable latent space representation of the group-level connectivity, (2) incorporates the anatomical knowledge of nodes and tests its co-varying relationship with connectivity and (3) quantifies the uncertainty and evaluates the likelihood of the estimated group-level effects against chance. We devise a novel Bayesian MCMC algorithm to estimate the model. We evaluate the performance of our model through extensive simulations. By applying the ABC model to study brain structural connectivity stratified by sex among Alzheimer's Disease (AD) subjects and healthy controls incorporating the anatomical attributes (volume, thickness and area) on nodes, our method shows superior predictive power on out-of-sample structural connectivity and identifies meaningful sex-specific network neuromarkers for AD.

---

## 1 Introduction

Brain structural connectivity captures the white matter fiber tracts among brain regions inferred via diffusion MRI (dMRI) and provides a unique characterization of brain anatomical network organization. Statistical analysis at the group and population level of structural connectivity represents a fundamental component to dissect brain structural alterations and uncover significant neural substrates for cognitive and disease processes. How to properly summarize and perform statistical inference for a group-level connectivity architecture, for instance under different sex groups or disease cohorts, are important statistical issues that warrant careful considerations.

Most of the existing methods estimating group-level connectivity are in light of an underlying assumption that all the connections across the brain are independent from each other. As a result, group-level brain connectivity is constructed based on the sample mean or median across the group with respect to each individual edge (Achard et al., 2006; Sinke et al., 2016; Song et al., 2009), ignoring the structural dependence and higher-order topological properties within brain networks (Rubinov & Sporns, 2010). Alternatively, some methods seek to select the “best” individual network based on certain selection criteria (Joyce et al., 2010; Meunier et al., 2009). However, the processes neglect the heterogeneity among the individual connectivity, and the resulting group-level brain connectivity likely includes features not present in the majority of the subjects. There is still a lack of methods that can properly estimate group-level connectivity accommodating the topological structure and are capable of quantification of uncertainty in order to evaluate the likelihood of the group difference effects against chance.

In addition, converging evidence indicates that the anatomical attributes including cortical volume, surface area and cortical thickness of brain regions are associated with the white matter fiber tracks among them (e.g., Cai et al., 2021; Feng et al., 2021; Hodel et al., 2020; Yee et al., 2018). However, existing approaches do not consider or offer an intuitive framework to incorporate anatomical knowledge of the nodes when modeling their corresponding structural connections. This is particularly crucial when studying

neurodegenerative diseases. For example, Cai et al. (2021) found that the surface area of the brain regions and their connectivity components change in a coherent way for Alzheimer's Disease (AD) patients. Uncovering the interplay between regional and network-level brain structural alternations could provide an unprecedented opportunity to enhance our understanding on the disease etiology and neurodegeneration onset. There is an urgent need to integrate regional and connectivity measures together and investigate their relationships and how they change under different subpopulations.

With these considerations and challenges, we propose an attributes-informed brain connectivity (ABC) modeling framework that estimates the group-level connectivity among the brain regions while incorporating the anatomical attributes of the brain regions. Different from the existing approaches, we borrow inspirations from the generative latent space network models that emphasize modeling the dynamic and complex higher-order topological properties of the individual networks (Hoff et al., 2002). Attention has been given to the use of generative network models in structural and functional analyses (Betzel et al., 2016), and the group-level brain connectivity has been estimated using exponential random graph models (Lehmann et al., 2021) and the latent distance models (Aliverti & Durante, 2019; Wilson et al., 2020). Different from these existing methods, we propose a joint modeling framework that allows for the incorporation and investigation of the anatomical attributes of the brain regions when estimating group-level connectivity, and in the proposed ABC model, the two modalities regulate and inform each other in the estimation process. By combining the connectivity and the anatomical attributes into one analysis, we leverage the strengths of each structural knowledge component to investigate the connectivity changes. With the added anatomical knowledge, we expect the proposed method to show superior estimation with better precision and interpretability.

The proposed ABC model contributes to the current literature in multiple ways. First, we offer an interpretable latent space representation of the structural connectivity which accommodates the topological structure. The latent space shows a clear differentiation of the left and right hemispheres when fitted to data in line with the intrinsic structural anatomy of the brain. Regions of the cingulate cortex wrapping around the corpus callosum are found in the direction separate from the left and right hemispheres reflecting the unique feature of fiber connectivity around the cingulate cortex. Second, the ABC model allows us to incorporate the related anatomical knowledge of the brain regions while estimating group-level brain connectivity. Such a process has been shown in our numerical studies to improve the estimation of group-level connectivity, resulting in superior predictive power and interpretability. The ABC model also allows us to uncover the co-varying relationship between regional anatomical knowledge and the corresponding structural connectivity. Furthermore, the ABC model allows us to quantify the uncertainty and evaluate the likelihood of the estimated group-level effects against chance. Existing heuristic approaches often apply the unrealistic assumption that the brain connectivity edges are static and fixed observations. In comparison, the ABC model, as a generative statistical network model, assumes the connectivity-induced elements to be realizations of random processes with noise and proposes a data generation process for the observed ones. This allows us to assess the uncertainty of the estimated group-level brain connectivity and investigate the group-level differences against chance. In this aim, we apply the ABC model to study

sex-specific AD neuromarkers based on a joint modeling of structural connectivity and volumetric measures.

The remaining paper is organized as follows. Section 2 describes the data and the proposed methodology. Section 3 details the optimization routine for estimating the group-level connectivity. In section 4, we assess whether the proposed method compares well against the existing methods based on the predictive power and whether/how the predictive power improves with the added anatomical knowledge. Section 5 shows the application of our method to Alzheimer's Disease Neuroimaging Initiative (ADNI) data. We conclude with a brief discussion in Section 6.

## 2 Data and Methods

### 2.1 ADNI data preprocessing

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (Weiner et al., 2013; Weiner et al., 2010). The ADNI was launched in 2003 as a public-private partnership with the primary goal of testing whether serial magnetic resonance imaging, positron emission tomography, other biological markers and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early AD. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

**Structural connectivity:** The ADNI T1-weighted structural MRI (sMRI) and Diffusion Tensor Imaging (DTI) were obtained from the ADNI Grand Opportunities and ADNI Phase 2 (ADNI-GO/2) databases. We downloaded the imaging data for 173 subjects along with their demographic information. There are 99 males and 74 females in the data with an average age of  $72.9 \pm 7.39$  years. No significant difference in age was detected across different groups (ANOVA:  $P = 0.805$ ,  $F = 0.217$ ).

The DTI data were first processed under standard steps including denoising, motion-correction and distortion-correction using an overcomplete local principal components analysis (Manjón et al., 2013). Probabilistic white matter fiber tractography was performed using a streamline tractography algorithm called fiber assignment by continuous tracking (FACT)(Moore & Sciacca, 2019). SMRI scans were registered to the lower resolution b0 volume of the DTI data using the FLIRT toolbox in the FMRIB Software Library (Jenkinson et al., 2012), and the cortical ROIs were defined based on the Lausanne 2008 parcellation with 68 regions of interest in the native FreeSurfer space (Cammoun et al., 2012). An average of 3.2 million fibers were extracted for each brain network. The number of the fibers (NOF) connecting each pair of ROIs was obtained as well as the regions' surface area (SA). The fiber density-based structural connectivity was calculated by dividing NOF between two ROIs with their average surface areas (Xu et al., 2022; Yan et al., 2020). The structural brain networks were constructed as the fiber density of tracts connecting pairs of ROIs.

The tractography strategy used in this manuscript was published in Xu et al. (2022), where the same set of the brain networks was studied in a different analysis. When processing the DTI data, we used the probabilistic approach to efficiently extract information from the

underlying data (Zolal et al., 2016). We provided visualizations for the fiber density and fiber counts using the proposed tractography (see supplementary materials), which showed similar sparsity as other existing literatures on DTI data (e.g., Sinke et al., 2016; Song et al., 2009). Nonetheless, the probabilistic approach may lead to more false positives that warrant caution for interpretation. Future investigations are needed to further explore the costs and benefits of the probabilistic vs deterministic tractography approaches.

**Anatomical attributes:** The regional anatomical measurements were acquired from the University of California San Francisco FreeSurfer 5.1 Cross-Sectional Study database. The Cross-Sectional study uses ADNIGO/2 T1.5 scans that have been motion-corrected, B1-corrected and N3-inhomogeneity corrected by MayoClinic (Hartig et al., 2014). Each scan was segmented in the native FreeSurfer space allowing for comparison between subjects at each time point (Fischl & Dale, 2000a). The cortical ROI attribute data were obtained from the screening sessions. The cortical attributes include the cortical volume, surface area, average thickness and thickness standard deviation. The highly folded nature of the cortex makes quantifying the cortex challenging (Fischl & Dale, 2000b). FreeSurfer methods address this difficulty and provide the thickness as an averaged measure across a reconstructed cortical surface (Fischl et al., 1999), along with its standard deviation both of which are utilized in our study.

## 2.2 The ABC model

In this study, we have a set of brain structural connectivity matrices  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$ , each with  $V \times V$  dimensions, where  $V$  is the number of nodes or brain regions, and  $N$  is the sample size. For subject  $i$ , each element  $x_{v,v',i}$  in the connectivity matrix measures the strength of the anatomical connection between brain regions  $v$  and  $v'$ ,  $v < v'$ ,  $v, v' = 1, 2, \dots, V$ ; and we have  $x_{v,v',i} = x_{v',v,i}$  following the symmetry of connectivity matrix and  $x_{v,v,i} = 0$ . Simultaneously, we also measure a series of anatomical information e.g. volume, area and thickness upon each of the nodes for each subject and summarize them by a set of anatomical attribute matrices  $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N\}$ , each with  $V \times P$  dimensions, where  $P$  is the number of anatomical metrics, or attributes. We assume the structural connections on the edges interact with the anatomical structure of the nodes. Given these two data components, the likelihood of the ABC model can be written as:

$$p(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N \mid \mathbf{Z}, a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, \sigma^2, \tau^2, \boldsymbol{\Theta}, \boldsymbol{\Sigma}) = \prod_{i=1}^N p(\mathbf{X}_i, \mathbf{Y}_i \mid \mathbf{Z}, a_i, b_i, \sigma^2, \tau^2, \boldsymbol{\Theta}, \boldsymbol{\Sigma}), \quad (2.1)$$

where we assume that the individual brain connectivity and attribute matrices are generated from a set of shared latent variables  $\mathbf{Z}$  and  $\boldsymbol{\Theta}$ . Note that different from a model that has only the brain connectivity or the brain attributes as the response variables, the proposed ABC model uses both the brain connectivity and the nodal attributes as the response variables and thus allows the two forms of information to mutually inform each other. To characterize the latent topological structure as well as link, we introduce latent space modeling for the connectivity component, the attribute component and the joint component linking the

previous two components. For subject  $i$ , the connectivity between nodes  $v$  and  $v'$  is modeled by

$$x_{v,v',i} = \mathbf{w}_i^T \boldsymbol{\beta} + a_i + \mathbf{z}_v^T \mathbf{z}_{v'} + e_{v,v',i}, \quad e_{v,v',i} \stackrel{iid}{\sim} N(0, \sigma^2), \quad (2.2)$$

where  $\mathbf{z}_v$  is the  $K$ -dimensional,  $\mathbf{z}_v \in \mathbb{R}^K$ ,  $K \ll V$  vector containing the latent variable values for node  $v$ ;  $a_i$  is the fixed intercept for individual  $i$  modeling the overall connection, and for identification purposes, the sum of the intercepts across subjects is constrained to be zero;  $e_{v,v',i}$  is the error term associated with the connectivity edge between nodes  $v$  and  $v'$  for individual  $i$ ; and  $\sigma^2$  is the error variance. We also adjust for  $Q$  covariates denoted by  $\mathbf{w}_i$  with the effects of the covariates on the overall connectivity characterized by  $\boldsymbol{\beta}$ . Region-specific effects such as gender and AD should be investigated via comparing the group-level connectivities. The approximation of the posterior distributions for the unknown quantities is facilitated by setting a  $N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}^{-1})$ ,  $\boldsymbol{\beta}_0 = (0, 0, \dots, 0, 0)^T$ ,  $\boldsymbol{\Sigma} = \mathbf{I}_Q$  prior distribution for  $\boldsymbol{\beta}$ , noninformative priors including a gamma (1/2, 1/2) distribution for  $\sigma_e^{-2}$  and a  $N(0, 1)$  distribution for  $a_i$ . We leave the description on the prior distributions for the latent network variables afterwards.

For subject  $i$ , the attribute value of node  $v$  on attribute  $p$  is modeled by

$$y_{v,p,i} = \mathbf{h}_i^T \boldsymbol{\gamma} + b_i + \theta_{v,p} + \epsilon_{v,p,i}, \quad \epsilon_{v,p,i} \stackrel{iid}{\sim} N(0, \tau^2), \quad (2.3)$$

where  $b_i$  is the fixed intercept for individual  $i$  accounting the overall difference across subjects, and for identification purposes, its sum across subjects is zero. The latent variable  $\theta_{v,p}$  represents the average attribute value for node  $v$  and attribute  $p$  across subjects. In the vector form,  $\boldsymbol{\theta}_v$  is a  $P$ -dimensional vector that follows a Multivariate Normal distribution defined in the joint component. The error term is denoted by  $\epsilon_{v,p,i}$ ; and  $\tau^2$  is the error variance. We adjust for  $Q'$  individual-specific covariates,  $\mathbf{h}_i$ , which can be the same as or different from  $\mathbf{w}_i$  in model (2.2). The effects of the covariates on the attributes of the brain regions are characterized by  $\boldsymbol{\gamma}$ . Parts of the attribute model mirror those of the connectivity model, and other parts are distinct. Each letter symbol is used to represent a distinct variable to maintain clarity. While the vector products of the latent variables are often used to model network connectivity data in order to capture the topological properties of the network (Hoff, 2008; Wasserman & Faust, 1994), they are not applicable for the attributes data. The attributes data usually do not have the same dependence structures as the networks, and when there are only a few attributes, latent dimensions are not warranted (Wang et al., 2023), as is the case for our study. Approximation of the posterior distribution for the  $\boldsymbol{\gamma}$  is facilitated by setting a  $N(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$ ,  $\boldsymbol{\mu}_\gamma = (0, 0, \dots, 0, 0)^T$ ,  $\boldsymbol{\Sigma}_\gamma = \mathbf{I}_{Q'}$ . Similarly,  $N(0, 1)$  and gamma (1/2, 1/2) prior distributions are assigned for  $b_i$  and  $\tau_i^{-2}$ , respectively. The prior distributions for the latent variables of the attributes are described in the following joint modeling.

To further characterize the association between structural connections and regional anatomical information, we integrate the latent components from the two parts through a joint model written as:

$$(\mathbf{z}_u, \theta_u) \overset{\text{Tiid}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0_K \\ 0_P \end{pmatrix}, \boldsymbol{\Sigma}\right), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_z & \boldsymbol{\Lambda}_{zy}^T \\ \boldsymbol{\Lambda}_{zy} & \boldsymbol{\Lambda}_y \end{pmatrix}, \quad (2.4)$$

where  $\boldsymbol{\Lambda}_z$  is the  $K \times K$  covariance matrix for the connectivity, and  $\boldsymbol{\Lambda}_y$  is the  $P \times P$  covariance matrix for the attributes. The off-diagonal  $P \times K$  matrix  $\boldsymbol{\Lambda}_{zy}$  captures the association between the connectivity and the node attributes, where each of their entries describes the relationship between a latent connectivity variable and a latent attribute variable. When there are non-zero elements in the  $\boldsymbol{\Lambda}_{zy}$  matrix, the connectivity and node attribute estimations regulate and inform each other, providing a more informative and plausible result. We assign a Wishart ( $\mathbf{I}_{K+P}, K+P+2$ ) prior distribution for  $\boldsymbol{\Sigma}$ , facilitating nonzero posterior correlation between brain structural connectivity and node anatomical structures.

To summarize, we demonstrate the ABC framework in Figure 1. The ABC model takes both brain connectivity and regional attributes as the inputs. Within the model, through shared group-level latent variables, we induce shared brain connectivity components across subjects. As illustrated by the existing precedents in network literature (Gollini & Murphy, 2016; Wang et al., 2023), such an approach using shared group-level latent variables is a succinct way to synthesize information across subjects and to improve the estimation and prediction of the model. To link brain connectivity and regional attributes, we propose a shared distribution of the latent variables and allow their relationship to be represented in an intuitive way through covariance parameters. The benefit of this approach is that the relationship is more realistically represented as co-varying, where both the regional attributes and the brain connectivity influence each other instead of only allowing one to influence the other. Current literature supports this representation with studies showing that structural connectivity explains regional attributes (Yee et al., 2018), and vice versa (Cai et al., 2021; Hodel et al., 2020; Yee et al., 2018). Using the ABC model, we produce three types of outputs: (1) the attributes-informed group-level connectivity and its latent space representation, (2) the connectivity difference between subject groups and (3) the dependence between the brain connectivity and the attributes. These outputs will be illustrated with the application. The code is available at <https://github.com/selenashuowang/ABCModel.git>.

### 2.3 Related Network Models

In recent years, we see a surge of methodology developments to analyze network data in many domains; see surveys of network methods in Desmarais and Cranmer (2017), Goldenberg et al. (2010), Kim et al. (2018), Smith et al. (2019), Snijders (2011), Sweet (2016), Vivar and Banks (2012), and Wang (2021). Latent variables-based method is a popular modeling choice to visualize the relational dependence structure and account for the dependence structure while performing analytical tasks (see discussions in Sweet, 2016). Popular latent variable network models include stochastic block models (also called latent

class models)(Holland et al., 1983) and latent space models with latent spaces defined by Euclidean distances (Hoff et al., 2002) and vector product-based latent space models (Hoff, 2008, 2021; Hoff, 2005, 2009). Between the latent class models (for identifying clusters) and the vector product-based latent space models, Hoff (2007) provides the theoretical evidence that the vector product-based network model generalizes the latent class model, not vice versa; and Sosa and Buitrago (2021) provides empirical evidence that the vector product-based model shows good prediction accuracy for different types of networks while the latent class model only shows good fit for networks with high levels of clustering. Given our modeling objective to capture group-level connectivity, it is optimal to use the vector product-based latent space model. The proposed ABC model extends the latent space model to accommodate multiplex brain network structure and identify significant links between brain and behavior.

Multiplex networks are networks with multiple layers of connections observed on a shared set of nodes. Structural connectivity networks can be seen as a type of multiplex networks with individual-specific white matter fiber connections on a shared set of brain region nodes. Latent space models for multiplex networks have been proposed via a shared set of latent variables across layers (D'Angelo et al., 2018; Gollini & Murphy, 2016) or by separating the shared and separate components across layers (MacDonald et al., 2022; Salter-Townshend & McCormick, 2017). To capture the group-level white matter fiber connectivity, we propose a shared structural latent space across individuals in the group; distinct from the latent space joint model by Gollini and Murphy (2016), we propose the ABC model for data integration, incorporating a secondary mode of information about the brain regions, the region-specific volumetric measures. We allow the volumetric measures to inform the estimated group-level structural connectivity in a data-driven fashion.

The proposed ABC model is also distinct from existing latent space network models in its integration of the regional or nodal attribute information. Existing latent variables-based network models often incorporate nodal attributes information as covariates to model their influences on link probabilities while conditioning on the latent variables (Hoff, 2009; Hoff et al., 2002; Ma et al., 2020). The resulting latent space is influenced by whether and which covariates are included in the model. Alternatively, the relationship between nodal attributes and the latent variables can be captured by a shared latent variable between the network and the attributes (Gu & Yu, 2022; Wang, Paul, et al., 2023; Zhang et al., 2022) or by modeling the latent network variables as functions of the covariates in the covariate-assisted modeling (Austin et al., 2013; Yang et al., 2013). A drawback of the joint latent space model is that there is a lack of quantification of the relationship between the latent variables and the attributes. On the other hand, the covariate-assisted modeling loses the flexibility of the latent space modeling, where the relative positions of the estimated latent variables are direct functions of covariates and do not reflect network dependence.

We overcome the limitations of these existing methods by proposing a joint data generation process for multiplex networks and the observed volumetric measures to allow the network and the nodal attributes to co-vary in a data-driven fashion. In this aim, we build on the work by Fosdick and Hoff (2015), where a joint model is used to test potential dependencies between a network and nodal attributes. However, Fosdick and Hoff (2015) does not account

for the multiplex brain network structures, nor does it take into account the potentially high dimensions of the attributes and potential noise and measurement errors. In sum, the proposed ABC model uniquely contributes to the current network modeling landscape and address challenges associated with the structural brain network data in neuroinformatics.

## 2.4 Advantages

In this paper, we use the latent space network models to capture complex dependence patterns in brain connectivity networks. Third-order dependencies such as transitivity exist in brain connectivity networks as connectivity edges are dependent among a triad of brain regions. When two brain regions are mutually connected with the same brain region, it is likely that these two brain regions are also connected. These type of dependence structures play importance roles in the overall communicative structure of the brain (Opsahl et al., 2010), as well as neurodevelopment and neurodegeneration (Bassett & Sporns, 2017; Bullmore & Sporns, 2009; Zuo et al., 2012). In addition, the communicative structure of the brain networks such as the centrality of the brain regions reflects the configuration and organization of the network and how they might be altered with disfunction and neurodegeneration (Bassett & Sporns, 2017; Bullmore & Sporns, 2009; Zuo et al., 2012). Simple linear additive effects are only able to capture the heterogeneity across the rows and the columns of the network, reflected as variations in node degrees or strength; but they cannot capture higher-order dependencies among three or more brain regions. Network-based methods are advantageous against linear additive effects models as they are able to capture distributed patterns of interactions and account for the complexity of integrated systems not possible with additive effects.

The proposed ABC model is advantageous against the decomposition methods such as the spectral decomposition of connectomes. First, because the proposed ABC model theorizes the data generation process for the observed connectivity, it separates meaningful information (captured by the latent connectivity variables) from random noise. In this way, the estimated connectivity is less noisy, and we are able to quantify the uncertainty of the estimated group-level connectivity for case-control comparisons. In other words, the probabilistic framework for the observed connectivity allows us to make statistical inference not possible with decomposition methods. Second, the ABC model is able to integrate multiple data types in a flexible way. For example, we are able to integrate volumetric measurements of the brain regions when estimating the group-level connectivity in an intuitive and data-driven fashion. In contrast, it is much more difficult and less intuitive to integrate two data types using the spectral decomposition of connectomes while preserving interpretability. Lastly, the proposed ABC model is able to estimate group-level connectivity while accounting for potential covariate effects which is less intuitive with the spectral decomposition.

## 3 Bayesian Inference

The existing latent space network modeling literature focuses on either modeling networks as the response and nodal attributes as the covariates (e.g., Hoff et al. (2002), Krivitsky and Handcock (2008), and Sweet (2015)) or modeling attributes as the response and

networks as the predictors (e.g., Sweet and Adhikari (2020)). Therefore, the proposed ABC model cannot be estimated by existing software. In this section, we propose a Bayesian posterior inference algorithm to estimate the proposed joint model using Markov chain Monte Carlo (MCMC) methods. With sufficient iterations, we obtain stable Markov chains to approximate various quantities of the targeted posterior distributions. With random initial values on the unknown parameters, we conduct posterior computation by iterating the following steps. We also briefly summarize the procedure in Algorithm 1.

- For connectivity parameters,
  - simulate  $\beta, \mathbf{a}$  from their full conditional distributions.
  - simulate  $\sigma^2$  given  $\beta, \mathbf{a}, \tau^2, \gamma, \mathbf{b}, \mathbf{Z}, \Theta, \Sigma, \mathbf{X}, \mathbf{Y}$ .
- For attribute parameters,
  - simulate  $\gamma, \mathbf{b}$  from their full conditional distributions.
  - simulate  $\tau^2$  given  $\beta, \mathbf{a}, \sigma^2, \gamma, \mathbf{b}, \mathbf{Z}, \Theta, \Sigma, \mathbf{X}, \mathbf{Y}$ .
- For the joint latent variables and the covariance,
  - simulate  $\{\mathbf{Z} \text{ and } \Theta\}$  from their full conditional distributions.
  - simulate  $\Sigma$  from its full conditional distribution.

### 3.1 Full Conditionals of Connectivity Parameters

To derive the full conditional distributions of  $\beta_{Q \times 1}$  and  $\mathbf{a}_{N \times 1}$ , we look at the connectivity component of the joint model without the vector product. For each connectivity matrix for each individual  $X_i$ , we subtract  $\mathbf{Z}^T \mathbf{Z}$  and obtain  $\mathbf{R}_i = \mathbf{X}_i - \mathbf{Z}^T \mathbf{Z}$ . We extract the unique off-diagonal elements of the resulting symmetric matrix and form a  $V(V-1)/2 \times 1$  vector. We stack  $N$  such vectors and form a  $NV(V-1)/2 \times 1$  vector,  $\mathbf{r}, \mathbf{r}_{NV(V-1)/2 \times 1} = \mathbf{G}\mathbf{a} + \mathbf{W}\beta + \mathbf{e}$ , where  $\mathbf{G}$  is the  $NV(V-1)/2 \times N$  matrix,  $\mathbf{G} = (\mathbf{I}_N \otimes \mathbf{1}_{V(V-1)/2}^T)^T$ , with  $\mathbf{I}_N$  being the  $N \times N$  identity matrix and  $\mathbf{1}_{V(V-1)/2}$  being the  $V(V-1)/2 \times 1$  vector of 1s, and  $\mathbf{W}$  is the  $NV(V-1)/2 \times Q$  appropriate design matrix. We can further transform  $\mathbf{r}$  such that the transformed error term is a standard normal distribution using  $\tilde{\mathbf{r}} = \mathbf{c}\mathbf{r}$ ,  $\mathbf{c} = \sigma^{-1}$ . Therefore, the model can be written as  $\tilde{\mathbf{r}} = \mathbf{c}\mathbf{G}\mathbf{a} + \mathbf{c}\mathbf{W}\beta + \tilde{\mathbf{e}}$ , where elements in  $\tilde{\mathbf{e}}$  follow a standard normal distribution.

In matrix notations, we use  $\mathbf{Z}$  to denote the  $V \times K$  matrix of the latent variable values and  $\mathbf{E}$  to denote the  $V \times V$  matrix of errors. Recall that the approximation of the posterior distribution for the  $a_i$  is facilitated by setting a  $N(0,1)$  prior distribution. Assume  $\Sigma_0$  as the identity matrix of  $N$  dimensions, then the prior distribution of  $\mathbf{a}_{N \times 1}$  follows a multivariate normal distribution  $N(\mathbf{0}, \Sigma_0)$  distribution. The first step of the iteration is implemented by first simulating  $\beta$  conditional on  $\{\tilde{\mathbf{r}}, \mathbf{W}\}$  and then simulating  $\mathbf{a}$  conditional on  $\{\tilde{\mathbf{r}}, \mathbf{W}, \beta\}$ . The latter distribution can be derived by letting  $\mathbf{m} = \tilde{\mathbf{r}} - \mathbf{c}\mathbf{W}\beta$ , and thus it can be shown that

$$\begin{aligned}
p(\mathbf{a} | \mathbf{m}) &\propto p(\mathbf{m} | \mathbf{a})p(\mathbf{a}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{m} - c\mathbf{G}\mathbf{a})^T(\mathbf{m} - c\mathbf{G}\mathbf{a})\right) \exp\left(-\frac{1}{2}\mathbf{a}^T \Sigma_0^{-1} \mathbf{a}\right) \\
&\propto \exp\left(-\frac{1}{2}\mathbf{a}^T (c^2\mathbf{G}^T\mathbf{G} + \Sigma_0^{-1})\mathbf{a} + \mathbf{a}^T(c\mathbf{G}^T\mathbf{m})\right).
\end{aligned} \tag{3.1}$$

This is the kernel of a multivariate normal distribution with variance  $\text{Var}[\mathbf{a} | \mathbf{m}] = (c^2\mathbf{G}^T\mathbf{G} + \Sigma_0^{-1})^{-1}$  and mean  $\mathbb{E}[\mathbf{a} | \mathbf{m}] = c(c^2\mathbf{G}^T\mathbf{G} + \Sigma_0^{-1})^{-1}\mathbf{G}^T\mathbf{m}$ . The distribution of  $\beta$  conditional on  $\{\tilde{\mathbf{r}}, \mathbf{W}\}$  is the product of  $p(\beta)$  and  $p(\tilde{\mathbf{r}} | \beta, \mathbf{W})$ , and it can be shown that

$$\begin{aligned}
p(\tilde{\mathbf{r}} | \beta, \mathbf{W}) &\propto \int p(\tilde{\mathbf{r}} | \beta, \mathbf{W}, \mathbf{a})p(\mathbf{a})d\mathbf{a} \\
&\propto \int \exp\left(-\frac{1}{2}(\mathbf{m} - c\mathbf{G}\mathbf{a})^T(\mathbf{m} - c\mathbf{G}\mathbf{a})\right) \exp\left(-\frac{1}{2}\mathbf{a}^T \Sigma_0^{-1} \mathbf{a}\right) d\mathbf{a} \\
&\propto \exp\left(-\frac{1}{2}\mathbb{E}[\mathbf{a} | \mathbf{m}]^T \text{Var}[\mathbf{a} | \mathbf{m}]^{-1} \mathbb{E}[\mathbf{a} | \mathbf{m}]\right) \\
&\propto \exp\left(-\frac{1}{2}\beta^T(c^2\mathbf{W}^T\mathbf{S}\mathbf{W})\beta + \beta^T(c^2\mathbf{W}^T\mathbf{S}\mathbf{r})\right),
\end{aligned} \tag{3.2}$$

where  $\mathbf{S} = \mathbf{G}(c^2\mathbf{G}^T\mathbf{G} + \Sigma_0^{-1})^{-1}\mathbf{G}^T$ . The prior distribution of  $\beta$  is proportional to  $\exp\left(-\frac{1}{2}\beta^T \mathbf{S}_0\beta + \beta^T \mathbf{S}_0\beta_0\right)$ , i.e., the prior distribution of  $\beta$  follows a multivariate normal distribution  $\mathcal{N}(\beta_0, \mathbf{S}_0^{-1})$  distribution. The distribution of  $\beta$  conditional on  $\{\tilde{\mathbf{r}}, \mathbf{W}\}$  thus follows a multivariate normal distribution with variance  $(c^2\mathbf{W}^T\mathbf{S}\mathbf{W} + \mathbf{S}_0)^{-1}$  and mean  $(c^2\mathbf{W}^T\mathbf{S}\mathbf{W} + \mathbf{S}_0)^{-1}(c^2\mathbf{W}^T\mathbf{S}\mathbf{r} + \mathbf{S}_0\beta_0)$ . The posterior for  $\sigma^{-2}$  is  $\text{Gamma}\left(\frac{N * V * (V - 1)/2 + 1}{2}, 1/2 * (1 + \sum_{r=1}^N \sum_{u=1}^V \sum_{u < v} e_{u,v,i}^2)\right)$ .

### 3.2 Full Conditionals of Attribute Parameters

The full conditional distributions of the attribute parameters can be similarly obtained as those of the connectivity parameters. In particular, we stack  $N$  vectorized attribute matrices and form a  $NVP \times 1$  vector,  $\mathbf{r}', \mathbf{r}'_{NVP \times 1} = \mathbf{G}'\mathbf{b} + \mathbf{H}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ , where  $\mathbf{G}'$  is the  $NVP \times N$  matrix,  $\mathbf{G}' = (\mathbf{I}_N \otimes \mathbf{1}_{VP}^T)^T$ , with  $\mathbf{I}_N$  being the  $N \times N$  identity matrix and  $\mathbf{1}_{VP}$  being the  $VP \times 1$  vector of 1s, and  $\mathbf{H}$  is the  $NVP \times Q'$  appropriate design matrix. We further transform  $\mathbf{r}'$  such that the transformed error term is a standard normal distribution using  $\tilde{\mathbf{r}}' = d\mathbf{r}'$ ,  $d = \tau^{-1}$ . Therefore, the model can be written as  $\tilde{\mathbf{r}}' = d\mathbf{G}'\mathbf{b} + d\mathbf{H}\boldsymbol{\gamma} + \tilde{\boldsymbol{\epsilon}}$ , where elements in  $\tilde{\boldsymbol{\epsilon}}$  follow a standard normal distribution.

Let  $\mathbf{m}' = \tilde{\mathbf{r}}' - d\mathbf{H}\boldsymbol{\gamma}$ . Following similar derivation steps as before, we can show that the distribution of  $\mathbf{b}$  conditional on  $\{\tilde{\mathbf{r}}', \mathbf{H}, \boldsymbol{\gamma}\}$  follows a multivariate normal distribution with variance  $\text{Var}[\mathbf{b} | \mathbf{m}'] = (d^2\mathbf{G}'^T\mathbf{G}' + \Sigma_0^{-1})^{-1}$  and mean  $\mathbb{E}[\mathbf{b} | \mathbf{m}'] = d(d^2\mathbf{G}'^T\mathbf{G}' + \Sigma_0^{-1})^{-1}\mathbf{G}'^T\mathbf{m}'$ . It can be further concluded that the distribution of  $\boldsymbol{\gamma}$  conditional on  $\{\tilde{\mathbf{r}}', \mathbf{H}\}$  also

follows a multivariate normal distribution with variance  $(d^2 \mathbf{H}^T \mathbf{S}' \mathbf{H} + \mathbf{S}'_0)^{-1}$  and mean  $(d^2 \mathbf{H}^T \mathbf{S}' \mathbf{H} + \mathbf{S}'_0)^{-1} (d^2 \mathbf{H}^T \mathbf{S}' \mathbf{r}' + \mathbf{S}'_0 \mathbf{y}_0)$ , where  $\mathbf{S}' = \mathbf{G}' (d^2 \mathbf{G}'^T \mathbf{G}' + \Sigma_0^{-1})^{-1} \mathbf{G}'^T$ . The posterior for  $\tau^{-2}$  is Gamma  $\left( \frac{N * V * P + 1}{2}, 1/2 * (1 + \sum_{i=1}^N \sum_{u=1}^V \sum_{p=1}^P \epsilon_{u,p,i}^2) \right)$ .

### 3.3 Full Conditionals of the Joint Latent Variables and the Covariance

The brain region attributes are related to the brain connectivity through the dependence between  $\Theta$  and  $\mathbf{Z}$ . We are interested in the joint full conditional distribution of  $\Theta$  and  $\mathbf{Z}$ . Let us take a look at the probability model for the  $u$ th rows of  $\mathbf{Z}$  and  $\Theta - \mathbf{z}_u$  and  $\theta_u$ . We first reparameterize the covariance matrix  $\Sigma$  as the block matrix:  $\Sigma = \begin{pmatrix} \Lambda_z & \Lambda_{zy}^T \\ \Lambda_{zy} & \Lambda_y \end{pmatrix}$ , and the inverse of this block matrix is

$$\Sigma^{-1} = \begin{pmatrix} \Lambda_z & \Lambda_{zy}^T \\ \Lambda_{zy} & \Lambda_y \end{pmatrix}^{-1} = \begin{pmatrix} (\Lambda_z - \Lambda_{zy}^T \Lambda_y^{-1} \Lambda_{zy})^{-1} & -(\Lambda_z - \Lambda_{zy}^T \Lambda_y^{-1} \Lambda_{zy})^{-1} \Lambda_{zy}^T \Lambda_y^{-1} \\ -(\Lambda_y - \Lambda_{zy} \Lambda_z^{-1} \Lambda_{zy}^T)^{-1} \Lambda_{zy} \Lambda_z^{-1} & (\Lambda_y - \Lambda_{zy} \Lambda_z^{-1} \Lambda_{zy}^T)^{-1} \end{pmatrix}. \quad (3.3)$$

For the ease of notations, we further define the inverse of the covariance matrix as another block matrix:  $\Sigma^{-1} = \begin{pmatrix} Q_z & Q_{yz} \\ Q_{zy} & Q_y \end{pmatrix}$  with each component as a function of  $\Lambda_s$ .

For individual  $i$ , the connectivity component of the joint model with only the vector products and the error term is:

$$F_i = X_i - a_i - \mathbf{w}_i \beta = \mathbf{Z} \mathbf{Z}^T + E_i$$

We can transform  $F_i$  in such a way that the transformed error term is a standard normal distribution using  $\tilde{F}_i = c F_i$ , where  $c = \sigma_e^{-1}$ . Therefore, the model for  $\tilde{F}_i$  can be written as  $\tilde{F}_i = c \mathbf{Z} \mathbf{Z}^T + \tilde{E}_i$ , where  $\tilde{e}_{u,v,i}$  follows a standard normal distribution. Consider the  $u$ th row of the matrix  $\tilde{F}_i$ :

$$\begin{aligned} \tilde{f}_{u,1,i} &= c \mathbf{z}_u^T \mathbf{z}_1 + \tilde{e}_{p,1,i}, \\ \tilde{f}_{u,2,i} &= c \mathbf{z}_u^T \mathbf{z}_2 + \tilde{e}_{p,2,i}, \\ &\dots \\ \tilde{f}_{u,V,i} &= c \mathbf{z}_u^T \mathbf{z}_V + \tilde{e}_{p,V,i}, \end{aligned} \quad (3.5)$$

where  $\mathbf{z}_u$  is the  $K \times 1$  vector of coordinates on the latent network dimensions for node  $u$ . In this way, the connectivity component is written in the more familiar simple regression form.

Similarly, we can rewrite the attribute component into another regression form. First, we consider the matrix form of the attribute component without the individual-specific intercepts and the covariate effects:

$$T_i = Y_i - b_i - h_i^T \gamma = \Theta + \Psi_i, \quad (3.6)$$

where, as mentioned before,  $\Theta$  is the  $V \times P$  matrix of latent variables; and  $\Psi_i$  is the  $V \times P$  matrix of random error. Consider the  $u$ th row of  $T_i$ :

$$t_{u,i} = \theta_u + \epsilon_{u,i}, \quad (3.7)$$

where  $\theta_u$  is the  $P \times 1$  vector of latent variable values for node  $u$ , and  $t_{u,i}$  is the  $P \times 1$  vector of attribute values for node  $u$  and individual  $i$ .

The joint full conditional distribution of  $z_u$  and  $\theta_u$  is:

$$\begin{aligned} p\left(\begin{pmatrix} z_u \\ \theta_u \end{pmatrix} \middle| t_{u,i}, \tilde{f}_{u,i}, \Sigma, \tau^2\right) &\propto \prod_{i=1}^N p(t_{u,i} | \theta_u, \tau^2) \prod_{i=1}^N p(\tilde{f}_{u,i} | z_u) p\left(\begin{pmatrix} z_u \\ \theta_u \end{pmatrix} \middle| \Sigma\right) \\ &\propto \exp\left(-\frac{1}{2}\tau^{-2} \sum_{i=1}^N (t_{u,i} - \theta_u)^T (t_{u,i} - \theta_u)\right) \exp\left(-\frac{1}{2} \sum_{i=1}^N \sum_{v=1, v \neq u}^V (\tilde{f}_{u,v,i} - c z_u^T z_v)^2\right) \\ &\quad \exp\left(-\frac{1}{2} \begin{pmatrix} z_u \\ \theta_u \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} z_u \\ \theta_u \end{pmatrix}\right). \end{aligned}$$

Using the previously defined block matrix, we can write  $p\left(\begin{pmatrix} z_u \\ \theta_u \end{pmatrix} \middle| \Sigma\right)$

as  $\exp\left(-\frac{1}{2}(\mathbf{z}_u^T \mathbf{Q}_z \mathbf{z}_u + \mathbf{z}_u^T \mathbf{Q}_{yz} \theta_u + \theta_u^T \mathbf{Q}_{zy} \mathbf{z}_u + \theta_u^T \mathbf{Q}_y \theta_u)\right)$ . Extracting relevant terms from  $p\left(\begin{pmatrix} z_u \\ \theta_u \end{pmatrix} \middle| t_{u,i}, \tilde{f}_{u,i}, \Sigma, \tau^2\right)$ , we can see that the full conditional distribution of  $z_u$  is

$$p(z_u | \tilde{f}_{u,i}, \Sigma, \theta_u) \propto \exp\left(-\frac{1}{2} \mathbf{z}_u^T \left( \sum_{v=1, v \neq u}^V N c^2 \mathbf{z}_v \mathbf{z}_v^T + \mathbf{Q}_z \right) \mathbf{z}_u + \mathbf{z}_u^T \left( \sum_{i=1}^N \sum_{v=1, v \neq u}^V c \tilde{f}_{u,v,i} \mathbf{z}_v - \frac{1}{2} \mathbf{Q}_{yz} \theta_u - \frac{1}{2} \mathbf{Q}_{zy}^T \theta_u \right)\right). \quad (3.8)$$

This is a multivariate normal distribution, with variance  $(\sum_{v=1, v \neq u}^V N c^2 \mathbf{z}_v \mathbf{z}_v^T + \mathbf{Q}_z)^{-1}$  and mean  $(\sum_{v=1, v \neq u}^V N c^2 \mathbf{z}_v \mathbf{z}_v^T + \mathbf{Q}_z)^{-1} \left( \sum_{i=1}^N \sum_{v=1, v \neq u}^V c \tilde{f}_{u,v,i} \mathbf{z}_v - \frac{1}{2} \mathbf{Q}_{yz} \theta_u - \frac{1}{2} \mathbf{Q}_{zy}^T \theta_u \right)$ . Similarly, the full conditional of  $\theta_u$  is

$$p(\theta_u | t_{u,i}, \Sigma, z_u, \tau^2) \propto \exp\left(-\frac{1}{2} \theta_u^T (N \tau^{-2} \mathbf{I} + \mathbf{Q}_y) \theta_u + \theta_u^T \left( \sum_{i=1}^N \tau^{-2} t_{u,i} - \frac{1}{2} \mathbf{Q}_{yz}^T z_u - \frac{1}{2} \mathbf{Q}_{zy} z_u \right)\right). \quad (3.9)$$

This is a multivariate normal distribution, with variance  $(N \tau^{-2} \mathbf{I} + \mathbf{Q}_y)^{-1}$  and mean  $(N \tau^{-2} \mathbf{I} + \mathbf{Q}_y)^{-1} \left( \sum_{i=1}^N \tau^{-2} t_{u,i} - \frac{1}{2} \mathbf{Q}_{yz}^T z_u - \frac{1}{2} \mathbf{Q}_{zy} z_u \right)$ .

**Algorithm 1:**

ABC Estimation procedure

1:	<b>Initialize</b> $\sigma^2, \beta, \mathbf{a}, \tau^2, \gamma, \mathbf{b}, \mathbf{Z}, \Theta$ and $\Sigma$ ;
2:	<b>while</b> $t < N_{\text{iter}}$ <b>do</b>
3:	Sample $\mathbf{a}$ from a multivariate normal distribution with variance $(c^2 \mathbf{G}^T \mathbf{G} + \Sigma_0^{-1})^{-1}$ and mean $c(c^2 \mathbf{G}^T \mathbf{G} + \Sigma_0^{-1})^{-1} \mathbf{G}^T \mathbf{m}$ ;
4:	Sample $\beta$ from a multivariate normal distribution with variance $(c^2 \mathbf{W}^T \mathbf{S} \mathbf{W} + \mathbf{S}_0)^{-1}$ and mean $(c^2 \mathbf{W}^T \mathbf{S} \mathbf{W} + \mathbf{S}_0)^{-1} (c^2 \mathbf{W}^T \mathbf{S} \mathbf{r} + \mathbf{S}_0 \beta_0)$ ;
5:	Sample $\sigma^{-2}$ from a gamma distribution $(\frac{N * V * (V - 1) / 2 + 1}{2}, 1/2 * (1 + \sum_{i=1}^N \sum_{u=1}^V \sum_{v < u}^V e_{u,v,i}^2))$ ;
6:	Sample $\mathbf{b}$ from a multivariate normal distribution with variance $(d^2 \mathbf{G}'^T \mathbf{G}' + \Sigma_0^{-1})^{-1}$ and mean $d(d^2 \mathbf{G}'^T \mathbf{G}' + \Sigma_0^{-1})^{-1} \mathbf{G}'^T \mathbf{m}'$ ;
7:	Sample $\gamma$ from a multivariate normal distribution with variance $(d^2 \mathbf{H}^T \mathbf{S}' \mathbf{H} + \mathbf{S}'_0)^{-1}$ and mean $(d^2 \mathbf{H}^T \mathbf{S}' \mathbf{H} + \mathbf{S}'_0)^{-1} (d^2 \mathbf{H}^T \mathbf{S}' \mathbf{r}' + \mathbf{S}'_0 \gamma_0)$ ;
8:	Sample $\tau^{-2}$ from a gamma distribution $(\frac{N * V * P + 1}{2}, 1/2 * (1 + \sum_{i=1}^N \sum_{u=1}^V \sum_{p=1}^P e_{u,p,i}^2))$ ;
9:	Sample $\theta_u$ from a multivariate normal distribution, with variance $(\sum_{v=1, v \neq u}^V N c^2 \mathbf{z}_v \mathbf{z}_v^T + \mathbf{Q}_z)^{-1}$ and mean $(\sum_{v=1, v \neq u}^V N c^2 \mathbf{z}_v \mathbf{z}_v^T + \mathbf{Q}_z)^{-1} (\sum_{i=1}^N \sum_{v=1, v \neq u}^V c \tilde{f}_{u,v,i} \mathbf{z}_v - \frac{1}{2} \mathbf{Q}_{yz} \theta_u - \frac{1}{2} \mathbf{Q}_{zy}^T \theta_u)$ ;
10:	Sample $\mathbf{z}_u$ from a multivariate normal distribution with variance $(N \tau^{-2} \mathbf{I} + \mathbf{Q}_y)^{-1}$ and mean $(N \tau^{-2} \mathbf{I} + \mathbf{Q}_y)^{-1} (\sum_{i=1}^N \tau^{-2} t_{u,i} - \frac{1}{2} \mathbf{Q}_{yz}^T \mathbf{z}_u - \frac{1}{2} \mathbf{Q}_{zy} \mathbf{z}_u)$ ;
11:	Sample $\Sigma$ from an inverse-Wishart $(\mathbf{I}_{K+P} + \mathbf{F}'^T \mathbf{F}', V + K + P + 2)$ ;
12:	$t \leftarrow t + 1$ ;
13:	<b>end while</b>
14:	<b>return</b> posterior samples of $\sigma^2, \beta, \mathbf{a}, \tau^2, \gamma, \mathbf{b}, \mathbf{Z}, \Theta$ and $\Sigma$ as approximated target distributions;

Recall that prior for  $\Sigma$  is  $\Sigma^{-1} \sim \text{Wishart}(\mathbf{I}_{K+P}, K + P + 2)$ . Let  $\mathbf{F}'$  be a  $V \times (K + P)$  matrix with  $u$ th row as  $(\mathbf{z}_u^T, \theta_u^T)$ . The full conditional for  $\Sigma$  follows an inverse-Wishart  $(\mathbf{I}_{K+P} + \mathbf{F}'^T \mathbf{F}', V + K + P + 2)$ .

Following equation 2.2, it can be proved that  $\mathbf{Z}$  is identified up to rotations while  $\mathbf{Z}^T \mathbf{Z}$  is directly identified following the condition that  $\mathbf{Z}$  is centered. To resolve the rotation indeterminacy issue, we perform a procrustean transformation of the estimated  $\mathbf{Z}$  at each iteration to its initial position,  $\mathbf{Z}_0$ . For each  $\mathbf{Z}$  sampled from the posterior, we find the rotated  $\mathbf{Z}^*$  of  $\mathbf{Z}$  such that  $\mathbf{Z}^*$  has the smallest sum of squared deviations from the target orientation  $\mathbf{Z}_0$ . We freely estimate the covariance matrix  $\Sigma$ .

## 4 Simulation

We conduct simulation studies to evaluate whether the proposed ABC model compared with the existing alternatives will facilitate a better prediction of brain connectivity from new subjects, and to assess how incorporating node-level attributes improves the performance. We compare the ABC model with two existing approaches, namely the commonly used averaging method (Average), where the group-level connectivity is estimated as the entry-wise sample mean of individual connectivity matrices, and the multiplex stochastic block model (MSBM, Barbillon et al. (2017)), a generative network model for estimating the group-level connectivity with discrete latent variables. To assess the improvement on estimation by incorporating the regional anatomical attributes, we compare the ABC model against the same model without incorporating node-wise attributes, denoted as the brain connectivity (BC) model.

The data are generated as follows. For simplicity and consistency, the number of attributes and the number of latent variables are both assigned as two in all generated data. We first generate the connectivity latent variables as well as the latent variables of the attributes from the multivariate normal distribution with the mean zero and the pre-specified correlation matrix (the variances are assumed to be 1). Following existing literature on the joint models with multiple sources of information (e.g., Gollini & Murphy, 2016; Wang et al., 2023), we believe that the benefit of adding attributes of the brain regions depends on the relationship between connections and attributes. Thus, we specify the correlation matrix between connectivity and attributes based on three levels of dependence: (a) they are independent, (b) they have moderate correlations of 0.5 and (c) they have strong correlations of 0.9. To create symmetric positive definite matrices, we assign correlation values to the corresponding dimensions between the connectivity and the attributes of the brain regions, e.g., for the strong dependence condition, we assign the 0.9 correlation between the first dimensions of connectivity and attributes and between the second dimensions of the connectivity and attributes. With these three levels of dependence, we randomly generate latent variable values for both the connectivity and the attributes.

We consider two sample sizes,  $N = 50$  and  $N = 100$  and two conditions for the number of nodes  $V = 20$  and  $V = 70$ , and we specify three levels of the signal-to-noise (S/N) ratio, 0.05, 0.02 and 0.01 controlled by modifying the error variance while keeping the variance of the latent variables constant. To compare ABC with BC, we add three additional levels of the signal-to-noise ratio, 0.1, 1 and 2, and we also add a smaller sample size  $N = 20$ . These additional conditions allow us to assess the predictive power of ABC with a fuller range of parameters. We randomly sample the errors for the connectivity from the normal distribution with the mean 0 and the variance defined by the desired S/N ratio. The errors for the attributes are sampled from the normal distribution with the mean 0 and variance 0.5. The individual-specific intercepts are set as 1 or  $-1$  with a sum of zero across subjects. In total, we will consider 24 different scenarios combining from different connectivity and attribute correlations, sample sizes, node numbers and S/Ns. Under each scenario, we simulate the data 100 times.

To evaluate prediction, each simulated dataset is randomly divided as an equally sized training set and test set. We implement the ABC model on the training set based on the developed MCMC algorithm with the convergence of the posterior computation confirmed by trace plots of the key parameters. As comparisons, we also apply the competing methods Average, MSBM and BC on the same training set. Specifically, for the Average method, we average over the individual connectivity in the training set and use the averaged connectivity as the estimated group connectivity; for MSBM, we fit the model using the `sbm` package (Barbillon et al., 2017) in R; and for BC, we implement it following the same process as ABC model except that the attribute component is excluded. We estimate the Mean Squared Error (MSE) of the predicted connectivity on the test set by calculating the average squared difference between the upper diagonal elements of the predicted values and corresponding observed values. We also calculate the correlation between the upper diagonal elements of the predicted values and corresponding observed values. We use the estimated model parameters and unknown quantities based on the training data to predict observations for the test data. The predicted observations are based on estimated model parameters from the training data. Conceptually, this process is similar to obtaining predictions from a simple linear regression model, where the regression coefficients are estimated based on the training data to predict observations in the test data. To predict a new subject's connectivity data, we use the group-level connectivity obtained via the estimated connectivity latent variables from the training data—the estimation process is detailed in the MCMC section via iterative samplings of full conditional distributions—and the other model parameters such as individual-level intercepts. With these estimated model parameters and unknown quantities, we obtain prediction for a new subject following the model equations.

Table 1 shows the MSEs and the correlations of the proposed ABC method against those of the Average method and the MSBM. The standard deviations of the squared differences between the predicted and the observed connectivity across replications are reported along with the standard deviations of the correlations. As an illustrate, we report the results from the independence and strong dependence conditions here, and the results for the moderate dependence condition are shown in Figure 2.

Based on Table 1, we can see that the group-level connectivity is more precisely and accurately recovered using ABC than the Average method and MSBM. The improvement becomes more evident when the estimation of the group-level connectivity becomes more difficult. As the signal-to-noise ratio decreases, the MSE of the Average method substantially increases, which can be attributed to the fact that the Average method does not and cannot separate signals from noise. As the noise becomes stronger, the Average method performs worse. This is true for both independent and dependent conditions. The correlations between the predicted and the observed connectivity show a similar trend. Between ABC and MSBM, ABC consistently outperforms MSBM in predicting the group-level connectivity. There are two potential reasons for this. First, the stochastic block model is inferior for capturing the topological properties of brain structural connectivity as generated by our settings. The block model is often used to capture potential clusters in the network, effective for networks with high levels of stochastic equivalence and less effective for networks with high levels of homophily (Hoff, 2007). Second, the ABC model outperforms because it incorporates the additional attributes information. Between

the independence and strong dependence conditions, the predictive power of ABC improves while the predictive powers of MSBM stay the same. This provides evidence for the benefit of incorporating the attributes information.

In Figure 2, we present the difference in the predictive power between ABC and BC when (a)  $V = 20$  and (b)  $V = 70$ . Positive values indicate that ABC outperforms BC with higher correlations between the predicted and the observed brain connectivity. The comparison between ABC and BC illustrates the improvement on connectivity prediction by incorporating node-wise attributes. Figure 2 shows that adding the attributes information is beneficial for estimating the group-level connectivity when the attributes are related to the connectivity with both moderate (red) and strong (green) dependence. The stronger the dependence, the more beneficial it is to incorporate attributes. When they are independent (blue), there is no observable improvement. When there is dependence, the level of improvement depends on how difficult it is to estimate the group-level connectivity or how much information there is to estimate the group-level connectivity. There is more benefit for incorporating the attributes when the signal-to-noise ratio is small and when the sample size is small. The results for when  $V = 20$  show a similar trend as the results for  $V = 70$ .

In Figure 3, we present the correlations between the predicted and observed connectivity in the test data across 100 random samples when the true dimensionality is 2 versus when the true dimensionality is 5. We focus on the data condition with small ( $N = 50$ ) and large sample size ( $N = 100$ ) when the number of brain region is 100, and the signal to noise ratio is 0.05. We also include a comparison between when the connectivity and attributes are independent versus dependent. The results show that the model shows similarly satisfactory recovery of connectivity for both low and high dimensionality. When connectivity and attributes are dependent, the high dimensionality shows slightly better prediction accuracy, which could be contributed to the fact that more connectivity dimensions are related to attributes in the high dimensionality condition than the low dimensionality condition.

## 5 Application

We apply the proposed ABC model to the ADNI data described in section 2.1. Our goal is to identify sex-specific structural connectivity differences between the healthy and AD population. To plausibly constructing the group-level connectivity and performance inference, we propose the ABC model to integrate brain node attributes information on the cortical volume, surface areas, average cortical thickness and standard deviation of the cortical thickness. Each of the four groups has 30, 49, 45, and 50 subjects, respectively. Alternatively, we can estimate the group-level effects of sex and disease status by incorporating them as covariates of the joint model. However, in such a model, no edge-level connectivity differences can be estimated. Such models limit the interpretability of the results by only allowing global connectivity differences between groups, e.g., healthy subjects overall tend to have higher brain connectivity than subjects with AD. Such conclusions, although statistically meaningful, make little contribution to our current understanding about the characteristics of neurodegenerative diseases such as AD. In this section, we first present the results of the estimated group-level connectivity. Second, we compare the estimated connectivity across different groups and identify their

differences. Lastly, we characterize and perform inference on the correlation between structural connections and regional anatomical information under different subgroups to understand their interplay.

## 5.1 Estimation and Model Fit

We perform posterior inference based on the MCMC algorithm described in section 3. We generate 20,000 iterations with a burn-in period of 500 and thinning every 10th sample. Each model takes about 1 hour to complete with 193.67 MB memory and one computing node with three cores on the Yale HPC cluster. The trace plots show no obvious signs of non-convergence; see examples in the supplementary material.

The samples in each group are randomly divided into 3 sets: training (90%), validation (5%) and test set (5%). To select the number of connectivity dimensions, each ABC model is trained on the training set with  $K = 2, 3, 4, 5, 6, 7, 8$ . To compare the fit of the model with varying numbers of dimensions, we assess the recovery of the observed connectivity based on the validation set. The number of dimensions with the best recovery of the observed connectivity in the validation set is selected. Lastly, after the dimensions are selected, the model fit index is calculated based on the test set. In Table 2, we present the predictive power, estimated as the correlation between the predicted and the observed connectivity, of the ABC model with varying dimensions based on the validation set. The results show that the predictive power increases as the number of dimensions increases from 2 to 5, and then decreases as  $K$  continues to increase. The highest predictive power is observed with  $K = 5$  suggesting that there are likely 5 connectivity dimensions. This result applies to all four subject groups. To assess the fit of the model to data, we calculate the predictive power of the ABC model with  $K = 5$  based on the test data. As shown in Table 2, these results show satisfactory fit of the ABC model to data.

## 5.2 Attributes-informed Group-level Connectivity

Figure 4 shows five estimated latent dimensions of the structural connectivity from (a) the lateral view of the right hemisphere, (b) the superior view of the brain, and (c) the lateral view of the left hemisphere. These latent dimensions are estimated based on the healthy female subjects, similar to those based on the healthy male subjects (see the supplementary materials for their similarities). The brain regions are colored based on the values of the latent dimensions with red indicating positive values and green indicating negative values. Each row of Figure 4 shows the estimated values of the corresponding latent dimension. For example, Figure 4a,1 shows the estimated values of the first latent dimension from the lateral view of the right hemisphere. Figure 5 shows the latent space of the structural connectivity between (a) dimensions 1 and 2, (b) dimensions 3 and 5 and (c) dimensions 1 and 4. See the latent spaces between each pair of dimensions in the supplementary materials. Each dot of the latent space represents a brain region. The color differentiates the left hemisphere (black) from the right hemisphere (red). A vector drawn from the origin (0, 0) to the dot is the vector presentation of the brain region. We use capital letters to refer to directions of the latent space represented by the region vectors. Following the model equation, the structural connectivity between two brain regions is represented by the vector product of the two regions' vectors. Strong connectivity is represented by the large vector product, and

vice versa. Thus, both the length of the vector and the angle between two vectors signal connectivity strength with a small angle indicating a strong connection. For a given brain region, the same vector length is used in the presentation of its connection with all other regions. Therefore, the vector length indicates an overall connectivity strength for the region.

In Figure 5, a grouping of the brain regions can be observed when these regions share a similar direction, distinct from directions of the other regions. For example, between dimensions 1 and 2 in Figure 5a, three groups of brain regions are found in the directions of A,B and C. In direction A, we find most of the regions in the left hemisphere; and in direction B, we find most of the regions in the right hemisphere. This differentiation of the left and right hemispheres can also be seen in rows 1 and 2 in Figure 4. In these two rows, edges in the same hemisphere are estimated with latent variable values of the same sign, shown with the same color; and edges in different hemispheres are estimated with latent variable values of opposite signs, shown with opposite colors. Therefore, the first two latent dimensions differentiate the left from the right hemisphere. This differentiation suggests that regions in the same hemisphere are more connected with higher fiber density than regions from different hemispheres. This could be attributed to the association fibers interconnecting areas of the cerebral cortex in the same hemisphere. These association fibers may be short, connecting adjacent regions, or long, connecting more distant areas of the cortex. In direction C, we mostly find regions of the cingulate cortex. As we know, the interconnections between hemispheres are based on the commissural fibers. Their bands link the hemispheres and include the corpus callosum, which is wrapped by the cingulate gyrus like a “belt”. Our results reflect the distinct nature of the fiber connectivity around regions of the cingulate. This distinction is captured by the latent space that would have been lost otherwise.

In Figure 5b, between dimensions 3 and 5 of the latent space, four groupings of regions are found in the directions of D,E,F and G. In directions D and G, regions in the posterior parietal cortex, the temporal and occipital lobes are found from the right (D) and left (G) hemispheres. In directions E and F, regions in the frontal cortex and the primary somatosensory cortex are found from the right (F) and left (E) hemispheres. The differentiation of the posterior parietal cortex, the temporal and occipital lobes (directions D and G) from the frontal cortex and the primary somatosensory cortex (directions F and E) can also be observed in rows 3 and 5 of Figure 4. We anticipate that directions D and G reflect the long association fibers including the inferior longitudinal fasciculus connecting the temporal and the occipital lobes and the superior longitudinal fasciculus connecting the parietal, occipital and temporal lobes with the frontal lobe (Felten et al., 2015). Meanwhile, directions E and F reflect the connectivity of the frontal cortex and the primary somatosensory cortex. The short association fibers interconnect the prefrontal cortex, the premotor area and the motor cortex with the primary somatosensory cortex (Arle & Shils, 2011).

In Figure 5c, dimension 4 of the latent space differentiates the prefrontal cortex regions including lateral orbitofrontal, medial orbitofrontal and frontal pole from the parietal lobe including postcentral, superior parietal and supramarginal. This differentiation of the prefrontal cortex from the parietal lobe can also be seen in row 4 of Figure 4. In directions

of H and I, the posterior cingulate regions are found with the parietal lobe regions. The posterior cingulate cortex connects regions of parietal lobe that receive inputs from the dorsal visual stream and somatosensory areas (Rolls & Wirth, 2018; Vogt & Laureys, 2009), and posterior cingulate cortex is concerned with spatial representation, spatial processing, orientation and certain types of memory (Baleydier & Mauguiere, 1980; Beckmann et al., 2009).

### 5.3 Group-level Connectivity Differences

In this section, we identify gender-specific AD biomarkers by comparing the group-level connectivity between different subject populations. Figure 6 shows the estimated attributes-informed group-level structural connectivity for (a) healthy male subjects, (b) male subjects with AD and (c) female subjects with AD. The estimated connectivity for healthy female subjects can be found in the supplementary material. Compared with the healthy male subjects, AD male subjects show increased connectivity for certain brain regions in the occipital lobe, highlighted in Figure 6b. Meanwhile, a different set of regions is found with increased connectivity for AD female subjects in certain regions of the parietal lobe, highlighted in Figure 6c. This result suggests that there is sex-based difference in the neuromarkers of AD.

To assess whether these sex differences associated with AD are significant, we estimate the uncertainty of the estimated edge values based on the posterior sampling. We selectively investigate the edges that show large connectivity differences between different subject groups based on Figure 6. Each edge is investigated based on the posterior distributions, and those that show significant sex differences are shown in Figures 8 and 7. In Figures 8 and 7, the first column shows the locations of the brain regions involved in the edges from the lateral view of the left hemisphere. The second column shows the posterior distributions of the associated edges for AD female, AD male, healthy female and healthy male subjects.

Among all edges identified with significant sex differences, the associated brain regions are found in the left hemisphere. This could be attributed to the asymmetric thinning of the cerebral cortex between the left and right hemispheres exacerbated by AD (Roe et al., 2021). Roe et al. (2021) found that the two hemispheres deteriorate at different rates with the left hemisphere shrinking faster in patients with AD. It is thus not surprising that large structural connectivity differences are observed between subject groups in the left hemisphere.

Figure 7 shows sex-based difference in the neuromarkers of AD for the structural connectivity between (a) pericalcarine and precuneus (b) lingual and cuneus, (c) lingual and pericalcarine and (d) pericalcarine and cuneus. Among these four edges, AD male subjects have the highest fiber density connectivity, followed by healthy male subjects. Females tend to have lower fiber density connectivity than males. The difference between AD females and AD males are statistically significant given that their posterior distributions do not overlap (the corresponding credible intervals do not overlap). There does not seem to be a difference between healthy females and AD females. It is interesting to note that the regions associated with these edges are uniformly located in the medial aspect of the occipital lobe and the precuneus of the parietal lobe, which includes the primary visual cortex. AD patients often report symptoms of visual deficits, which are linked to damages to the

occipital lobe (Berkovitch et al., 2021; Crayton et al., 1977; Nielsen, 1955). The abnormality in the precuneus and the occipital lobe for the male AD patients could also be linked to psychosis (Carter et al., 2015; Rikandi et al., 2017; Sireteanu et al., 2008; Zmigrod et al., 2016), which is more often observed with male patients (Barajas et al., 2015; Chiu et al., 2018).

Figure 8 shows sex-based differences in the structural connectivity between regions of (a) superior parietal and supramarginal, (b) inferior parietal and supramarginal, and (c) inferior parietal and superior parietal. Among these three edges, healthy female subjects have the highest fiber density connectivity, followed by AD female, healthy male and AD male subjects. Between males and females, females have higher fiber density connectivity in these regions; and both sexes suffer a loss of connectivity with AD. The difference between AD females and AD males are statistically significant given that their posterior distributions do not overlap (the corresponding credible intervals do not overlap). It is interesting to note that the regions associated with these edges are found in the parietal lobe, more specifically, the posterior parietal lobe, a region implicated in spatial attention and eye movements (Groh & Werner-Reiss, 2002).

The parietal lobe is known to have structural differences between males and females (e.g., Frederikse et al., 1999; Kosciak et al., 2009; Salinas et al., 2012; Studholme et al., 2020). Females have a larger ratio of parietal lobe cortex to white matter than males in children between 7 and 17 years of age (Salinas et al., 2012). This sex-based structural difference in the parietal lobe can also be traced back to the early developmental stages, with females having greater relative proportion of the brain surface area occupied by the parietal lobe than males in both fetuses and neonates (Lehtola et al., 2019; Studholme et al., 2020). The parietal lobe is involved in the visuospatial processing, which is known to suffer damage with AD. Compared with the healthy elderly, AD patients tend to experience visuospatial deficits (Binetti et al., 1998; Cahn-Weiner et al., 1999; Cushman & Duffy, 2007; DeIpoli et al., 2007; Ricker et al., 1994), faring worse on visuospatial tasks such as the tests of mental rotation (Kurylo et al., 1996; Lineweaver et al., 2005; Mendola et al., 1995). It is thus not surprising that the parietal lobe is found to suffer extensive damage in AD patients, in line with existing research on AD such as Lineweaver et al. (2005).

The proposed ABC model provides consistent and interpretable evidence for sex-based differences in the neuromarkers of AD such as the role of the parietal lobe in AD. By contrast, existing research generally show a lack of success in finding consistent sex-based visuospatial differences for AD patients (Laws et al., 2016). Some show no significant sex differences in visuospatial deficits for patients with mild AD (Pernecky et al., 2007), and others show male superiority on some visuospatial tasks, but not others (Millet et al., 2009). Thus, it can be seen that the proposed ABC model is a comprehensive tool for identifying group-level differences based on sex and disease types.

In Table 3, we show results for generating group-level connectivity for AD females, AD males, healthy females and healthy males using the Average method; and we show the results for the case-control comparison between AD female and AD males for identified imaging biomarkers via the ABC model. In this way, we evaluate the ABC model by

comparing its performance in the group-level connectivity generating against the commonly used Average method. The results show smaller fiber density in AD females than AD males for pericalcarine—cuneus, lingual—pericalcarine and lingual—cuneus connectivity edges though only pericalcarine—cuneus,  $t(67.259) = 3.585, p < 0.001$  and lingual—pericalcarine,  $t(69.807) = 1.9949, p = 0.05$  connectivity edges show significant differences based on two-sample t tests; and only the pericalcarine—cuneus edge shows significant difference when controlling for multiple comparisons. The results show larger fiber density in AD females than AD males for superior parietal—supramarginal and inferior parietal—superior parietal connectivity edges though neither shows significant difference. It is important to note that univariate tests assuming independence of the connectivity edges are costly and often suffer low power (Wang, Liu, et al., 2023).

#### 5.4 Dependence between Connectivity and Attributes

Lastly, we discuss the dependence between the regional anatomical attributes and structural connections among them reported by the ABC model. Through ABC, we estimate the covariances (and correlations) between each of the attributes and the latent connectivity dimensions for each group of subjects. Among the four anatomical attributes, the cortical volume was consistently found to be significantly related to the structural connectivity among the brain regions. The 95% credible intervals for the correlation in absolute values between the cortical volume and the fourth dimension of the brain connectivity are [0.056, 0.484], [0.063, 0.484], [0.096, 0.509], and [0.036, 0.452] for AD female, AD male, healthy female and healthy male, respectively. The result shows that the cortical volume is significantly related to the white matter fiber density with a relatively stronger correlation for healthy females. The corresponding credible intervals of the other three attributes contain zero, and further research is needed to conclude their relationship with structural connectivity.

## 6 Discussion

The ABC model outlined in this article constitutes a principal strategy along with a valid inference process for estimating group-level brain connectivity while incorporating node attributes. Characterizing brain connectivity specific to a subpopulation in an analytically and biologically plausible way plays a crucial role to understand neurological architectures under different phenotypic conditions. Filling a literature gap, we have proposed a generative latent space network model as a comprehensive and extendable strategy for estimating attributes-informed brain connectivity. Based on the proposed modeling framework, we dissect interpretable latent space structures of group-level structural connectivity, informed by regional anatomical knowledge. Through a valid inference process, we quantify the uncertainty of each unknown connectivity and regional component and evaluate the likelihood of group-level difference against chance. Through extensive simulations and applications on AD, we have shown the superior performance of our model in understanding and comparing group-level brain structural networks and their interplay with anatomical structures, as well as predicting connectivity traits for new subjects.

Predicting structural connectivity is an important task with many studies having access to MRI for anatomical information but without or with missing DTI data. In such cases, being able to predict missing structural connectivity based on available anatomical information allows us to fully utilize accessible information to deal with a missing data issue and to improve prediction accuracy. The ABC model outperforms existing approaches in predictive tasks by incorporating anatomical knowledge as demonstrated by the simulation studies. In particular, the advantages of the ABC model become more evident as the estimation task becomes more difficult with decreasing signal to noise ratio and smaller sample size. The Average method is not able to differentiate signal from noise, and its predictive power deteriorates when more noise is found in the data. Meanwhile, the ABC method shows minimal loss in predictive power with increasing noise outperforming both the Average method and the MSBM. In practice, rarely do we observe negligible noise in the brain imaging data, thus making the use of the Average method problematic.

The proposed ABC model assumes additive Gaussian structure for the error term given the continuous fiber density structural connections. The Gaussian error term is one of the most common error assumptions for statistical model given its flexibility and simplicity. Data transformations often accompany continuous observations that deviate from the symmetric normal distribution, which makes the proposed model more flexible, but warrants caution for interpretation cite. Nonetheless, recent research have began to investigate potential multiplicative nature of the error term cite, where a larger fiber bundle exhibits a greater variance in term of number of streamlines than a small bundle when repeating measurements. The error variance of the diffusion tractography connectomes can be further decomposed into region or edge specific component and independent noise component, fully reflecting potential multiplicative nature of the errors. Given the complexity of the issue, a future study is warranted to investigate the noise associated with diffusion tractography connectomes.

We have demonstrated the practical benefits of the ABC model using the ADNI data. The estimated latent structural connectivity space offers a reduced-dimension and interpretable representation of the fiber density structural connectivity. For future direction, the ABC model can also be applied to the Adolescent Brain Cognitive Development (ABCD) study to investigate the biological basis of cognitive development through adolescence into young adulthood. The model allows us to estimate group-level connectivity to identify brain region changes following adolescents' cognitive developmental stages. The ABC model allows us to quantify the uncertainty of the potential group-level differences and test the possible dependence between cognition and anatomy of the brain.

Although our focus is on jointly modeling structural connectivity and anatomical attributes, converging literature suggests that anatomical attributes may also impact functional connectivity (Wilson et al., 2020). To investigate the relationship between anatomical attributes and functional connectivity, we can apply the ABC model to estimate the group-level functional brain connectivity while incorporating the anatomical attributes of the corresponding brain regions. We can identify and compare the estimated brain connectivity between different sexes or disease conditions and test the impact of the anatomical attributes on functional connectivity. In addition, we also intend to consider

regional functional attributes including measurements from PET imaging (e.g. AV45, FDG-PET) when modeling group-level functional connectivity. This will allow us to characterize how regional metrics from PET interact with functional connections measured by functional MRI, which has received growing attention and become an active research direction for AD study, especially with more multi-modal PET imaging data becoming available.

Future research will also be considered to extend our analytical framework to accommodate alternative data conditions. For example, the ABC model can be extended to model dynamic functional connectivity in order to query how the connectivity metric changes along the time course. The temporal dependence between the functional brain imaging data can be accounted for to facilitate the latent connectivity space to be time-dependent, and then the dynamics of connectivity across time windows can be explored. The ABC model can also be extended to incorporate behavior measurements by allowing multiple latent dimensions in the attributes model. In general mental health and behavior studies, behavior outcomes are typically collected from multiple spectra, and we would expect different data spectra to be highly correlated while offering distinct information. To integrate different attribute spectra, multi-dimensional latent space modeling needs to be developed to characterize their shared latent components. Finally, the ABC model is currently limited to continuous connection metrics, and it can be easily extended to binary metrics with Bernoulli distributions or count data with Poisson distributions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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.

## Funding

This study was supported in part by NIH grants RF1AG081413, P30AG021342 R01EB034720 and RF1AG068191.

## References

- Achard S, Salvador R, Whitcher B, Suckling J, & Bullmore E (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1), 63–72. [PubMed: 16399673]

- Aliverti E, & Durante D (2019). Spatial modeling of brain connectivity data via latent distance models with nodes clustering. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(3), 185–196.
- Arle JE, & Shils JL (2011). *Essential neuromodulation*. Academic Press.
- Austin A, Linkletter C, & Wu Z (2013). Covariate-defined latent space random effects model. *Social networks*, 35(3), 338–346.
- Baleyrier C, & Mauguier F (1980). The duality of the cingulate gyrus in monkey. *neuroanatomical study and functional hypothesis*. *Brain: a journal of neurology*, 103(3), 525–554. [PubMed: 6774795]
- Barajas A, Ochoa S, Obiols JE, & Lalucat-Jo L (2015). Gender differences in individuals at high-risk of psychosis: A comprehensive literature review. *The Scientific World Journal*, 2015.
- Barbillon P, Donnet S, Lazega E, & Bar-Hen A (2017). Stochastic block models for multiplex networks: An application to a multilevel network of researchers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 295–314.
- Bassett DS, & Sporns O (2017). Network neuroscience. *Nature neuroscience*, 20(3), 353–364. [PubMed: 28230844]
- Beckmann M, Johansen-Berg H, & Rushworth MF (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, 29(4), 1175–1190. [PubMed: 19176826]
- Berkovitch L, Charles L, Del Cul A, Hamdani N, Delavest M, Sarrazin S, Mangin J-F, Guevara P, Ji E, d'Albis M-A, et al. (2021). Disruption of conscious access in psychosis is associated with altered structural brain connectivity. *Journal of Neuroscience*, 41(3), 513–523. [PubMed: 33229501]
- Betzler R, Avena-Koenigsberger A, Goñi J, He Y, de Reus M, Griffa A, Vértes P, Mišić B, Thiran J, Hagmann P, et al. (2016). Generative models of the human connectome. *neuroimage*, 124 (part a), 1054–1064. [PubMed: 26427642]
- Binetti G, Cappa SF, Magni E, Padovani A, Bianchetti A, & Trabucchi M (1998). Visual and spatial perception in the early phase of alzheimer's disease. *Neuropsychology*, 12(1), 29. [PubMed: 9460732]
- Bullmore E, & Sporns O (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3), 186–198. [PubMed: 19190637]
- Cahn-Weiner DA, Sullivan EV, Shear PK, Fama R, Lim KO, Yesavage JA, Tinklenberg JR, & Pfefferbaum A (1999). Brain structural and cognitive correlates of clock drawing performance in alzheimer's disease. *Journal of the International Neuropsychological Society*, 5(6), 502–509. [PubMed: 10561930]
- Cai LY, Kerley CI, Yu C, Aboud KS, Beason-Held L, Shafer AT, Resnick SM, Jordan LC, Anderson AW, Schilling KG, et al. (2021). Joint cortical surface and structural connectivity analysis of alzheimer's disease. *Medical Imaging 2021: Image Processing*, 11596, 824–833.
- Cammoun L, Gigandet X, Meskaldji D, Thiran JP, Sporns O, Do KQ, Maeder P, Meuli R, & Hagmann P (2012). Mapping the human connectome at multiple scales with diffusion spectrum mri. *Journal of Neuroscience Methods*, 203(2), 386–397. 10.1016/j.jneumeth.2011.09.031 [PubMed: 22001222]
- Carter R, et al. (2015). On visual hallucinations and cortical networks: A trans-diagnostic review. *Journal of neurology*, 262(7), 1780–1790. [PubMed: 25761375]
- Chiu P-Y, Teng P-R, Wei C-Y, Wang C-W, & Tsai C-T (2018). Gender difference in the association and presentation of visual hallucinations in dementia with lewy bodies: A cross-sectional study. *International journal of geriatric psychiatry*, 33(1), 193–199. [PubMed: 28295599]
- Crayton JW, Stålberg E, & Hilton-Brown P (1977). The motor unit in psychotic patients: A single fibre emg study. *Journal of Neurology, Neurosurgery & Psychiatry*, 40 (5), 455–463. [PubMed: 197211]
- Cushman LA, & Duffy CJ (2007). The sex specificity of navigational strategies in alzheimer disease. *Alzheimer Disease & Associated Disorders*, 21(2), 122–129. [PubMed: 17545737]
- D'Angelo S, Murphy TB, & Alfö M (2018). Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. *The Annals of Applied Statistics*. <https://api.semanticscholar.org/CorpusID:88523060>

- DePoloyi A, Rankin K, Mucke L, Miller B, & Gorno-Tempini M (2007). Spatial cognition and the human navigation network in ad and mci. *Neurology*, 69(10), 986–997. [PubMed: 17785667]
- Desmarais BA, & Cranmer SJ (2017). *Statistical inference in political networks research*. Oxford University Press Oxford.
- Felten DL, O'Banion MK, & Maida ME (2015). *Netter's atlas of neuroscience*. Elsevier Health Sciences.
- Feng F, Huang W, Meng Q, Hao W, Yao H, Zhou B, Guo Y, Zhao C, An N, Wang L, et al. (2021). Altered volume and structural connectivity of the hippocampus in alzheimer's disease and amnesic mild cognitive impairment. *Frontiers in Aging Neuroscience*, 13, 705030. [PubMed: 34675796]
- Fischl B, & Dale AM (2000a). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Science*, 97(20), 11050–11055. 10.1073/pnas.200033797
- Fischl B, & Dale AM (2000b). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20), 11050–11055.
- Fischl B, Sereno MI, Tootell RB, & Dale AM (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, 8(4), 272–284. [PubMed: 10619420]
- Fosdick BK, & Hoff PD (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, 110 (511), 1047–1056. [PubMed: 26848204]
- Frederikse ME, Lu A, Aylward E, Barta P, & Pearlson G (1999). Sex differences in the inferior parietal lobule. *Cerebral Cortex*, 9(8), 896–901. [PubMed: 10601007]
- Goldenberg A, Zheng AX, Fienberg SE, & Airoldi EM (2010). A survey of statistical network models.
- Gollini I, & Murphy TB (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1), 246–265.
- Groh JM, & Werner-Reiss U (2002). Visual and auditory integration. *Encyclopedia of the Human Brain*, 4, 739–752.
- Gu J, & Yu PL (2022). Joint latent space models for ranking data and social network. *Statistics and Computing*, 32(3), 51.
- Hartig M, Truran-Sacrey D, Raptentsetsang S, Simonson A, Mezher A, Schuff N, & Weiner M (2014). Ucsf freesurfer methods. *Alzheimer's Disease Neuroimaging Initiative*, 1–8.
- Hodel J, Leclerc X, Zuber M, Gerber S, Besson P, Marcaud V, Roubeau V, Brasme H, Ganzoui I, Ducreux D, et al. (2020). Structural connectivity and cortical thickness alterations in transient global amnesia. *American Journal of Neuroradiology*, 41(5), 798–803. [PubMed: 32381542]
- Hoff P (2007). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in neural information processing systems*, 20.
- Hoff P (2008). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in neural information processing systems*, 657–664.
- Hoff P (2021). Additive and multiplicative effects network models. *Statistical Science*, 36(1), 34–50.
- Hoff PD (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469), 286–295.
- Hoff PD (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and mathematical organization theory*, 15(4), 261.
- Hoff PD, Raftery AE, & Handcock MS (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.
- Holland PW, Laskey KB, & Leinhardt S (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2), 109–137.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, & Smith SM (2012). Fsl. *NeuroImage*, 62(2), 782–790. 10.1016/j.neuroimage.2011.09.015 [PubMed: 21979382]
- Joyce KE, Laurienti PJ, Burdette JH, & Hayasaka S (2010). A new measure of centrality for brain networks. *PloS one*, 5(8), e12200. [PubMed: 20808943]

- Kim B, Lee KH, Xue L, & Niu X (2018). A review of dynamic network models with latent variables. *Statistics surveys*, 12, 105. [PubMed: 31428219]
- Koscik T, O'Leary D, Moser DJ, Andreasen NC, & Nopoulos P (2009). Sex differences in parietal lobe morphology: Relationship to mental rotation performance. *Brain and cognition*, 69(3), 451–459. [PubMed: 18980790]
- Krivitsky PN, & Handcock MS (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of statistical software*, 24.
- Kurylo DD, Corkin S, Rizzo III JF, & Growdon JH (1996). Greater relative impairment of object recognition than of visuospatial abilities in alzheimer's disease. *Neuropsychology*, 10(1), 74.
- Laws KR, Irvine K, & Gale TM (2016). Sex differences in cognitive impairment in alzheimer's disease. *World journal of psychiatry*, 6(1), 54. [PubMed: 27014598]
- Lehmann B, Henson R, Geerligs L, White S, et al. (2021). Characterising group-level brain connectivity: A framework using bayesian exponential random graph models. *NeuroImage*, 225, 117480. [PubMed: 33099009]
- Lehtola S, Tuulari J, Karlsson L, Parkkola R, Merisaari H, Saunavaara J, Lähdesmäki T, Scheinin N, & Karlsson H (2019). Associations of age and sex with brain volumes and asymmetry in 2–5-week-old infants. *Brain Structure and Function*, 224 (1), 501–513. [PubMed: 30390153]
- Lineweaver TT, Salmon DP, Bondi MW, & Corey-Bloom J (2005). Differential effects of alzheimer's disease and huntington's disease on the performance of mental rotation. *Journal of the International Neuropsychological Society*, 11(1), 30–39. [PubMed: 15686606]
- Ma Z, Ma Z, & Yuan H (2020). Universal latent space model fitting for large networks with edge covariates. *The Journal of Machine Learning Research*, 21(1), 86–152.
- MacDonald PW, Levina E, & Zhu J (2022). Latent space models for multiplex networks with shared structure. *Biometrika*, 109(3), 683–706.
- Manjón JV, Coupé P, Concha L, Buades A, Collins DL, & Robles M (2013). Diffusion weighted image denoising using overcomplete local pca. *PLoS ONE*, 8(9). 10.1371/journal.pone.0073021
- Mendola JD, Cronin-Golomb A, Corkin S, & Growdon JH (1995). Prevalence of visual deficits in alzheimer's disease. *Optometry and vision science: official publication of the American Academy of Optometry*, 72(3), 155–167. [PubMed: 7609938]
- Meunier D, Lambiotte R, Fornito A, Ersche K, & Bullmore ET (2009). Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, 3, 37. [PubMed: 19949480]
- Millet X, Raoux N, Le Carret N, Bouisson J, Dartigues J-F, & Amieva H (2009). Gender-related differences in visuospatial memory persist in alzheimer's disease. *Archives of Clinical Neuropsychology*, 24 (8), 783–789. [PubMed: 19889648]
- Moore C, & Sciacca F (2019). Fiber assignment by continuous tracking algorithm (fact). [Radiopaedia.org](https://radiopaedia.org). 10.53347/rid-72014
- Nielsen J (1955). Occipital lobes, dreams and psychosis. *The Journal of Nervous and Mental Disease*, 121(1), 50–52. [PubMed: 14368313]
- Opsahl T, Agneessens F, & Skvoretz J (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3), 245–251.
- Perneczky R, Drzezga A, Diehl-Schmid J, Li Y, & Kurz A (2007). Gender differences in brain reserve. *Journal of neurology*, 254 (10), 1395–1400. [PubMed: 17934882]
- Ricker JH, Keenan P, & Jacobson MW (1994). Visuo-perceptual-spatial ability and visual memory in vascular dementia and dementia of the alzheimer type. *Neuropsychologia*, 32(10), 1287–1296. [PubMed: 7845568]
- Rikandi E, Pamilo S, Mäntylä T, Suvisaari J, Kiesepää T, Hari R, Seppä M, & Raji T (2017). Precuneus functioning differentiates first-episode psychosis patients during the fantasy movie alice in wonderland. *Psychological medicine*, 47 (3), 495–506. [PubMed: 27776563]
- Roe JM, Vidal-Piñero D, Sørensen Ø, Brandmaier AM, Düzel S, Gonzalez HA, Kievit RA, Knights E, Kühn S, Lindenberger U, et al. (2021). Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in alzheimer's disease. *Nature communications*, 12(1), 1–11.
- Rolls ET, & Wirth S (2018). Spatial representations in the primate hippocampus, and their functions in memory and navigation. *Progress in neurobiology*, 171, 90–113. [PubMed: 30219248]

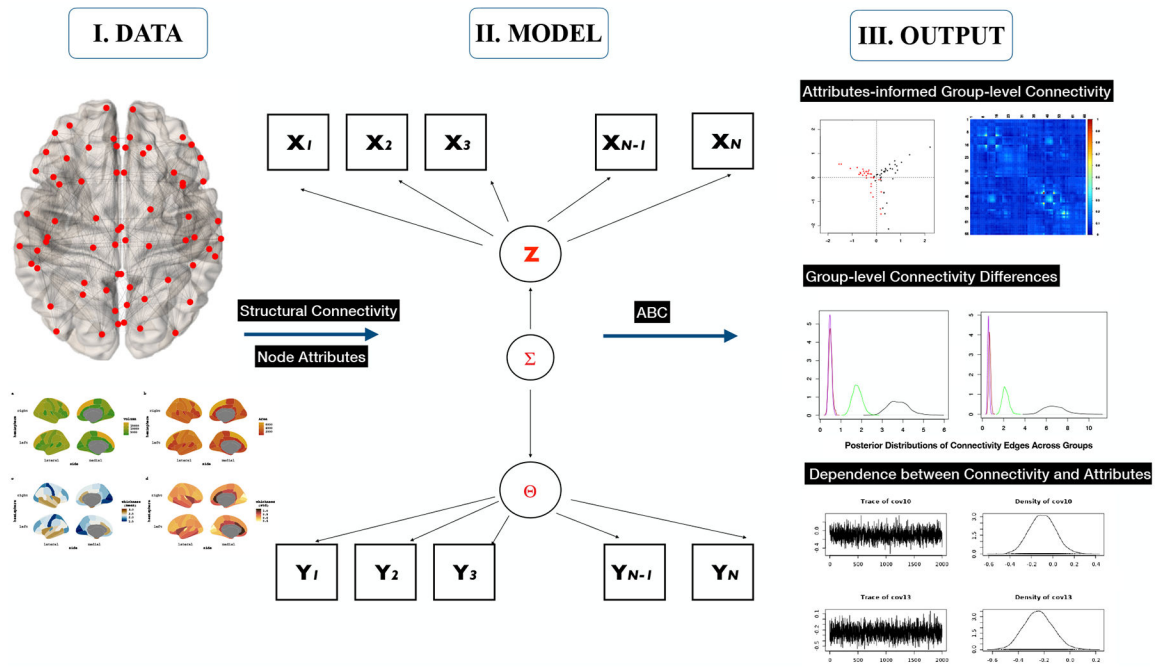
- Rubinov M, & Sporns O (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52(3), 1059–1069. [PubMed: 19819337]
- Salinas J, Mills ED, Conrad AL, Kosciak T, Andreasen NC, & Nopoulos P (2012). Sex differences in parietal lobe structure and development. *Gender Medicine*, 9(1), 44–55. [PubMed: 22333522]
- Salter-Townshend M, & McCormick TH (2017). Latent space models for multiview network data. *The annals of applied statistics*, 11(3), 1217. [PubMed: 29721127]
- Sinke MR, Dijkhuizen RM, Caimo A, Stam CJ, & Otte WM (2016). Bayesian exponential random graph modeling of whole-brain structural networks across lifespan. *NeuroImage*, 135, 79–91. [PubMed: 27132542]
- Sireteanu R, Oertel V, Mohr H, Linden D, & Singer W (2008). Graphical illustration and functional neuroimaging of visual hallucinations during prolonged blindfolding: A comparison to visual imagery. *Perception*, 37(12), 1805–1821. [PubMed: 19227374]
- Smith AL, Asta DM, & Calder CA (2019). The geometry of continuous latent space models for network data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 34(3), 428. [PubMed: 33235407]
- Snijders TA (2011). Statistical models for social networks. *Annual review of sociology*, 37, 131–153.
- Song M, Liu Y, Zhou Y, Wang K, Yu C, & Jiang T (2009). Default network and intelligence difference. *IEEE Transactions on autonomous mental development*, 1(2), 101–109.
- Sosa J, & Buitrago L (2021). A review of latent space models for social networks. *Revista Colombiana de Estadística*, 44(1), 171–200.
- Studholme C, Kroenke CD, & Dighe M (2020). Motion corrected mri differentiates male and female human brain growth trajectories from mid-gestation. *Nature communications*, 11(1), 1–16.
- Sweet T, & Adhikari S (2020). A latent space network model for social influence. *psychometrika*, 85(2), 251–274. [PubMed: 32221792]
- Sweet TM (2015). Incorporating covariates into stochastic blockmodels. *Journal of Educational and Behavioral Statistics*, 40(6), 635–664.
- Sweet TM (2016). Social network methods for the educational and psychological sciences. *Educational Psychologist*, 51(3–4), 381–394.
- Vivar JC, & Banks D (2012). Models for networks: A cross-disciplinary science. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4 (1), 13–27.
- Vogt BA, & Laureys S (2009). The primate posterior cingulate gyrus: Connections, sensorimotor orientation, gateway to limbic processing. *Cingulate neurobiology and disease*, 275–308
- Wang S (2021). Recent integrations of latent variable network modeling with psychometric models. *Frontiers in Psychology*, 12.
- Wang S, Liu Y, Xu W, Tian X, & Zhao Y (2023). Inference-based statistical network analysis uncovers star-like brain functional architectures for internalizing psychopathology in children. *arXiv preprint arXiv:2309.11349*.
- Wang S, Paul S, & De Boeck P (2023). Joint latent space model for social networks with multivariate attributes. *Psychometrika*, 1–31. [PubMed: 35687222]
- Wasserman S, & Faust K (1994). *Social network analysis: Methods and applications*.
- Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuciak JA, Soares H, Toga AW, ... Alzheimer's Disease Neuroimaging, I. (2013). The alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimers Dement*, 9(5), e111–94. [PubMed: 23932184]
- Weiner MW, Aisen PS, Jack CR Jr, Jagust WJ, Trojanowski JQ, Shaw L, Saykin AJ, Morris JC, Cairns N, Beckett LA, et al. (2010). The alzheimer's disease neuroimaging initiative: Progress report and future plans. *Alzheimer's & Dementia*, 6(3), 202–211.
- Wilson JD, Cranmer S, & Lu Z-L (2020). A hierarchical latent space network model for population studies of functional connectivity. *Computational Brain & Behavior*, 3(4), 384–399.
- Xu F, Garai S, Duong-Tran D, Saykin AJ, Zhao Y, & Shen L (2022). Consistency of graph theoretical measurements of alzheimer's disease fiber density connectomes across multiple parcellation

- scales. (BIBM) 2022 IEEE International Conference on Bioinformatics and Biomedicine IEEE, Regular Paper.
- Yan J, Raja VV, Huang Z, Amico E, Nho K, Fang S, Sporns O, Wu YC, Saykin A, Goñi J, & Shen L (2020). Brain-wide structural connectivity alterations under the control of alzheimer risk genes. *International Journal of Computational Biology and Drug Design*, 13(1), 58. 10.1504/ijcbdd.2020.105098 [PubMed: 32095160]
- Yang J, McAuley J, & Leskovec J (2013). Community detection in networks with node attributes. 2013 IEEE 13th international conference on data mining, 1151–1156.
- Yee Y, Fernandes DJ, French L, Ellegood J, Cahill LS, Vousden DA, Noakes LS, Scholz J, van Eede MC, Nieman BJ, et al. (2018). Structural covariance of brain region volumes is associated with both structural connectivity and transcriptomic similarity. *Neuroimage*, 179, 357–372. [PubMed: 29782994]
- Zhang X, Xu G, & Zhu J (2022). Joint latent space models for network data with high-dimensional node variables. *Biometrika*, 109(3), 707–720.
- Zmigrod L, Garrison JR, Carr J, & Simons JS (2016). The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 69, 113–123. [PubMed: 27473935]
- Zolal A, Sobottka SB, Podlesek D, Linn J, Rieger B, Juratli TA, Schackert G, & Kitzler HH (2016). Comparison of probabilistic and deterministic fiber tracking of cranial nerves. *Journal of Neurosurgery*, 127(3), 613–621. [PubMed: 27982771]
- Zuo X-N, Ehmke R, Mennes M, Imperati D, Castellanos FX, Sporns O, & Milham MP (2012). Network centrality in the human functional connectome. *Cerebral cortex*, 22(8), 1862–1875. [PubMed: 21968567]

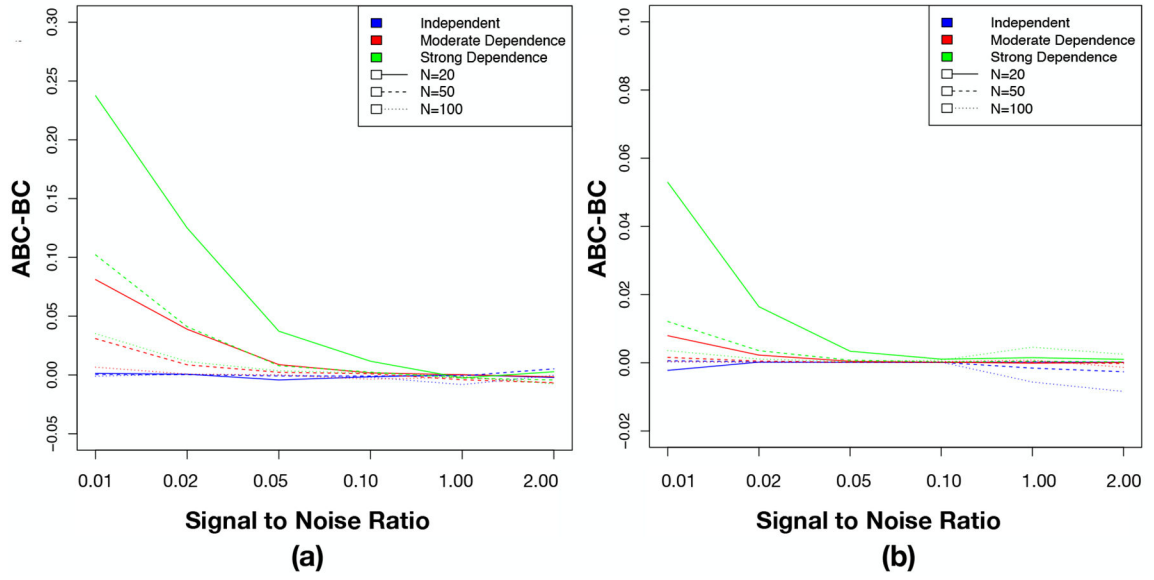
An interpretable latent space representation of the group-level connectivity

An incorporation of the anatomical knowledge of nodes and the testing of its co-varying relationship with connectivity

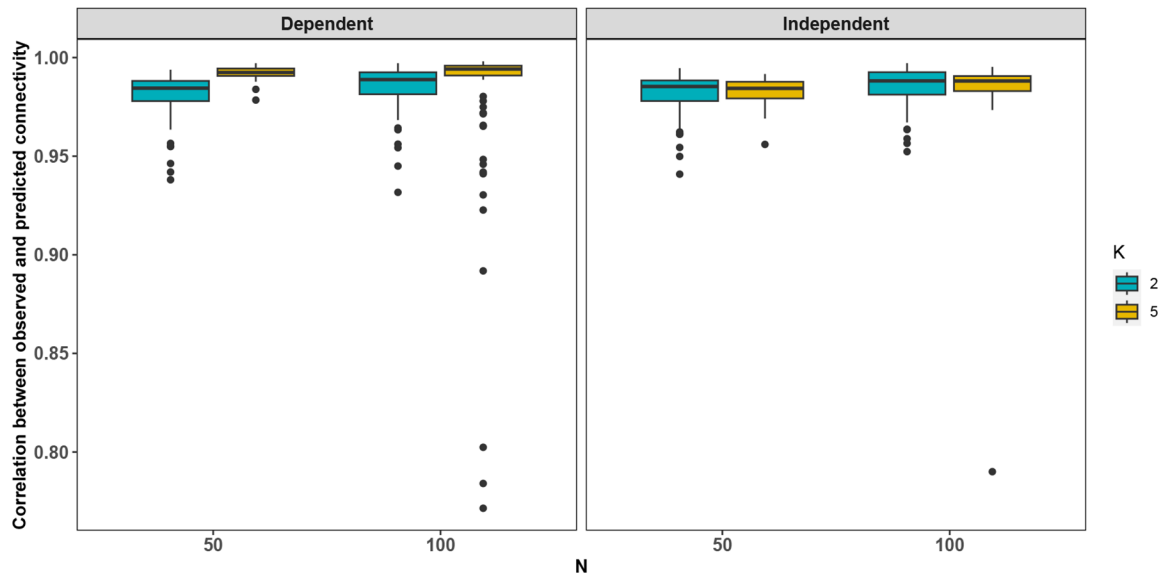
A quantification of the uncertainty and an evaluation of the likelihood of the estimated group-level effects against chance



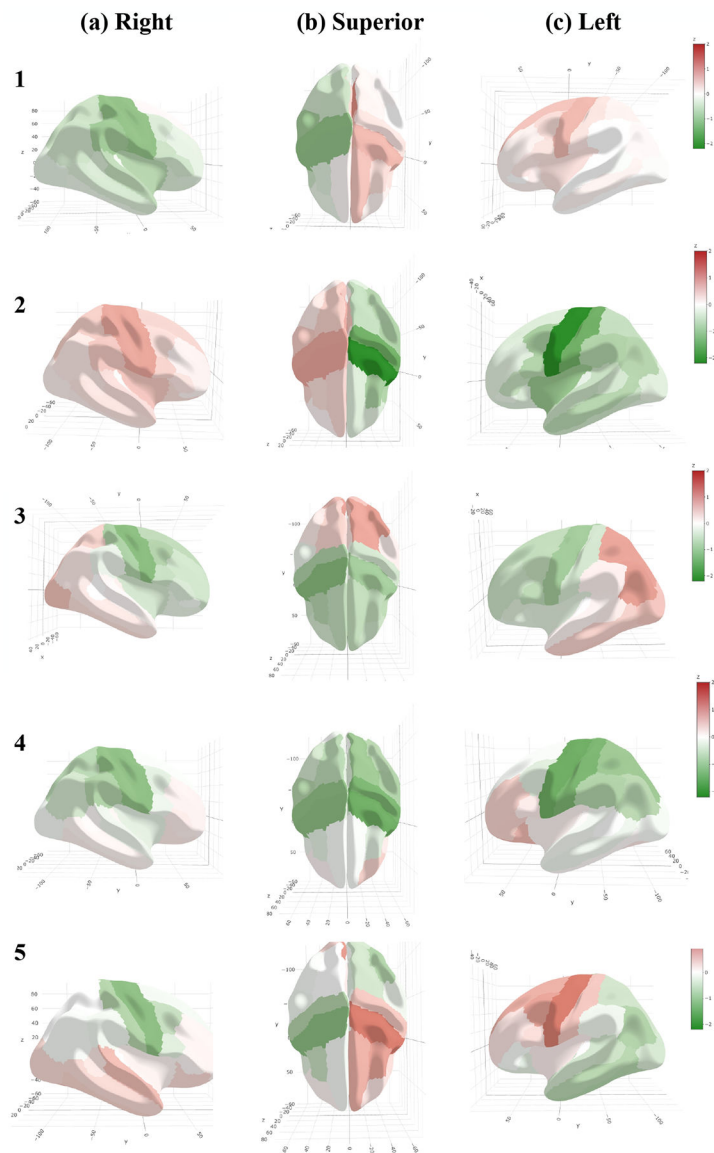
**Figure 1:** The schematic diagram of ABC pipeline. With the input of structural connectivity matrices,  $\{X_1, X_2, \dots, X_N\}$  and volumetric measures,  $\{Y_1, Y_2, \dots, Y_N\}$ , we propose a joint latent variable model, where we allow the latent connectivity variables,  $Z$  and latent behavior variables,  $\Theta$  to co-vary with a shared covariance matrix,  $\Sigma$  in the generative Bayesian framework. Via the ABC model, we simultaneously estimate the group-level structural connectivity informed by the volumetric attributes, identify imaging biomarkers comparing the healthy and patient population and infer the dependence between structural connectivity and attributes.



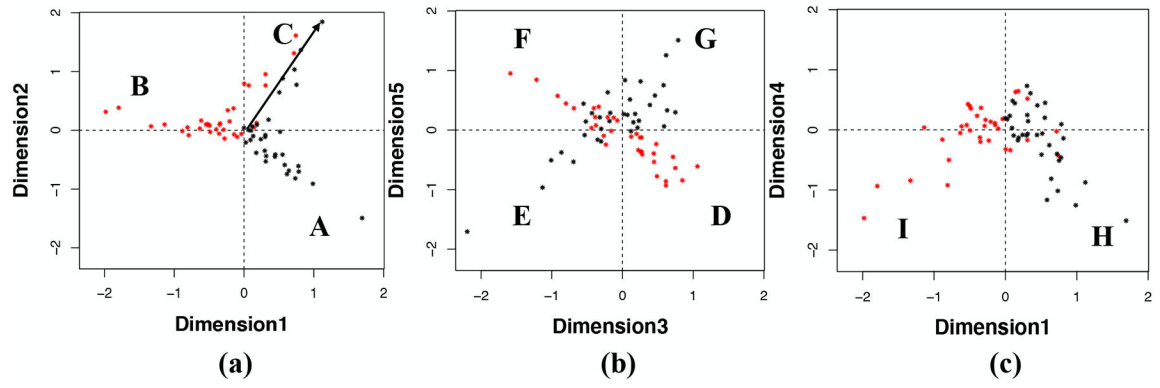
**Figure 2:** The difference in the correlational predictive power between ABC and BC when (a)  $V = 20$  and (b)  $V = 70$ . Positive values indicate that ABC outperforms BC with higher correlations between the predicted and the observed brain connectivity. Negative values indicate that BC outperforms ABC with higher correlations between the predicted and the observed brain connectivity.



**Figure 3:**  
The correlations between the estimated and predicted connectivity in independent data across 100 random samples when the true dimensionality is 2 (blue) versus 5 (yellow).

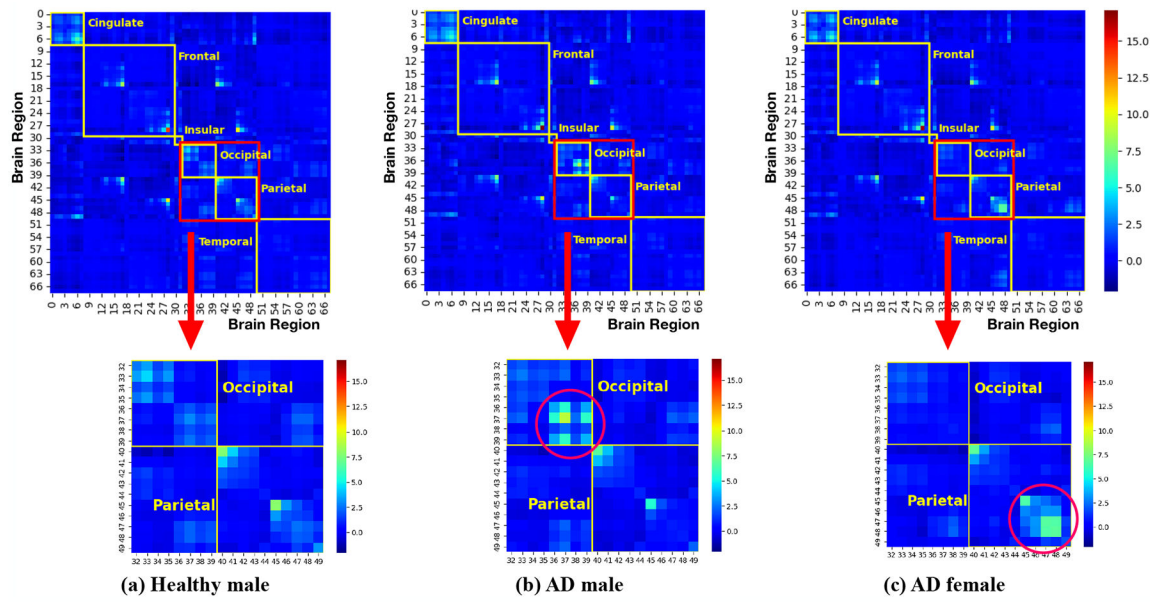


**Figure 4:** Five estimated latent connectivity dimensions from (a) the lateral view of the right hemisphere, (b) the superior view of the brain and (c) the lateral view of the left hemisphere. Brain regions are colored based on the values of the latent dimensions with red indicating positive values and green indicating negative values.

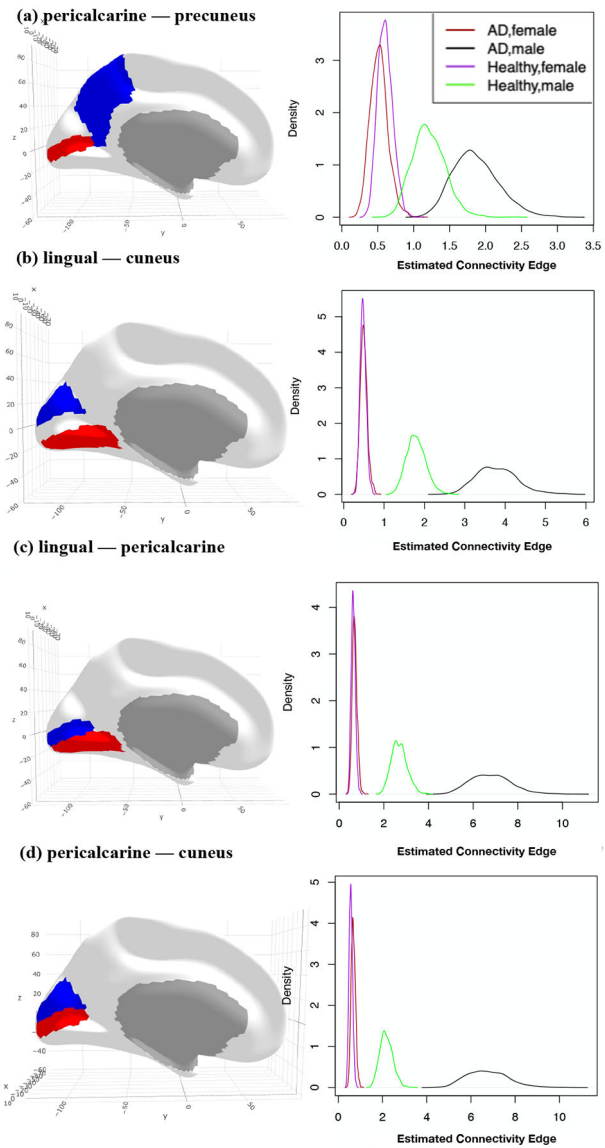


**Figure 5:**

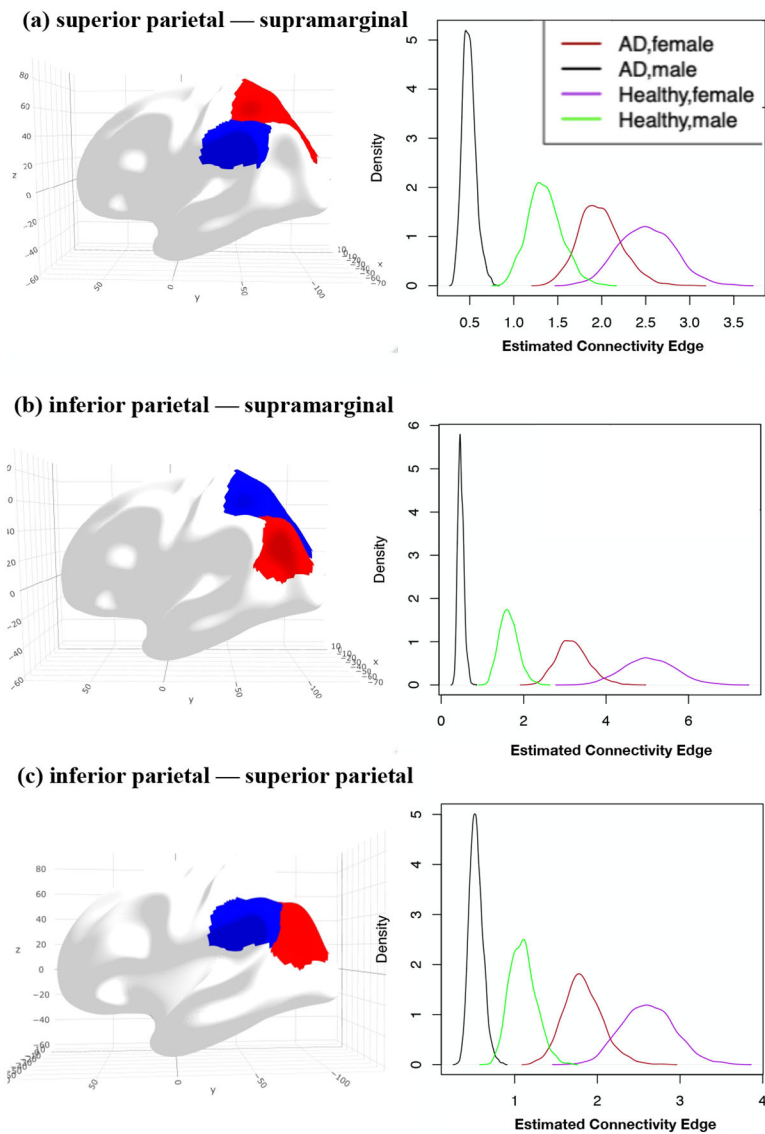
The latent space representations of the structural connectivity between (a) dimensions 1 and 2, (b) dimensions 3 and 5 and (c) dimensions 1 and 4. Each dot of the latent space represents a brain region. The color differentiates the left hemisphere (black) from the right hemisphere (red).



**Figure 6:**  
 Estimated attributes-informed brain connectivity for (a) healthy male subjects, (b) male subjects with AD, and (c) female subjects with AD.



**Figure 7:** Sex-based difference in AD phenotype for the structural connectivity between regions of (a) pericalcarine and precuneus (b) lingual and cuneus, (c) lingual and pericalcarine, and (d) pericalcarine and cuneus.



**Figure 8:** Sex-based difference in AD phenotype for the structural connectivity between regions of (a) superior parietal and supramarginal, (b) inferior parietal and supramarginal, and (c) inferior parietal and superior parietal.

**Table 1:**

Prediction of the group-level connectivity

<i>N</i>	<i>V</i>	Dependence	<i>S/N</i>	MSE			Correlation	
				ABC	Average	MSBM	ABC	Average
50	20	IND	0.05	0.71 (0.31)	0.82 (0.09)	1.90 (0.86)	0.90 (0.06)	0.82 (0.07)
			0.02	0.87 (0.34)	2.02 (0.21)	1.90 (0.91)	0.81 (0.11)	0.67 (0.10)
			0.01	1.12 (0.39)	4.02 (0.40)	1.93 (0.90)	0.69 (0.16)	0.54 (0.10)
		DEP	0.05	0.68 (0.28)	0.82 (0.09)	1.94 (0.80)	<b>0.91</b> (0.06)	<b>0.82</b> (0.06)
			0.02	0.80 (0.31)	2.01 (0.20)	2.00 (0.96)	<b>0.86</b> (0.08)	<b>0.68</b> (0.09)
			0.01	0.94 (0.35)	4.01 (0.40)	2.01 (0.96)	<b>0.80</b> (0.10)	<b>0.56</b> (0.10)
	70	IND	0.05	0.57 (0.13)	0.82 (0.04)	1.44 (0.42)	0.97 (0.01)	0.84 (0.03)
			0.02	0.62 (0.13)	2.02 (0.07)	1.80 (0.36)	0.95 (0.02)	0.70 (0.04)
			0.01	0.71 (0.14)	4.02 (0.13)	2.03 (0.50)	0.92 (0.02)	0.57 (0.05)
		DEP	0.05	0.56 (0.14)	0.82 (0.04)	1.36 (0.29)	<b>0.97</b> (0.02)	<b>0.84</b> (0.03)
			0.02	0.59 (0.14)	2.02 (0.07)	1.73 (0.35)	<b>0.96</b> (0.02)	<b>0.69</b> (0.04)
			0.01	0.66 (0.15)	4.02 (0.13)	1.99 (0.48)	<b>0.93</b> (0.02)	<b>0.57</b> (0.04)
100	20	IND	0.05	0.63 (0.29)	0.40 (0.04)	1.88 (0.85)	0.93 (0.04)	0.90 (0.05)
			0.02	0.72 (0.29)	0.99 (0.09)	1.90 (0.90)	0.89 (0.07)	0.79 (0.08)
			0.01	0.86 (0.30)	1.98 (0.19)	1.91 (0.90)	0.82 (0.11)	0.68 (0.10)
		DEP	0.05	0.63 (0.27)	0.40 (0.04)	1.92 (0.79)	<b>0.94</b> (0.04)	<b>0.90</b> (0.04)
			0.02	0.70 (0.28)	0.99 (0.09)	1.98 (0.97)	<b>0.91</b> (0.07)	<b>0.80</b> (0.07)
			0.01	0.79 (0.30)	1.98 (0.19)	1.99 (0.97)	<b>0.86</b> (0.10)	<b>0.69</b> (0.08)
	70	IND	0.05	0.55 (0.13)	0.41 (0.02)	1.36 (0.30)	0.98 (0.01)	0.91 (0.02)
			0.02	0.57 (0.13)	1.00 (0.03)	1.76 (0.35)	0.97 (0.02)	0.81 (0.03)
			0.01	0.63 (0.14)	2.00 (0.06)	2.03 (0.50)	0.95 (0.02)	0.70 (0.04)
		DEP	0.05	0.54 (0.13)	0.41 (0.02)	1.36 (0.35)	<b>0.98</b> (0.02)	<b>0.91</b> (0.02)
			0.02	0.55 (0.14)	1.00 (0.03)	1.75 (0.34)	<b>0.97</b> (0.01)	<b>0.80</b> (0.03)
			0.01	0.60 (0.14)	2.00 (0.06)	1.98 (0.48)	<b>0.96</b> (0.02)	<b>0.70</b> (0.04)

**Table 2:**

The correlation between the predicted and the observed connectivity

Dimension ( $K$ )	AD Female	AD Male	Healthy Female	Healthy Male
2	0.396	0.370	0.368	0.424
3	0.448	0.466	0.459	0.453
4	0.515	0.515	0.523	0.529
5	<b>0.547</b>	<b>0.526</b>	<b>0.556</b>	<b>0.542</b>
6	0.414	0.306	0.513	0.344
7	0.363	0.352	0.239	0.200
8	0.323	0.437	0.191	0.228
Test	0.508	0.561	0.540	0.570

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

The mean of the identified imaging biomarkers via the ABC model

	AD		Healthy		AD Female/Male <sup>1</sup>
	Female	Male	Female	Male	
pericalcarine—cuneus	8.708	11.437	10.530	10.065	-2.728***
lingual—pericalcarine	8.471	9.924	8.693	9.521	-1.452*
lingual—cuneus	1.501	1.617	1.663	1.868	-0.116
pericalcarine—precuneus	1.326	1.278	1.077	1.026	0.048
inferior parietal—supramarginal	2.800	3.412	3.405	3.051	-0.611
superior parietal—supramarginal	3.580	3.456	3.553	3.918	0.124
inferior parietal—superior parietal	11.249	10.650	10.615	10.204	0.599

<sup>1</sup>Mean Difference between AD females and AD males.