

Classification of longitudinal brain networks with an application to understanding superior aging

Lu Wang¹  | Zhengwu Zhang²

¹Department of Statistics, Central South University, Changsha, 410083, China

²Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA

Correspondence

Lu Wang, Department of Statistics, Central South University, Changsha 410083, China.
Email: wanglu_stat@csu.edu.cn

Funding information

National Institutes of Health, Grant/Award Number: MH118927; National Natural Science Foundation of China, Grant/Award Number: 11901583; Roberta K. Courtman Revocable Trust

This paper studies the problem of classifying longitudinal structural brain networks to identify meaningful substructures and their time-varying effects. The problem is motivated by a subpopulation of healthy older adults who can maintain excellent cognitive functions across time, while others usually have cognitive decline in aging. It is of substantial scientific interest to study neurological mechanisms behind this successful aging phenomena; however, existing statistical tools for longitudinal networks are very limited. We propose a structured classification method that could identify a set of small outcome-relevant subgraphs and estimate the age effect of each signal subgraph from the longitudinal network predictors, as well as an efficient algorithm for model estimation. Application of this method to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data identifies a small set of brain regions whose connectivity strengths are predictive of successful cognitive aging, which has more appealing interpretation and better predictive performance compared with unstructured classification methods.

KEYWORDS

classification, generalized linear models, networks, statistical modelling, variable selection

1 | INTRODUCTION

With advanced neuroimaging technologies, more and more large neuroscience studies start collecting longitudinal brain scans, for example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack Jr et al., 2008) and the recent Adolescent Brain Cognitive Development (ABCD) study (Casey et al., 2018). Compared with cross-sectional data, longitudinal network data have the advantage of revealing dynamic changes across time and their effect on the final outcome. In this paper, we are interested in studying a subset of healthy older adults in ADNI who, instead of having a declined trend, can maintain their excellent cognitive trajectories across time. This subpopulation is often referred as *supernormals* in the literature (Lin, Ren, et al., 2017; Lin, Wang, et al., 2017). Studying neural mechanisms of this subpopulation can provide valuable therapeutic targets to delay cognitive decline and prevent severe cognitive impairment (Alzheimer's or another dementia).

Using a state-of-the-art connectome extraction pipeline (Zhang, Descoteaux, et al., 2018), we extracted longitudinal structural brain networks over a 5-year span for 40 supernormals and 45 cognitively normal controls with matched age, gender and education. Each subject's diffusion magnetic resonance imaging (MRI) and structural MRI were used in this preprocessing process, and more preprocessing details are described in Section 4. The structural brain network corresponds to a set of white matter connections among predefined brain regions. In our data, the number of observed structural brain networks for each individual ranges from 1 to 5 since not all the participants visited every year during the 5-year programme. Each structural brain network is characterized by a weighted adjacency matrix where each element denotes the connectivity strength of fibers between a pair of brain regions. Figure 1 shows a profile of the dataset we used.

We aim to study the relationship between successful cognitive aging and structural brain networks in hope of finding neurologically interpretable structural markers in brain connectome predictive of supernormals, for example, a subgraph formed by a small set of closely interrelated brain

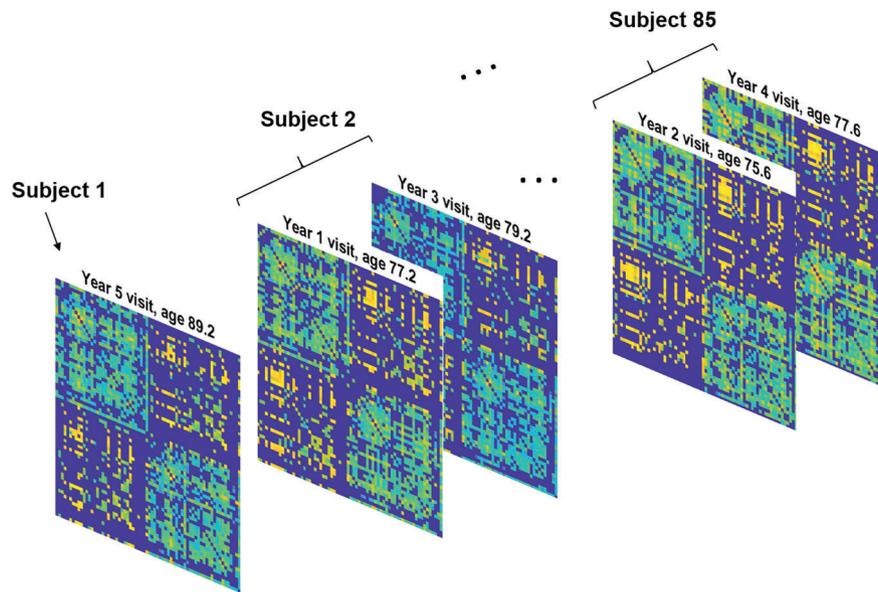


FIGURE 1 The profile of longitudinal structural brain networks for individuals in the dataset

regions. With age information of the subjects, we also aim to estimate the age effect of each outcome-relevant subgraph, that is the dynamic importance of each subgraph in preserving cognition, which may improve cognition diagnosis and provide insights on the role of aging in normal and diseased brains.

The signal subgraph learning is basically a variable selection problem where the number of predictors—connections in the longitudinal brain networks—could easily exceed the number of observations (subjects). One typical approach to this large p small n problem would be some generalized linear model with certain regularization, such as LASSO (Tibshirani, 1996), Elastic-Net (Zou & Hastie, 2005) and SCAD (Fan & Li, 2001). But these methods require first transforming each adjacency matrix into a long vector and do not guarantee any structure among the selected connections in the brain network, making the results hard to interpret.

There are some advanced statistical methods taking into account network structure in the variable selection problem. Arroyo Reli3n et al. (2019) proposed a graph classification method that incorporates L_1 and group lasso penalty in selection of edges and nodes. Guha and Rodriguez (2020) and Wang et al. (2019) both built on the idea of sparse and low-rank matrix decomposition and used Bayesian shrinkage or penalized optimization to identify important nodes and edges. But these methods are not applicable to longitudinal network classification.

The longitudinal networks can be easily formed into a multidimensional array (tensor), and hence, tensor regression models (Zhou et al. 2013) are relevant, which directly work with a tensor covariate without vectorization. Zhang, Li, et al. (2019) proposed tensor generalized estimating equations for longitudinal imaging analysis, but they consider a general-order tensor, which is not suitable for the case of symmetric adjacency matrices, and the response is time-varying. There is a line of research following tensor regression in imaging and network data analysis (Sun & Li, 2019; Wang et al., 2021; Zhang, Sun, et al., 2018, 2019), but none can be directly applied to our problem.

We propose a longitudinal network classification model with partial time-varying coefficients and elastic-net penalty to learn a set of small signal subgraphs and estimate their temporal effects (age effects in our application). The signal subgraphs are cliques (fully-linked graphs) which have appealing interpretation in neurological studies, as many complex cognitive processes are the product of coordinated activities among several brain regions. The estimated age effect of each subgraph reflects how its predictive effect on the outcome varies over time, which can improve our understanding of the interaction between the aging of brain structure and cognitive functioning. The rest of the paper is organized as follows. Our proposed model and model estimation algorithm are introduced in Section 2. Section 3 presents a simulation study evaluating the performance of our model in recovering true signal subgraphs and age effects. We apply our method to a longitudinal dataset of ADNI in Section 4, and Section 5 concludes.

2 | METHOD

The data can be summarized as $(y_i, \{(W_i^{(s)}, g_{is}) : s = 1, \dots, T_i\})$ for each subject i , $i = 1, \dots, n$. The binary response $y_i = 1$ denotes that subject i is a *supernormal* and $y_i = 0$ for a normal control. We use the clinical criteria in Lin, Ren, et al. (2017) and Lin, Wang, et al. (2017) to

identify supernormals from healthy older subjects. Note that the most important feature for a subject to be a supernormal is his or her ability to maintain a stable (or an increasing) cognitive function trajectory across a 5-year period. Multiple brain scans were taken in this period, and $W_i^{(s)}$ is the weighted adjacency matrix of the structural brain network measured for subject i at the s th visit ($s = 1, \dots, T_i$), and g_{is} is the age of subject i when the network $W_i^{(s)}$ was observed. Note that each $W_i^{(s)}$ measures the connectivity strengths among the same set of V brain regions and is an undirected network without self loops. Therefore, each $W_i^{(s)}$ is a $V \times V$ symmetric matrix with zero diagonal. Our goal is to learn a set of small signal subgraphs from the brain network predictive of the outcome and estimate their age effects.

2.1 | Model

We propose the following longitudinal network classification model (LNCM):

$$y_i | \{(g_{is}, W_i^{(s)}) : s = 1, \dots, T_i\} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_i), i = 1, \dots, n, \quad (1)$$

$$\text{logit}(p_i) = \alpha_0 + \sum_{h=1}^K \frac{1}{T_i} \sum_{s=1}^{T_i} \lambda_h(g_{is}) \beta_h^\top W_i^{(s)} \beta_h,$$

where $\beta_h \in \mathbb{R}^V$ and $\lambda_h(g)$ is a function of age g . Model (1) assumes that the binary outcome y_i of each individual follows an independent Bernoulli distribution given his/her longitudinal network observations and the corresponding age information $\{(W_i^{(s)}, g_{is}) : s = 1, \dots, T_i\}$. The bilinear part in the logit link of (1) is divided by T_i because not all the subjects have the same number of network observations. Note that we impose a linear relationship between the log odds of the response and the edges of longitudinal networks. While this model may have the risk of losing some global topological information encoded in $W_i^{(s)}$, it does not lose any information at the edge level and brings significant convenience for our signal subgraph learning problem. This linear assumption has been widely used in classification or regression with network predictors (Arroyo Reli3n et al., 2019; Guha & Rodriguez, 2020; Zhou et al., 2013).

The coefficients in model (1) are assumed to have K components, where each component matrix $\beta_h \beta_h^\top$ selects a predictive subgraph. For ease of interpretation, the logit link of (1) can be written in the following matrix dot product form:

$$\text{logit}(p_i) = \alpha_0 + \frac{1}{T_i} \sum_{s=1}^{T_i} \sum_{h=1}^K \langle \lambda_h(g_{is}) \beta_h \beta_h^\top, W_i^{(s)} \rangle \quad (2)$$

where $\langle B, W \rangle = \text{trace}(B^\top W) = \text{vec}(B)^\top \text{vec}(W)$. Moreover, Equation (2) indicates that the nonzero entries in each component matrix $\beta_h \beta_h^\top$ locate an outcome-relevant clique subgraph in the network predictor. The dynamic contribution of each subgraph to the outcome is captured by $\lambda_h(g)$, which also avoids constraining the coefficient matrix in (2) to be positive semi-definite. We do not let β_h vary with age in (2) for two reasons. First, the brain regions in each subgraph associated with y are assumed to be stable across time for healthy adults. Second, β_h is fixed over time for model simplicity. Otherwise a random function has to be assigned to each entry of β_h ($h = 1, \dots, K$), leading to intractable estimation and overfitting issues.

Suppose that the connection strengths of a subgraph corresponding to $\beta_h \beta_h^\top$ tend to *decrease* for older adults with normal aging, while to maintain unchanged for supernormals. Then *higher* weight should be put on the term $\beta_h^\top W_i^{(s)} \beta_h$ observed at an older age in predicting y . In this case, $\lambda_h(g)$ would be expected to increase with age and may be assumed as a linear function $\lambda_h(g) = \rho_h g + \alpha_h$. Alternatively, $\lambda_h(g)$ could be a higher order polynomial function to increase flexibility. In simulations and applications, we let $\lambda_h(g)$ be up to the second order, that is, $\lambda_h(g) = \gamma_h g^2 + \rho_h g + \alpha_h$, for interpretation and model simplicity, because higher order terms of age are difficult to interpret and prone to overfitting. In addition, we rarely observed nonzero coefficient estimates for the quadratic terms of age in real data analysis.

Plugging in the quadratic function $\lambda_h(g) = \gamma_h g^2 + \rho_h g + \alpha_h$ into (2), we have

$$\text{logit}(p_i) = \alpha_0 + \sum_{h=1}^K \left[\langle \alpha_h \beta_h \beta_h^\top, \frac{1}{T_i} \sum_{s=1}^{T_i} W_i^{(s)} \rangle + \langle \rho_h \beta_h \beta_h^\top, \frac{1}{T_i} \sum_{s=1}^{T_i} g_{is} W_i^{(s)} \rangle + \langle \gamma_h \beta_h \beta_h^\top, \frac{1}{T_i} \sum_{s=1}^{T_i} g_{is}^2 W_i^{(s)} \rangle \right]. \quad (3)$$

Equation (3) implies that the actual covariates for each subject i in model (1) are not the raw longitudinal network observations, but the average of his/her networks, the weighted average of his/her networks by age and squared age, respectively. To ensure both the identifiability of the model and the sparsity of coefficient matrices $\{(\alpha_h \beta_h \beta_h^\top, \rho_h \beta_h \beta_h^\top, \gamma_h \beta_h \beta_h^\top) : h = 1, \dots, K\}$, we penalize the magnitude of the lower triangular entries in these coefficient matrices with the following elastic-net penalty:

$$\delta \sum_{h=1}^K \sum_{u=1}^V \sum_{v < u} [\eta(|\alpha_h| + |\rho_h| + |\gamma_h|) |\beta_{hu}| |\beta_{hv}| + (1-\eta)(\alpha_h^2 + \rho_h^2 + \gamma_h^2) \beta_{hu}^2 \beta_{hv}^2 / 2] \quad (4)$$

where the overall penalty factor $\delta > 0$ and $\eta \in [0, 1]$ controlling the fraction of L_1 penalty.

The goal of model (2) is to identify more interpretable signal subgraphs (in the form of clique graphs) for classifying subjects from longitudinal observed networks. The model will have the most power if the true signals are in the form of clique graphs; but what will happen if the true signal subgraphs are not in the form of clique graphs? Without considering the penalty term, there are possibilities: (1) we will need to have a large K to gain flexibility due to the clique constraint; or (2) the selected clique graphs have overlapping edges that are canceling out each other to approximate the true regression coefficients. The penalty term in (4) actually encourages the former solution over the later one.

2.2 | Model estimation

The parameters in LNCM (1) are estimated by minimizing the loss function below:

$$\text{Loss function} = -\frac{1}{n} \sum_{i=1}^n l_i + \sum_{h=1}^K \sum_{u=1}^V \sum_{v < u} \delta [\eta(|\alpha_h| + |\rho_h| + |\gamma_h|) |\beta_{hu}| |\beta_{hv}| + (1-\eta)(\alpha_h^2 + \rho_h^2 + \gamma_h^2) \beta_{hu}^2 \beta_{hv}^2 / 2] \quad (5)$$

where l_i is the log-likelihood of subject i .

The adjacency matrices $\{W_i^{(s)}\}$ and age $\{g_{is}\}$ of subjects are standardized as follows before fitting the model. Each cell of the adjacency matrix is normalized to have mean 0 and variance 1 with respect to all $\sum_{i=1}^n T_i$ observations. Age and squared age are also standardized in a similar way. Let $\{\tilde{g}_{is}\}$ and $\{\tilde{g}_{is}^2\}$ denote the standardized $\{g_{is}\}$ and $\{g_{is}^2\}$, respectively. Then the entries in the matrix predictors $\frac{1}{T_i} \sum_{s=1}^{T_i} W_i^{(s)}$, $\frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is} W_i^{(s)}$ and $\frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is}^2 W_i^{(s)}$ in (3) are roughly of the same magnitude. After obtaining the estimated $(\hat{\alpha}_h, \hat{\rho}_h, \hat{\gamma}_h)$ with the standardized covariates, it is easy to recover age effects in the original scale through $\hat{\lambda}_h(g_{is}) = \hat{\gamma}_h^{(o)} g_{is}^2 + \hat{\rho}_h^{(o)} g_{is} + \hat{\alpha}_h^{(o)}$ with

$$\hat{\gamma}_h^{(o)} = \frac{\hat{\gamma}_h}{\sigma_2}, \hat{\rho}_h^{(o)} = \frac{\hat{\rho}_h}{\sigma_1}, \hat{\alpha}_h^{(o)} = \hat{\alpha}_h - \hat{\rho}_h \frac{\mu_1}{\sigma_1} - \hat{\gamma}_h \frac{\mu_2}{\sigma_2} \quad (6)$$

where μ_1, μ_2 are the means of $\{g_{is}\}$ and $\{g_{is}^2\}$, respectively, and σ_1, σ_2 their standard deviations.

Although there is scaling indeterminacy between $\lambda_h(g)$ and $\beta_h \beta_h^\top$ within each component such that

$$\lambda_h(g_{is}) \beta_h \beta_h^\top = c_h \lambda_h(g_{is}) \frac{1}{c_h} \beta_h \beta_h^\top$$

for any $c_h \neq 0$, this scaling of $\lambda_h(g)$ does not change the maximizer or minimizer of $\lambda_h(g)$ if $\lambda_h(g)$ is a quadratic function, or the linear trend if $\lambda_h(g)$ is a linear or constant function. In practice, we always report the estimated age effect $\lambda_h(g)$ after scaling $\beta_h \beta_h^\top$ so that the off-diagonal element with the largest magnitude is 1 for each nonempty component.

The logit(p_i) in model (1) is a quadratic function of each β_h and the second derivative of l_i with respect to β_h is

$$\frac{\partial^2 l_i}{\partial \beta_h \partial \beta_h^\top} = (y_i - p_i) \frac{2}{T_i} \sum_{s=1}^{T_i} \lambda_h(g_{is}) W_i^{(s)} - p_i (1 - p_i) \frac{4}{T_i^2} \sum_{s=1}^{T_i} \sum_{t=1}^{T_i} \lambda_h(g_{is}) \lambda_h(g_{it}) W_i^{(s)} \beta_h \beta_h^\top W_i^{(t)}$$

which may not be negative semi-definite. Therefore the loss function (5) is not a convex function of β_h when fixing the other parameters, and there is no closed form solution for β_h when block updating each component vector β_h , $h = 1, \dots, K$. Notice that the networks in this case are undirected without self loops. The diagonal entries of each adjacency matrix $W_i^{(s)}$ are zero. Then the loss function (5) is actually a convex function of each entry β_{hu} in β_h when fixing the others. So we employ the coordinate descent algorithm to minimize (5). The technical details of deriving the analytic form update for each parameter are discussed in the supporting information. The coordinate descent algorithm for model estimation is summarized in Algorithm 1.

Algorithm 1 Coordinate descent algorithm for minimizing (5).

- 1: **Input:** Standardized symmetric adjacency matrices $\{W_i^{(s)}\}$, standardized age $\{\tilde{g}_{is}\}$ and squared age $\{\tilde{g}_{is}^2\}$, binary outcomes $\{y_i\}$; the number of components K , overall penalty factor $\delta > 0$, L_1 fractional penalty factor $\eta \in [0, 1]$, tolerance $\epsilon > 0$.
 - 2: **Output:** Estimates of $\{(\alpha_h, \rho_h, \gamma_h, \beta_h) : h = 1, \dots, K\}$ and α_0 .
 - 3: **Initialize:** each parameter of $\{(\alpha_h, \rho_h, \gamma_h, \beta_h) : h = 1, \dots, K\} \sim U(-0.1, 0.1)$; α_0 is initialized at 0.
 - 4: **repeat**
 - 5: **for** $h = 1 : K$ **do**
 - 6: **for** $u = 1 : V$ **do**
 - 7: update β_{hu} by (A.11)
 - 8: **end for**
 - 9: **end for**
 - 10: **for** $h = 1 : K$ **do**
 - 11: update α_h by (A.15)
 - 12: update ρ_h by (A.16)
 - 13: update γ_h by (A.18)
 - 14: **end for**
 - 15: update α_0 by (A.20)
 - 16: **until** relative change of loss function (5) $< \epsilon$.
-

Since the loss function (5) is lower bounded by 0 and each update always decreases the function value, Algorithm 1 is guaranteed to converge. The estimates for $\{(\alpha_h, \rho_h, \gamma_h, \beta_h) : h = 1, \dots, K\}$ will all be zero under a sufficiently large penalty factor δ , but they cannot be initialized at zero because the results will get stuck at zero. In practice, we recommend to initialize all the parameters as nonzero (except α_0) in case some components get degenerated unexpectedly at the beginning. Each parameter excluding α_0 is randomly initialized from the uniform distribution $U(-0.1, 0.1)$ in Algorithm 1. In general, Algorithm 1 should be run from multiple initializations to locate a good local solution. We use five initializations for Algorithm 1 in simulations and applications.

2.3 | Model selection

The penalty factors δ and η in regularization (4) can be tuned by cross validation (CV) for a fixed K . We use the “one-standard-error” rule (Hastie et al., 2009) to select the optimal pair (δ, η) , which corresponds to the most parsimonious model whose mean cross-validated deviance is within one standard-error of the minimum. Cross validation can be used to select K as well. After the optimal solutions for a sequence of K values are obtained, the best model is the one with the smallest mean CV deviance.

3 | SIMULATION STUDY

We conduct simulations to evaluate the performance of LNCM in recovering true signal subgraphs and age effects as well as prediction. Algorithm 1 of LNCM is implemented in Matlab (R2018a) and the code is publicly available at <https://github.com/wangronglu/LNCM>.

We compare the results of LNCM to the following unstructured logistic regression with the elastic-net penalty:

$$\text{logit}(p_i) = \alpha_0 + \langle B_1, \frac{1}{T_i} \sum_{s=1}^{T_i} W_i^{(s)} \rangle + \langle B_2, \frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is} W_i^{(s)} \rangle + \langle B_3, \frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is}^2 W_i^{(s)} \rangle. \quad (7)$$

where B_1 , B_2 and B_3 are $V \times V$ symmetric coefficient matrices with entrywise elastic-net penalty on them. Edges corresponding to the nonzero entries in B_1 , B_2 and B_3 have constant, linear and quadratic age effects, respectively. In fact, only the upper triangular entries of $\frac{1}{T_i} \sum_{s=1}^{T_i} W_i^{(s)}$, $\frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is} W_i^{(s)}$ and $\frac{1}{T_i} \sum_{s=1}^{T_i} \tilde{g}_{is}^2 W_i^{(s)}$ are entered in the regression. We write (7) in the matrix dot product form for the convenience of displaying results. This method is fitted with the `glmnet` toolbox in Matlab (https://www.stanford.edu/~hastie/glmnet_matlab). The penalty factors (λ, α) of `glmnet` are also tuned by cross validation and selected by the one-standard-error rule. The set of values for the L_1 fractional penalty factor α is chosen as $\{0.1, 0.2, \dots, 1\}$ and for each α , the overall penalty factor λ is tuned over a sequence of 100 equally spaced values on the logarithmic scale.

We simulate a synthetic dataset of $n = 200$ subjects. The number of network observations for each subject T_i ranges from 1 to 5 randomly and the age varies over a 5-year span with an initial age $g_{i1} \sim U(60, 90)$. Each network has 10 nodes. The first network observation $W_i^{(1)}$ for each subject is generated from a set of basis subgraphs with individual loadings as

$$W_i^{(1)} = \sum_{h=1}^8 a_{ih} \mathbf{q}_h \mathbf{q}_h^\top + \Delta_i \tag{8}$$

where $\mathbf{q}_h \in \{0, 1\}^{10}$ is a random binary vector with $\|\mathbf{q}_h\|_0 = h + 1, h = 1, \dots, 8$. The loadings $\{a_{ih}\}$ in (8) are generated independently from $N(0, 1)$ and Δ_i is a 10×10 symmetric noise matrix with each entry $\Delta_{ij|u,v} \stackrel{iid}{\sim} N(0, 0.1^2), u < v$. This generating process (8) produces dense networks with complex correlation structures. The follow-up adjacency matrix $W_i^{(s)}$ for each subject is generated by adding $N(0, 0.1^2)$ random noise to each edge weight in $W_i^{(s-1)}, s = 2, \dots, T_i$. The generated adjacency matrices $\{W_i^{(s)} : s = 1, \dots, T_i; i = 1, \dots, n\}$ are standardized to have mean 0 and variance 1 for each cell and the diagonals are set to zero.

The binary response y_i is generated from $\text{Bernoulli}(p_i)$ independently with

$$\text{logit}(p_i) = \frac{1}{T_i} \sum_{s=1}^{T_i} \lambda(g_{is}) \mathbf{q}_3^\top W_i^{(s)} \mathbf{q}_3 \tag{9}$$

where \mathbf{q}_3 is defined in (8) and the function $\lambda(g)$ is set as the right plot of Figure 2. The generating process (9) indicates that the true signal subgraph associated with y is a 4-node clique and its predictive effect on the outcome increases with age as displayed in Figure 2.

The input parameters of Algorithm 1 for LNCM are set as follows. The tolerance $\epsilon = 10^{-6}$. The number of components K is tuned over 1, 2, 3. We use fivefold cross validation to tune δ and η under each K , where the L_1 fractional penalty factor $\eta \in \{0.1, 0.2, \dots, 1\}$ and a sequence of 10 equally spaced δ values are chosen on the log scale for each η . The same fivefold cross validation is used for tuning (λ, α) in glmnet.

Figure 3 displays the estimated results of LNCM where the best model is selected at $K = 1$. In this case, LNCM correctly identifies the true signal subgraph and its increasing age effect. Figure 4 shows the estimated results of glmnet, which partially recovers the true signal subgraph, and we have to analyze their age effects edge by edge. Although glmnet selects most true signal edges corresponding to B_1 , these edges do not appear under B_2 or B_3 , indicating that these edges are only estimated to have constant age effects. Compared with the ground truth, glmnet only correctly identifies one signal edge (8–10) with linearly increasing age effect, but falsely selects two non-signal edges with linear age effects and one signal edge (6–8) with a fake quadratic age effect.

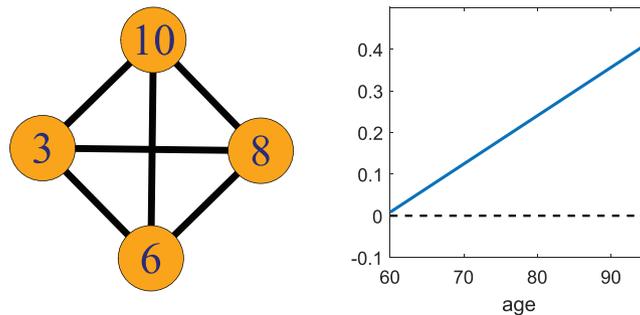


FIGURE 2 True signal subgraph and its age effect $\lambda(g)$

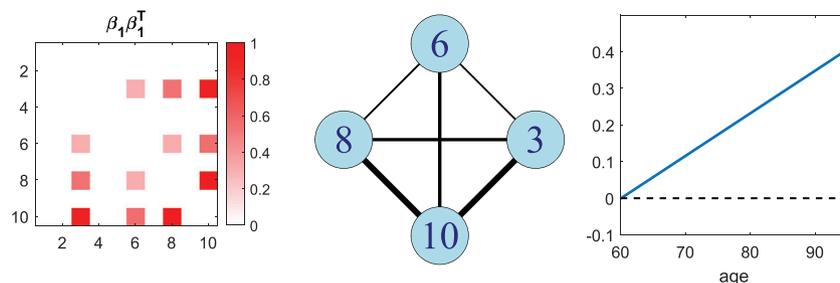


FIGURE 3 Estimated results of LNCM. Left: the estimated nonzero component matrix $\beta_1 \beta_1^\top$. Middle: the selected subgraph, where the thickness of each edge is proportional to the magnitude of its estimated λ coefficient in $\beta_1 \beta_1^\top$. Right: the estimated age effect $\lambda_1(g)$

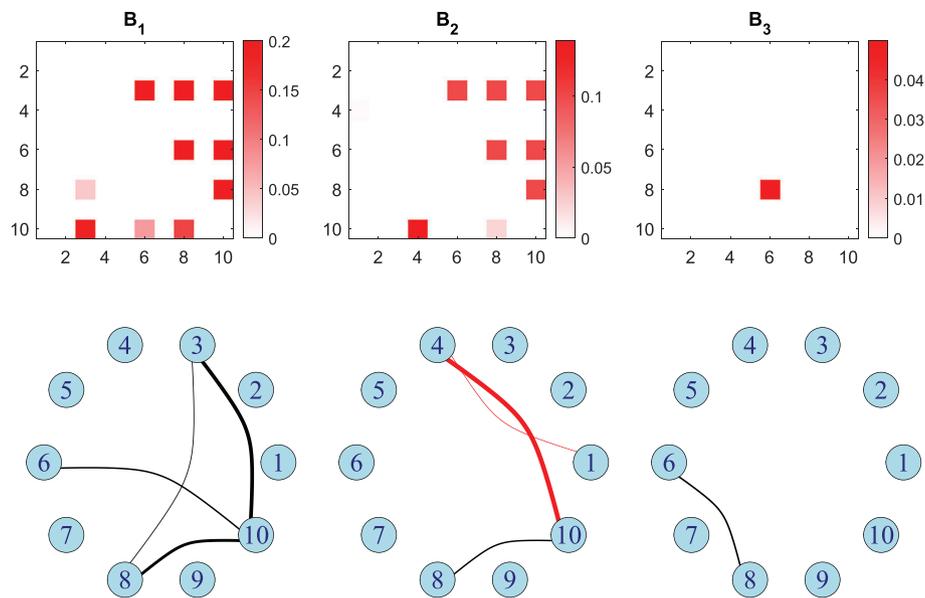


FIGURE 4 Estimated results of glmnet. Upper panel: estimated entries (lower triangular) versus true values (upper triangular) of B_1 , B_2 and B_3 in (7). Lower panel: the selected edges in the network where the thickness of each edge is proportional to the magnitude of its estimated coefficient in B_1 , B_2 or B_3 ; black edges denote true signal edges and red ones falsely identified edges

TABLE 1 Mean and standard deviation of TPR, FPR and mean cross-validated AUC across 100 simulations

	TPR	FPR	AUC
LNCM	0.7233 ± 0.2970	0.0369 ± 0.0715	0.8383 ± 0.0293
glmnet	0.6183 ± 0.1762	0.0431 ± 0.0506	0.8311 ± 0.0310

The procedure above is repeated 100 times. Each time a synthetic dataset is generated based on (8) and (9), and we record the true positive rate (TPR) for LNCM and glmnet, representing the proportion of true signal edges that are correctly identified, and the false positive rate (FPR), denoting the proportion of non-signal edges that are falsely identified. We also record the mean cross-validated AUC at the selected penalty factors for each method as a measure of predictive performance. Table 1 displays the mean and standard deviation of TPR, FPR and the mean cross-validated AUC for LNCM and glmnet across 100 simulations, which shows that LNCM on average achieves higher TPR, lower FPR and competitive predictive performance compared to the unstructured method.

4 | APPLICATION

We applied our method to a dataset from the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) database as described in Section 1 to better understand the neural mechanism underlying successful cognitive aging. The tuning methods for LNCM and glmnet are the same as in Section 3.

The dataset contains dMRI data for 40 supernormals and 45 cognitively normal controls over a 5-year span. Although ADNI2 contains many more subjects, here we focus on healthy subjects who have both longitudinal cognitive measures and brain imaging data in a span of 5 years. Such data are difficult to find from existing data repositories. A state-of-the-art DTI processing pipeline (Zhang, Descoteaux, et al., 2018) was applied to extract structural brain networks of subjects. More specifically, we first used a reproducible probabilistic tractography algorithm (Girard et al., 2014; Maier-Hein et al., 2017) to generate the whole-brain tractography data for each dMRI scan in the dataset. Then the popular Desikan-Killiany atlas (Desikan et al., 2006) was used to define the brain regions of interest (ROIs) in the structural connectivity network. The Desikan-Killiany parcellation has 68 cortical surface regions with 34 nodes in each hemisphere. For each pair of ROIs, the metric—connected surface area (CSA)—is extracted for the reconstructed streamlines as a measure of connectivity strength between two ROIs. Then each weighted adjacency matrix $W_i^{(s)}$ is made up of the CSA of streamlines connecting each pair of brain regions. Figure 5 shows the preprocess of extracting structural brain networks from the dMRI data.

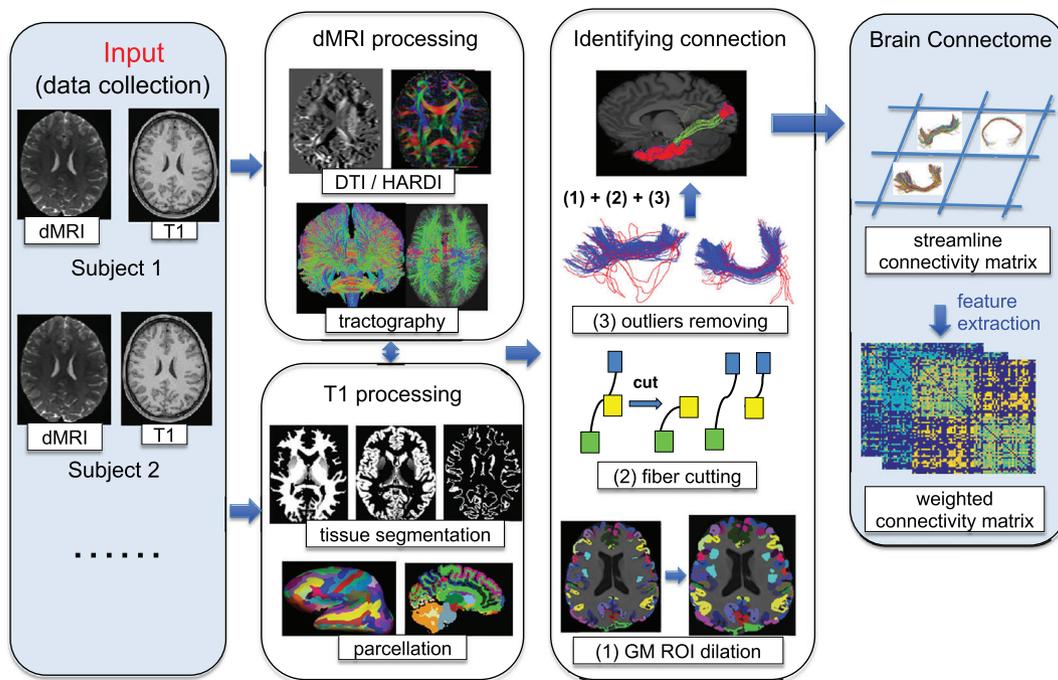


FIGURE 5 Structural brain network extraction pipeline

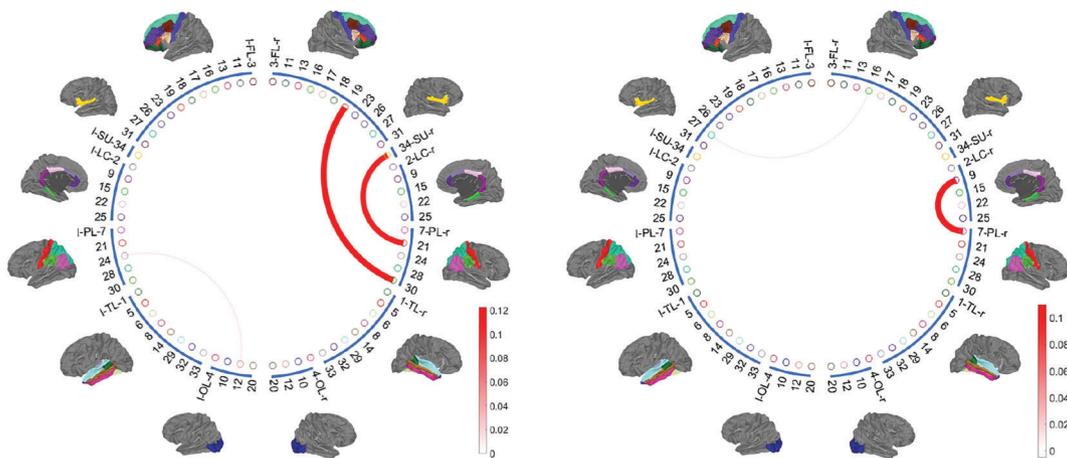


FIGURE 6 Connections selected by glmnet with constant (left) and linear (right) age effects, where the thickness of each edge is proportional to the magnitude of its estimated coefficient in B_1 or B_2 of (7); the colour goes from blue to red as the coefficient goes from negative to positive

Figure 6 displays the connections selected by glmnet with constant (left panel) and linear (right panel) age effects. LNCM identifies one signal subgraph with constant age effect as shown in Figure 7. The structured solution of LNCM has better prediction accuracy with the fivefold cross-validated AUC of 0.7081 ± 0.0576 , while that of glmnet is 0.6588 ± 0.0680 .

The subgraph in Figure 7 features Node 30r as the central node since the coefficients of the edges linking to 30r have larger magnitudes. In the Desikan-Killiany atlas, 30r is the right supramarginal gyrus, which has been found to engage in language processing (Price, 2010), phonological decision-making (Hartwigsen et al., 2010) and verbal working memory (Deschamps et al., 2014). It also appears to be affected by the normal aging process, including volumetric atrophy (Fjell et al., 2009) and decreased functional connectivity with right middle frontal cortex (Wu et al., 2011).

Figure 7 implies that older adults with stronger white matter connections among 30r (right supramarginal), 17r (right pars opercularis), 19r (right pars triangularis), 26r (right rostral middle frontal), 34r (right insula) and 21r (right postcentral) are more likely to be supernormals, while the connection strengths between 32r (right temporal pole) and the other regions may have very slight negative effects. The right pars triangularis (19r) and the right pars opercularis (17r) have emerged as important to various aspects of language and motor function (Molnar-Szakacs et al., 2004). The right rostral middle frontal gyrus (26r) is critical for executive function (Quan et al., 2013). The right insula (34r) is involved in emotion

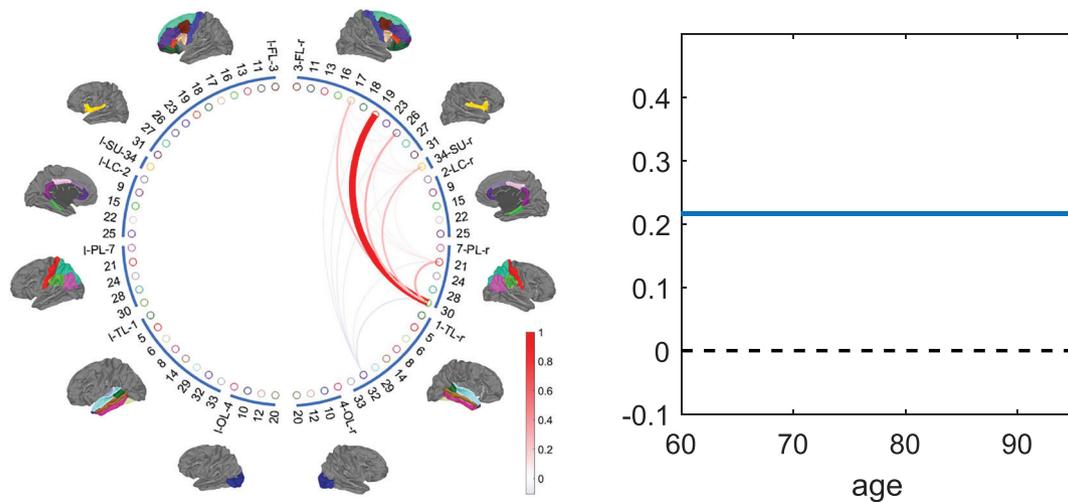


FIGURE 7 The signal subgraph (left) selected by LNCM and its estimated age effect (right), where the thickness of each edge is proportional to the magnitude of its estimated coefficient in $\beta_1\beta_1^T$, and the color goes from blue to red as the coefficient goes from negative to positive

regularization and sensor-motor processing (Chang et al., 2013). The right postcentral gyrus (21r) constitutes the somatosensory cortex (Dharani, 2015). The identified subgraph in Figure 7 represents a high level integration subnetwork of various brain functions from language, motor, emotion to sensation, which is critical to maintain cognition. The constant age effect implies that there is no specific age between 60 and 95 that has particularly large predictive effect on the outcome (supernormal or normal aging).

5 | CONCLUSION AND DISCUSSION

We have proposed a useful model for learning signal subgraphs and their age effects in classification of longitudinal networks. An effective algorithm was developed for model estimation which outputs a set of small clique subgraphs and a smooth function of age for each subgraph indicating its dynamic predictive effect on the outcome. Our method contributes to finding small sets of nodes in the network whose interconnections are associated with the outcome and avoids evaluating age effects edge by edge. Application of this method to a longitudinal dataset of ADNI shows interesting discovery of a small set of brain regions whose connectivity strengths are predictive of successful cognitive aging, which has more appealing interpretation and better predictive performance than unstructured classification methods.

There are also limitations to the proposed method. So far we have focused on a simple parametric form for the age effect function as a tradeoff between interpretation and model flexibility. A future direction would be to generalize the approach to accommodate more complex functions for age effects, for example nonparametric or piecewise constant functions. Moreover, in some applications, our model assumption might not hold, for example, the signal graphs might not be clique subgraphs. In this case, the identified clique subgraphs from our model may have overlapping connections with opposite signs and the out-of-sample predictive performance may be worse than that of an unstructured classifier (e.g. glmnet). In practice, we encourage the user to take into account the identified subgraphs and the out-of-sample predictive performance to decide whether the clique assumption is reasonable. For example, in our real data analysis, only one clique subgraph is selected where K is tuned over $\{1, 2, 3\}$ and our method achieves higher cross-validated AUC than glmnet does. This identified signal subgraph leads to the discovery of a small set of brain regions whose interconnections are associated with the outcome, which provides a more insightful understanding of the mechanism underlying supernormals.

ACKNOWLEDGEMENTS

We thank Dr. Feng Lin for providing us the supernormal data. The research of Lu Wang was supported by the National Natural Science Foundation of China (grant number 11901583). Zhengwu Zhang was supported by the National Institutes of Health (award number MH118927) and Roberta K. Courtman Revocable Trust.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

ORCID

Lu Wang  <https://orcid.org/0000-0001-7472-7008>

REFERENCES

- Arroyo Reli3n, J. D., Kessler, D., Levina, E., & Taylor, S. F. (2019). Network classification with applications to brain connectomics. *The Annals of Applied Statistics*, 13(3), 1648–1677.
- Casey, B., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C., Wager, T., Banich, M., Speer, N., Sutherland, M., Riedel, M., Dick, A., Bjork, J., Thomas, K., Chaarani, B., Mejia, M. H., Hagler, D., Daniela Cornejo, M., Scat, C. S., Harms, M., Dosenbach, N., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J., Kuperman, J., Fair, D., & Dale, A. (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32, 43–54.
- Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex*, 23(3), 739–749.
- Deschamps, I., Baum, S. R., & Gracco, V. L. (2014). On the role of the supramarginal gyrus in phonological processing and verbal working memory: Evidence from rTMS studies. *Neuropsychologia*, 53, 39–46.
- Desikan, R. S., S3gonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980.
- Dharani, K. (2015). *The biology of thought*. San Diego: Academic Press.
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.
- Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., & Dale, A. M. (2009). One-year brain atrophy evident in healthy aging. *Journal of Neuroscience*, 29(48), 15223–15231.
- Girard, G., Whittingstall, K., Deriche, R., & Descoteaux, M. (2014). Towards quantitative connectivity analysis: Reducing tractography biases. *NeuroImage*, 98, 266–278.
- Guha, S., & Rodriguez, A. (2020). Bayesian regression with undirected network predictors with an application to brain connectome data. *Journal of the American Statistical Association*, 116, 581–593.
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences*, 107(38), 16494–16499.
- Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., . . . , & Weiner, M. W. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685–691.
- Lin, F., Ren, P., Mapstone, M., Meyers, S. P., Porsteinsson, A., Baran, T. M., & Initiative, A. D. N. (2017). The cingulate cortex of older adults with excellent memory capacity. *Cortex*, 86, 83–92.
- Lin, F., Wang, X., Wu, R., Rebok, G. W., Chapman, B. P., & Initiative, A. D. N. (2017). Identification of successful cognitive aging in the Alzheimer's disease neuroimaging initiative study. *Journal of Alzheimer's Disease*, 59(1), 101–111.
- Maier-Hein, K. H., Neher, P. F., Houde, J.-C., C3t3, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., He, R., Li, Q., . . . , & Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1), 1349.
- Molnar-Szakacs, I., Iacoboni, M., Koski, L., & Mazziotta, J. C. (2004). Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. *Cerebral Cortex*, 15(7), 986–994.
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191(1), 62–88.
- Quan, M., Lee, S.-H., Kubicki, M., Kikinis, Z., Rathi, Y., Seidman, L. J., Mesholam-Gately, R. I., Goldstein, J. M., McCarley, R. W., Shenton, M. E., & Levitt, J. J. (2013). White matter tract abnormalities between rostral middle frontal gyrus, inferior frontal gyrus and striatum in first-episode schizophrenia. *Schizophrenia Research*, 145(1-3), 1–10.
- Sun, W. W., & Li, L. (2019). Dynamic tensor clustering. *Journal of the American Statistical Association*, 114(528), 1894–1907.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 58(1), 267–288.
- Wang, L., Lin, F. V., Cole, M., & Zhang, Z. (2021). Learning clique subgraphs in structural brain network classification with application to crystallized cognition. *NeuroImage*, 225, 117493.
- Wang, L., Zhang, Z., & Dunson, D. B. (2019). Symmetric bilinear regression for signal subgraph estimation. *IEEE Transactions on Signal Processing*, 67, 1929–1940.
- Wu, J.-T., Wu, H.-Z., Yan, C.-G., Chen, W.-X., Zhang, H.-Y., He, Y., & Yang, H.-S. (2011). Aging-related changes in the default mode network and its anti-correlated networks: A resting-state fMRI study. *Neuroscience Letters*, 504(1), 62–67.
- Zhang, J., Sun, W. W., & Li, L. (2018). Network response regression for modeling population of networks with covariates. arXiv: Methodology.
- Zhang, J., Sun, W. W., & Li, L. (2019). Mixed-effect time-varying network model and application in brain connectivity analysis. *Journal of the American Statistical Association*, 115, 2022–2036.
- Zhang, X., Li, L., Zhou, H., Zhou, Y., & Shen, D. (2019). Tensor generalized estimating equations for longitudinal imaging analysis. *Statistica Sinica*, 29, 1977–2005.
- Zhang, Z., Descoteaux, M., Zhang, J., Girard, G., Chamberland, M., Dunson, D., Srivastava, A., & Zhu, H. (2018). Mapping population-based structural connectomes. *NeuroImage*, 172, 130–145.

- Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), 540–552.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, L., & Zhang, Z. (2021). Classification of longitudinal brain networks with an application to understanding superior aging. *Stat*, 10(1), e402. <https://doi.org/10.1002/sta4.402>