

Learning Clique Subgraphs in Structural Brain Network Classification with Application to Crystallized Cognition

Lu Wang^{a,*}, Feng Vankee Lin^{b,c,d,e,f}, Martin Cole^g, Zhengwu Zhang^{e,g,*}

^a Department of Statistics, Central South University, China

^b Elaine C. Hubbard Center for Nursing Research On Aging, School of Nursing, University of Rochester Medical Center, USA

^c Department of Psychiatry, School of Medicine and Dentistry, University of Rochester Medical Center, USA

^d Department of Brain and Cognitive Sciences, University of Rochester, USA

^e Department of Neuroscience, University of Rochester Medical Center, USA

^f Department of Neurology, School of Medicine and Dentistry, University of Rochester Medical Center, USA

^g Department of Biostatistics and Computational Biology, University of Rochester Medical Center, USA

ARTICLE INFO

Keywords:

Network classification
Signal subgraph learning
Clique subgraphs
Structural brain networks
Symmetric bilinear logistic regression

ABSTRACT

Structural brain networks constructed from diffusion MRI are important biomarkers for understanding human brain structure and its relation to cognitive functioning. There is increasing interest in learning differences in structural brain networks between groups of subjects in neuroimaging studies, leading to a variable selection problem in network classification. Traditional methods often use independent edgewise tests or unstructured generalized linear model (GLM) with regularization on vectorized networks to select edges distinguishing the groups, which ignore the network structure and make the results hard to interpret. In this paper, we develop a symmetric bilinear logistic regression (SBLR) with elastic-net penalty to identify a set of small clique subgraphs in network classification. Clique subgraphs, consisting of all the interconnections among a subset of brain regions, have appealing neurological interpretations as they may correspond to some anatomical circuits in the brain related to the outcome. We apply this method to study differences in the structural connectome between adolescents with high and low crystallized cognitive ability, using the crystallized cognition composite score, picture vocabulary and oral reading recognition tests from NIH Toolbox. A few clique subgraphs containing several small sets of brain regions are identified between different levels of functioning, indicating their importance in crystallized cognition.

1. Introduction

Recent advances in magnetic resonance imaging (MRI) techniques enable us to noninvasively probe the human brain at higher resolutions than ever before (Glasser et al., 2016) and reconstruct connectomes with distinct physiological meanings (Park and Friston, 2013). Among them, the diffusion MRI (dMRI) infers the locations and directions of white matter (WM) fiber tracts via measuring the water molecular movement along major fiber bundles in WM. Diffusion MRI data are now collected in almost all major cohort-based neuroimaging studies, e.g., the Human Connectome Project (Van Essen et al., 2013), the UK Biobank (Miller et al., 2016) and the Adolescent Brain Cognitive Development (ABCD) Study (Casey et al., 2018). Structural connectivity (SC) analysis is among the most important applications of dMRI (Park and Friston, 2013; Yeh et al., 2016; Zhang et al., 2018; Zhao et al., 2015), where individual-level microstructural brain networks are constructed to delineate anatomical connections between brain regions. Fig. 1 il-

lustrates the pipeline we used for extracting SC (Zhang et al., 2018) (Fig. 1a) and an SC matrix extracted from one subject in the ABCD data (Fig. 1b).

Brain network classification and identification of predictive subnetworks are probably among the most important applications of SC into the mechanistic understanding of neuroscience phenomena. One typical approach to this network classification and variable selection problem is to treat all the connections of SC as a long feature vector, and apply existing classification methods for vectors, such as generalized linear models with L_1 or elastic-net penalty (Zou and Hastie, 2005), support vector machines (Zhu et al., 2004) etc. Another popular approach in the neuroscience literature is to perform massive univariate tests at each edge with multiple testing correction to identify edges that are significantly different between two groups. While these methods may have good prediction, they ignore the network nature of the data and do not guarantee any structure among the selected individual edges, making the results hard to interpret.

* Corresponding authors.

E-mail addresses: wanglu_stat@csu.edu.cn (L. Wang), Zhengwu_Zhang@urmc.rochester.edu (Z. Zhang).

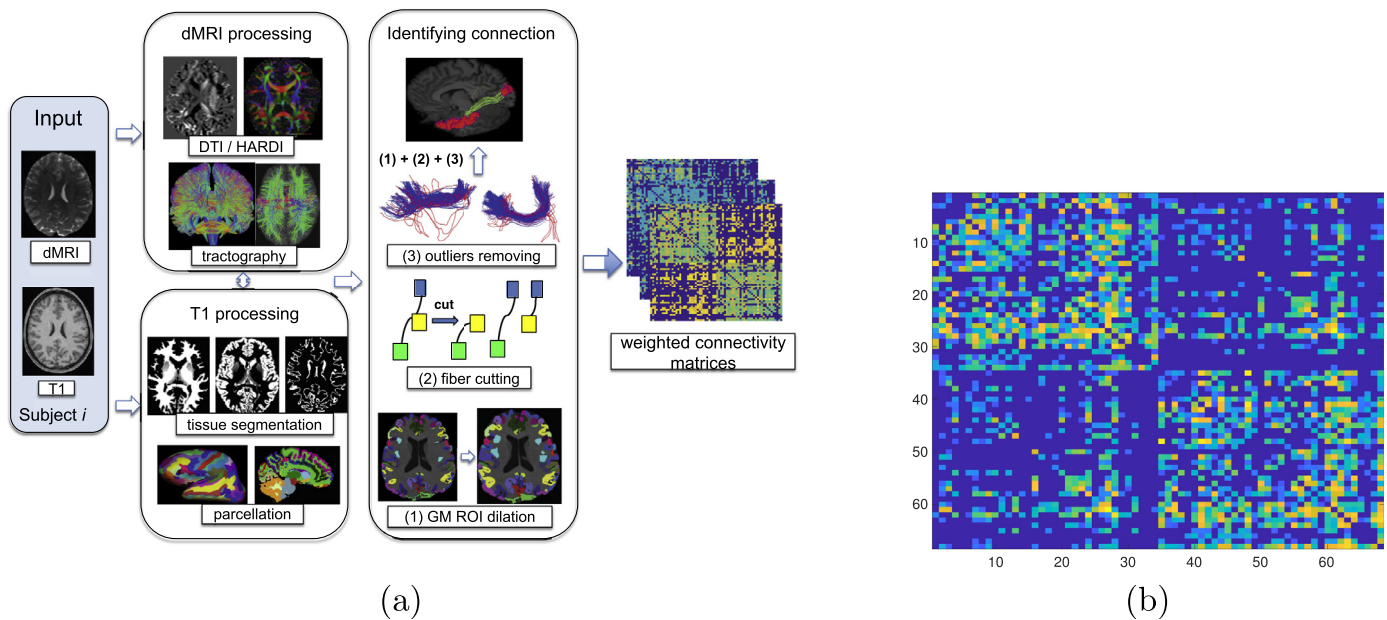


Fig. 1. (a) shows the pipeline we used to extract the network and (b) shows a structural brain network extracted from the ABCD data.

Other feature extraction approaches often employ a two-step procedure, where some unsupervised dimension reduction is first applied on the networks, and then a regression model is fitted on the extracted low-dimensional features. Zhang et al. (2019) proposed to use tensor decomposition approach to map networks to low-dimensional vectors. Simpson et al. (2011) proposed to reduce the networks to some topological summary measures such as clustering coefficient, network density, etc. However, such connectome simplification leads to an enormous loss of information and brings troubles in truly understanding which part of the brain network is responsible for the group difference.

There are a few advanced statistical methods considering the network structure when analyzing group differences in connectome. Vogelstein et al. (2012) proposed to search for a minimum set of vertices and edges distinguishing groups, but this method only applies to binary edges and involves solving a combinatorial optimization problem. Durante et al. (2018) developed a method for testing global and local changes in SC between groups, which gains power by accounting for dependence structure among edges through a Bayesian nonparametric modeling of the networks. Arroyo Reli6n et al. (2019) proposed a graph classification method that uses edge weights as predictors and incorporates L_1 and group lasso penalty to penalize both the number of edges and the number of nodes selected.

We developed a symmetric bilinear logistic regression (SBLR) model with elastic-net penalty to select a set of small clique signal subgraphs in network classification. A clique subgraph in graph theory refers to a subset of nodes of an undirected graph such that every two different nodes in the clique are connected (Seidman and Foster, 1978). SBLR puts symmetry constraint on the coefficient matrix in the logistic regression, because the adjacency matrices of structural brain networks are symmetric. The novelties and significance of our approach reside on:

- The small clique subgraphs identified by our method have appealing interpretations in neuroscience field. Clique subgraphs (containing all the interconnections among a subset of nodes) potentially aligns with the physiological meaning of subgraphs that should be strongly interrelated in order to provide the most efficient neural support of a behavior (Bassett and Sporns, 2017). Despite the clique structure imposed on the selected subgraphs, our model maintains the flexibility of identifying individual edges in network classification and good classification rate.

- The elastic-net penalty penalizes the size of each identified clique subgraph, producing a parsimonious representation of differences in brain connectome. We develop a coordinate descent algorithm for model estimation where analytical solutions are derived for a sequence of conditional convex optimizations. The code for implementing the algorithm is publicly available at <https://github.com/wangronglu/SBLR>.
- The SBLR approach is applied to examine structural network classification for crystallized intelligence among 4213 right-handed adolescents aged 9-10 years in the ABCD study. Emerging literature suggests unique roles of white matter in supporting general cognition and differentiating crystallized and fluid intelligence (G6ngora et al., 2020; Penke et al., 2012; Simpson-Kent et al., 2020). Of note, age 9-10 represents a critical change point for the relationship between white matter and crystallized intelligence (Simpson-Kent et al., 2020). Thus far, however, no study has specified the role of SC in crystallized intelligence of adolescents. We extensively analyzed SCs of subjects in ABCD with high vs. low crystallized intelligence using our SBLR model, aiming at learning interpretable SC-based brain connectome maps that can differentiate the levels of crystallized intelligence.

The rest of the paper is organized as follows. Section 2 introduces the data preprocessing steps, our method and the algorithm for model estimation. Section 3 presents simulation studies demonstrating good performance of SBLR in recovering true clique signal subgraphs compared with other methods. Application of this method to ABCD data in Section 4 shows coherent subgraphs of crystallized intelligence across composite score and individual domains. We conclude with a brief discussion in Section 5.

2. Methods

2.1. ABCD data preprocessing

We focus on the ABCD dataset downloaded from NIH Data Archive (<https://nda.nih.gov>) (Casey et al., 2018). The main goal of the ABCD study is to track the brain development from childhood through adolescence to understand biological and environmental factors that can affect the brain's developmental trajectory. ABCD recruits approximately 10,000 9 – 10 years old children. Longitudinal measures of the brain

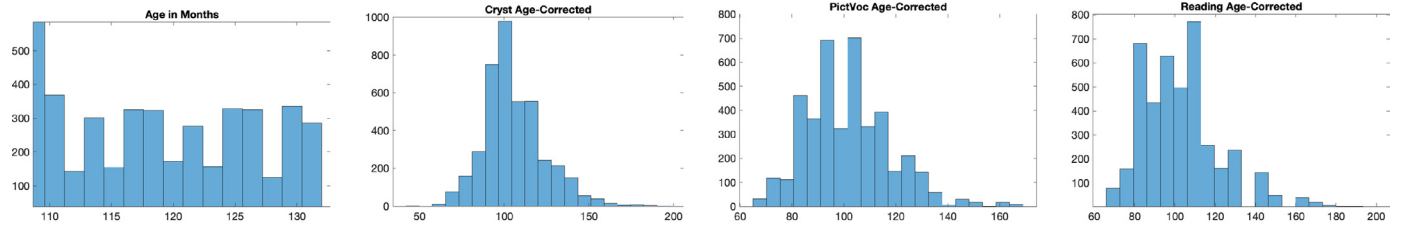


Fig. 2. Histograms of ages, age-corrected crystallized cognition composite scores, picture vocabulary scores and reading scores of the 4213 subjects involved in this study.

structure and function as well as rich behavior measures and genetic factors are collected across 21 sites in the United States (Auchter et al., 2018).

Imaging data and preprocessing: The ABCD imaging protocol is harmonized for three types of 3T scanners: Siemens Prisma, General Electric (GE) 750 and Philips. We downloaded the structural T1 MRI and diffusion MRI (dMRI) data for 5253 subjects from the ABCD 2.0 release. The structural T1 image was acquired with isotropic resolution of 1 mm³. The diffusion MRI image was obtained based on imaging parameters: 1.7 mm³ resolution, four different b-values ($b = 500, 1000, 2000, 3000$) and 96 diffusion directions. There are 6 directions at $b = 500$, 15 directions at $b = 1000$, 15 directions at $b = 2000$, 60 directions at $b = 3000$. Multiband factor 3 is used for dMRI acceleration. See Casey et al. (2018) for more details about the data acquisition and preprocessing of the ABCD data.

To obtain structural connectome, we used a state-of-the-art dMRI data preprocessing framework named population-based structural connectome (PSC) mapping (Zhang et al., 2018). PSC uses a reproducible probabilistic tractography algorithm (Girard et al., 2014; Maier-Hein et al., 2017) to generate the whole-brain tractography. This tractography method borrows anatomical information from high-resolution T1-weighted imaging to reduce bias in reconstruction of tractography. On average, about 10⁶ streamlines were generated for each subject. We used the popular Desikan–Killiany atlas (Desikan et al., 2006) to define the brain regions of interest (ROIs) corresponding to the nodes in the structural connectivity network. The Desikan–Killiany parcellation has 68 cortical surface regions with 34 nodes in each hemisphere. Freesurfer software (Dale et al., 1999; Fischl et al., 2004) is used to perform brain registration and parcellation. Next, for each pair of ROIs, we extracted the streamlines connecting them. In this process several procedures were used to increase the reproducibility: (1) each gray matter ROI is dilated to include a small portion of white matter region, (2) streamlines connecting multiple ROIs are cut into pieces so that we can extract the correct and complete pathway and (3) apparent outlier streamlines are removed. Extensive experiments have illustrated that these procedures can significantly improve the reproducibility of the extracted weighted networks, and readers can refer to Zhang et al. (2018) for more details. To analyze the brain as a network, a scalar number is usually extracted to summarize each connection. Here we use fiber count, but other measures, such as mean fractional anisotropy or connected surface area (Zhang et al., 2018) can be easily extracted using PSC.

Applying PSC, we processed 5253 subjects downloaded from NIH Data Archive. To control for handedness, we only focused on the 4213 right-handed subjects in our analysis. The distributions of the age, crystallized cognition composite score, picture vocabulary score and reading score are shown in Fig. 2. For each composite or domain-specific crystallized intelligence score, subsets of subjects with age-adjusted scale scores 1 standard deviation above or below the national average are categorized into high vs. low crystallized intelligence group.

2.2. Symmetric bilinear logistic regression (SBLR)

Our data can be summarized as $\{(y_i, \mathbf{x}_i, W_i) : i = 1, \dots, n\}$, where y_i is a binary response; $\mathbf{x}_i \in \mathbb{R}^m$ contains the regular covariates of subject i

(age, gender, etc.) with the first entry being 1; W_i denotes the structural brain network measured for subject i , which is a $V \times V$ symmetric matrix with zero diagonal entries. Our goal is to learn a set of small clique subgraphs from the brain network that are relevant to the outcome.

With this aim in mind, we propose the following symmetric bilinear logistic regression (SBLR):

$$y_i | W_i, \mathbf{x}_i \stackrel{\text{ind}}{\sim} \text{Bern}(p_i), i = 1, \dots, n, \\ \text{logit}(p_i) = \mathbf{x}_i^\top \alpha + \sum_{h=1}^K \lambda_h \beta_h^\top W_i \beta_h \quad (1)$$

where $\alpha \in \mathbb{R}^m$ with the first entry being the intercept; $\beta_h \in \mathbb{R}^V$ and $\lambda_h \in \mathbb{R}$, $h = 1, \dots, K$. We do not restrict the component vectors β_h 's to be orthogonal because such constraint may discourage the sparsity of these vectors. Model (1) assumes that the binary outcome y_i of each individual follows an independent Bernoulli distribution given the network observation W_i and other covariates \mathbf{x}_i .

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

$$\text{logit}(p_i) = \mathbf{x}_i^\top \alpha + \sum_{h=1}^K \langle \lambda_h \beta_h \beta_h^\top, W_i \rangle \quad (2)$$

where $\langle B, W \rangle = \text{trace}(B^\top W) = \text{vec}(B)^\top \text{vec}(W)$. The parameters λ_h 's in (2) are necessary to avoid constraining the coefficient matrix of W_i to be positive semi-definite. From (2), we can see that the nonzero entries in each component matrix $\lambda_h \beta_h \beta_h^\top$ locate an outcome-relevant clique subgraph in the network predictor. If β_h only contains two nonzero entries, the corresponding subgraph comes down to a single edge.

To ensure both identifiability of model (2) and sparsity of coefficient matrices $\{\lambda_h \beta_h \beta_h^\top : h = 1, \dots, K\}$, we penalize the magnitude of the lower-triangular entries in these coefficient matrices with the following elastic-net penalty:

$$\sum_{h=1}^K \sum_{u=1}^V \sum_{v < u} \delta [\eta |\lambda_h| |\beta_{hu}| |\beta_{hv}| + (1 - \eta) \lambda_h^2 \beta_{hu}^2 \beta_{hv}^2 / 2] \quad (3)$$

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

2.3. Estimation

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

$$\text{Loss function} = -\frac{1}{n} \sum_{i=1}^n \ell_i + \sum_{h=1}^K \sum_{u=1}^V \sum_{v < u} \delta [\eta |\lambda_h| |\beta_{hu}| |\beta_{hv}| + (1 - \eta) \lambda_h^2 \beta_{hu}^2 \beta_{hv}^2 / 2] \quad (4)$$

where ℓ_i is the log-likelihood of subject i . Plugging in the logit link of (1), we have

$$\begin{aligned} \ell_i &= y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \\ &= y_i \left(\mathbf{x}_i^\top \alpha + \sum_{h=1}^K \lambda_h \beta_h^\top W_i \beta_h \right) + \log(1 - p_i) \end{aligned} \quad (5)$$

where \tilde{W}_i represents the normalized network observation with mean 0 and variance 1 for each edge.

The algorithm of block updating each component vector β_h in tensor regression (Zhou et al., 2013) is not applicable to minimizing the loss function (4), because l_i is not a concave function of β_h when fixing the other parameters. Notice that

$$\frac{\partial^2 l_i}{\partial \beta_h \partial \beta_h^\top} = 2(y_i - p_i)\lambda_h \tilde{W}_i - 4p_i(1 - p_i)\lambda_h^2 \tilde{W}_i \beta_h \beta_h^\top \tilde{W}_i,$$

which may not be negative semi-definite. However, since the diagonal entries of each adjacency matrix \tilde{W}_i are zero, the loss function (4) is actually a convex function of each entry β_{hu} in β_h when fixing the others, making the coordinate descent algorithm very appealing (Wang et al., 2019). The technical details of deriving the analytic form update for each parameter are discussed in Appendix A. The coordinate descent algorithm for minimizing (4) is summarized in Algorithm 1, where all

Algorithm 1 Coordinate descent for SBLR model with elastic-net penalty.

```

1: Input: Normalized  $V \times V$  symmetric matrix observations  $\{\tilde{W}_i : i = 1, \dots, n\}$ ,  $n \times m$  design matrix  $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ , binary outcome  $\{y_i : i = 1, \dots, n\}$ ; rank  $K$ , overall penalty factor  $\delta > 0$ ,  $L_1$  fractional penalty factor  $\eta \in [0, 1]$ , tolerance  $\epsilon > 0$ .
2: Output: Estimates of  $\alpha$ ,  $\{(\lambda_h, \beta_h) : h = 1, \dots, K\}$ .
3: Initialize:  $\alpha = 0$  and each parameter of  $\{(\lambda_h, \beta_h) : h = 1, \dots, K\} \sim U(-0.1, 0.1)$ .
4: repeat
5:   for  $h = 1 : K$  do
6:     for  $u = 1 : V$  do
7:       update  $\beta_{hu}$  by (A.11)
8:     end for
9:   end for
10:  for  $h = 1 : K$  do
11:    update  $\lambda_h$  by (A.15)
12:  end for
13:  update  $\alpha$  by (A.16)
14: until relative change of loss function (??)  $< \epsilon$ .
```

the parameters are iteratively updated until the relative change of the loss function (4) is smaller than a tolerance number ϵ . Typical value for ϵ is 10^{-5} or 10^{-6} . Since the loss function (4) is lower bounded by 0 and each update always decreases the function value, Algorithm 1 is guaranteed to converge.

In general, the algorithm should be run from multiple initializations to locate a good local solution. Although the estimates for $\{(\lambda_h, \beta_h) : h = 1, \dots, K\}$ will all be zero under sufficiently large penalty factor δ , we cannot initialize them at zero because the results will then get stuck at zero. The updating rules (A.11) and (A.15) imply that each parameter will be set to 0 given others being zero. In fact, we recommend to initialize all the parameters in $\{(\lambda_h, \beta_h) : h = 1, \dots, K\}$ to be nonzero in case some components get degenerated unexpectedly at the beginning. In practice, we initialize each parameter in $\{(\lambda_h, \beta_h) : h = 1, \dots, K\}$ from the uniform distribution $U(-0.1, 0.1)$ as discussed at the end of Section A.1.

2.4. Model selection

The penalty factor δ in the regularization (3) can be tuned on a validation set over a grid of values on $[\delta_{\min}, \delta_{\max}]$ for a fixed η , where δ_{\max} is a roughly smallest value for which all the parameters $\{(\lambda_h, \beta_h)\}_{h=1}^K$ become zero, and $\delta_{\min} = \epsilon \delta_{\max}$ (e.g. $\epsilon = 0.01$). The optimal δ produces the smallest deviance (minus twice the log-likelihood) on validation set. The fractional parameter of L_1 penalty, $\eta \in [0, 1]$, can also be selected by validation.

Our proposed model (1) assumes a known number of components K . In practice, we choose a large enough number for K and allow the elastic-net penalty (3) to discard unnecessary components, leading to a data-driven estimate for the number of signal subgraphs. We assess the performance of the procedure and verify its lack of sensitivity to the chosen upper bound K in simulations.

3. Simulation study

We use simulations to compare the performance of recovering true signal subgraphs and predictive performance among SBLR and the following methods:

1. Logistic regression with elastic-net penalty on vectorized networks, where the upper triangular entries of each adjacency matrix W_i are entered into a regularized logistic regression. The grid of values for the L_1 fractional penalty factor α is picked as $\{0.1, 0.2, \dots, 1\}$. For each α , the optimal penalty factor λ is chosen from a sequence of 100 equally spaced λ values on the logarithmic scale. This method is fitted with glmnet toolbox in Matlab (http://www.stanford.edu/~hastie/glmnet_matlab).
2. Penalized graph classification (GC) approach (Arroyo Reli3n et al., 2019), which also uses edge weights as predictors, but incorporates L_1 and group lasso penalty to promote sparsity in selection of edges and nodes. Their penalty factor pair (λ, ρ) is tuned over a 5×11 grid, where $\lambda \in \{10^{-4}, 10^{-3}, \dots, 1\} \times \lambda_{\max}$ and $\rho \in \{1, 10, 20, \dots, 100\}$, with $(\lambda_{\max}, \rho = 100)$ ensuring that all the coefficients are penalized to zero. As Arroyo Reli3n et al. (2019) suggest, values of $\rho < 1$ do not result in node selection. This method is fitted with graphclass package in R.
3. Screening method based on multiple testing with false discovery rate control (MT-FDR), where a two-sample t -test is performed on each edge in the network between the two groups. Multiple comparisons are adjusted by rejecting all local nulls having a p -value below the Benjamini-Yekutieli threshold (Benjamini et al., 2001) under arbitrary dependence assumptions on the multiple tests. The false discovery rate is controlled at level $\alpha = 0.05$. A logistic regression is then fitted on the significant edge weights.
4. Tensor network factorization analysis (TNFA) (Zhang et al., 2019), which embeds the $V \times V$ symmetric adjacency matrices $\{W_i : i = 1, \dots, n\}$ into a low dimensional $n \times K$ matrix U ($K \leq V$). Each row i of U contains the K principal component scores of subject i , and each column of U corresponds to a basis network. A logistic regression is then fitted on the low embedding matrix U . The basis networks corresponding to the significant coefficients are selected as signal subnetworks. We use full rank $K = V$ for TNFA in simulations, which explains around 99% variation in the generated networks on average.

In the data generating process, each adjacency matrix W_i is a 20×20 symmetric matrix generated from a set of 15 basis subgraphs with individual loading vectors

$$W_i = \sum_{h=1}^{15} \lambda_{ih} \mathbf{q}_h \mathbf{q}_h^\top + \Delta_i, \quad (6)$$

where $\mathbf{q}_h \in \{0, 1\}^{20}$ is a random binary vector with $\|\mathbf{q}_h\|_0 = h + 1$, $h = 1, 2, \dots, 15$. The loadings $\{\lambda_{ih}\}$ in (6) are generated independently from $U(0, 1)$. Δ_i is a symmetric 20×20 noise matrix that adds 5% noise to each edge in W_i . Specifically, the (u, v) -th entry in Δ_i (denoted as $\Delta_{i[uv]}$) is sampled from a normal distribution with mean zero and standard deviation $0.05 * sd\left(\left(\sum_{h=1}^{15} \lambda_{ih} \mathbf{q}_h \mathbf{q}_h^\top\right)_{[uv]}\right)$. The generating process (6) produces dense networks with complex correlation structure. These generated adjacency matrices $\{W_i : i = 1, \dots, n\}$ are further standardized for each edge across subjects and the diagonals are set to zero.

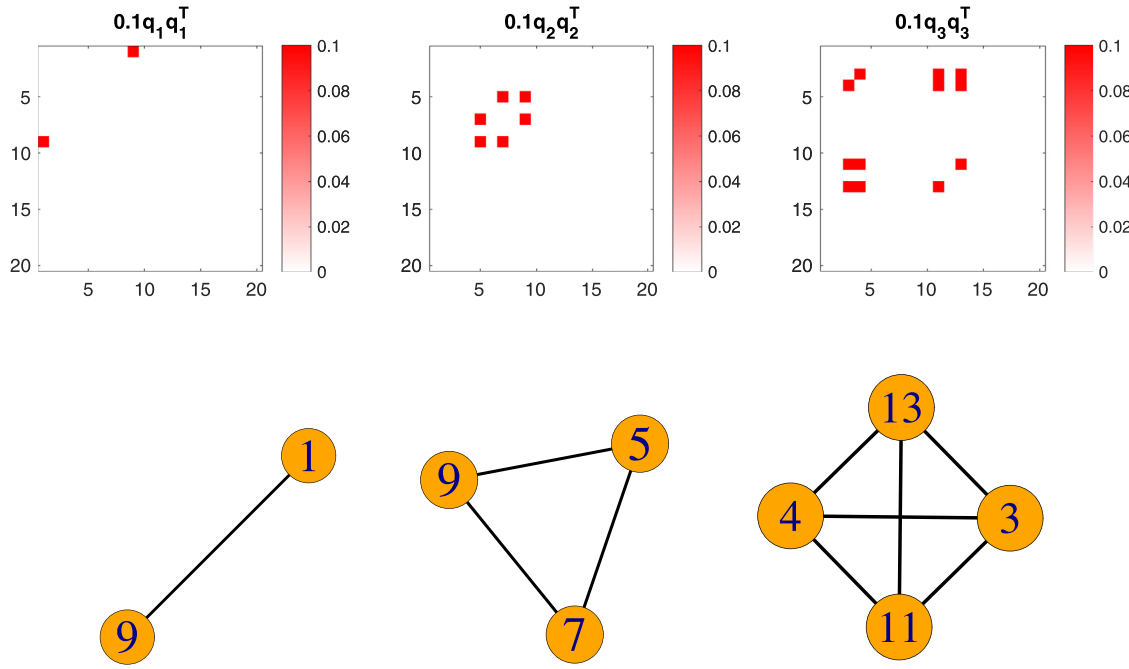


Fig. 3. True signal subgraphs (lower panel) corresponding to $\{0.1q_hq_h^T : h = 1, 2, 3\}$ (upper panel) in one simulation.

3.1. Clique signal subgraphs

In this setting, the binary response y_i is associated with 3 clique signal subgraphs in the network. Specifically, y_i is generated from $\text{Bernoulli}(p_i)$ with

$$\text{logit}(p_i) = 0.1q_1^T W_i q_1 + 0.1q_2^T W_i q_2 + 0.1q_3^T W_i q_3, i = 1, \dots, n. \quad (7)$$

The generating process (7) indicates that the true signal subgraphs relevant to y correspond to the first three basis subgraphs $\{q_hq_h^T : h = 1, 2, 3\}$ as displayed in Fig. 3.

We consider two sample sizes in this simulation: $n = 500$ and $n = 1000$. For each sample size n , a dataset $\{(W_i, y_i) : i = 1, \dots, n\}$ is drawn from the generating process (6),(7). The dataset is then divided into two parts: training set (70%) and validation set (30%). Each method (SBLR, glmnet, GC) is fitted with training set and the optimal penalty pair is selected corresponding to the lowest deviance on validation set. We then refit the model with full dataset under the optimal penalty pair. For each dataset, we additionally generate 200 pairs $\{(W_i, y_i)\}$ as test data and measure the predictive performance of each method by computing the area under an ROC curve (AUC) (DeLong et al., 1988) for test data.

Some input parameters of Algorithm 1 for SBLR model are set as follows. The tolerance $\epsilon = 10^{-6}$. The L_1 fractional penalty factor η is tuned over 5 fixed values $\{0.6, 0.7, \dots, 1\}$. For each value of η , we set $\delta_{\min} = 0.01\delta_{\max}$ and choose a sequence of 11 equally spaced δ values on the logarithmic scale. 5 initializations are used in Algorithm 1. Performances under two different choices of K are compared for SBLR: $K = 5$ and $K = 10$.

The estimated results under optimal penalty pair from *glmnet* are displayed in Fig. 4. As can be seen, the accuracy of *glmnet* improves as the sample size n increases, in terms of selecting more true signal edges and fewer non-signal edges. But it is difficult to identify meaningful structure among the selected edges. Fig. 5 displays the estimated results under optimal penalty pair for the GC approach (Arroyo Reli3n et al., 2019). Although this approach identifies all the true signal edges under larger sample size, it also selects a larger number of false edges. As with *glmnet*, this approach does not guarantee any structure among the selected edges.

Multiple t -test screening method (MT-FDR) deems many more edges to be significant in this case, taking up about 86% and 95% of the total number of edges in the network under $n = 500$ and $n = 1000$, respectively. Tensor factorization approach (TNFA) on the contrary selects no basis networks under each sample size, because none of the 20 components are significant in the logistic regression at the significance level of 0.05.

The estimated results under the optimal penalty pair of SBLR with $K = 5$ and $K = 10$ are displayed in Fig. 6. The performance of SBLR improves with larger sample size. Compared to Fig. 3, Fig. 6 shows that SBLR recovers all the true signal subgraphs under $n = 1000$ with fewer wrong edges compared to $n = 500$. The cliques identified by SBLR may change under different K due to random initialization for component vectors. As can be seen in Fig. 6, the selected subgraphs are different yet with many overlaps under the sample size $n = 500$ between $K = 5$ and $K = 10$. The results become more stable under the larger sample size $n = 1000$, where the node memberships of the selected cliques are identical between $K = 5$ and $K = 10$ with very slight differences in the estimated coefficients.

The procedure above is repeated 100 times. For either sample size n , 100 datasets are generated based on the generating process (6),(7). We record for each method the true positive rate (TPR) representing the proportion of true signal edges that are correctly identified, the false positive rate (FPR) representing the proportion of non-signal edges that are falsely identified, and the false discovery rate (FDR) describing the proportion of selected edges that are false, as well as AUC for test data. Fig. 7 displays the mean and standard deviation (SD) of the TPR, FPR, FDR and AUC for each method across 100 datasets. Note that higher values for TPR and AUC are better, and lower scores of FPR and FDR are better.

Fig. 7 shows that the mean and SD of the four measures are very similar for SBLR under $K = 5$ and $K = 10$, with the same type of penalty (L_1 or elastic-net) and the same sample size. This implies that the performance of SBLR is robust to the chosen upper bound for the rank. In addition, Fig. 7 shows that the mean TPR and AUC improve with larger sample size for all the methods. But with increasing sample size, the mean FPR and FDR of GC, multiple testing (MT-FDR) and tensor factorization (TNFA) increase considerably. The mean FPR of GC and MT-FDR

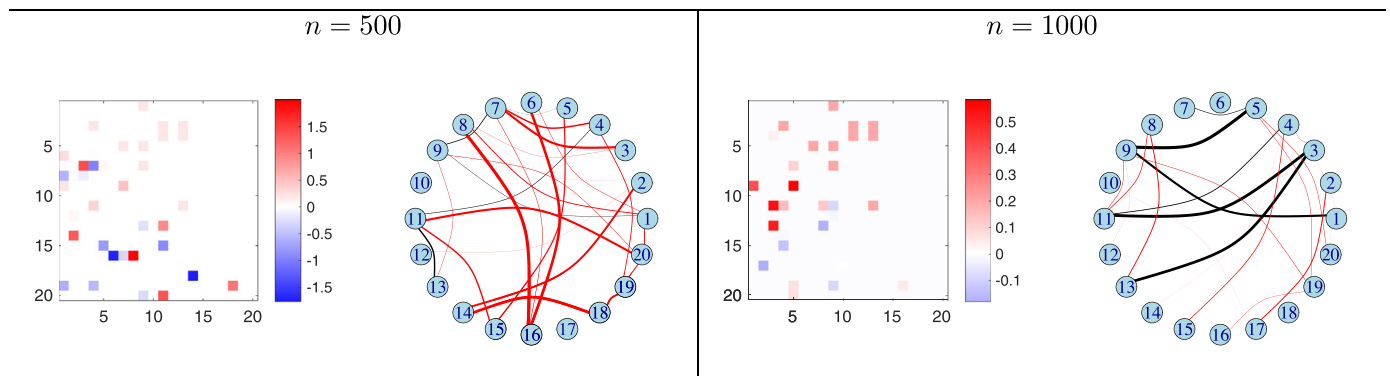


Fig. 4. Estimated results of glmnet under $n = 500$ and $n = 1000$. In each panel, the left plot shows estimated coefficients (lower triangular) versus true values (upper triangular); the right plot shows the selected edges in the network, where black edges denote true signal edges and red ones falsely identified edges; the thickness of each edge is proportional to the magnitude of its estimated coefficient.

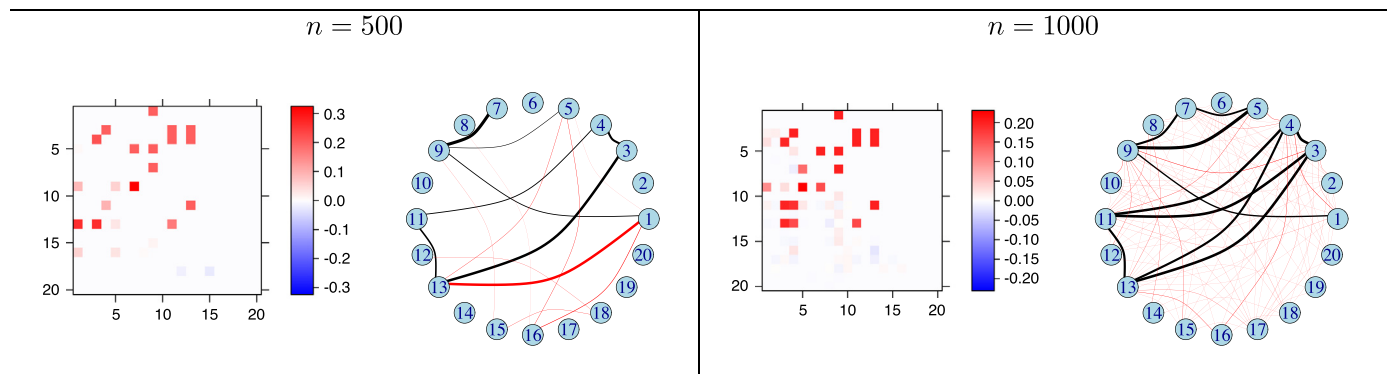


Fig. 5. Estimated results of penalized graph classification (GC) approach under $n = 500$ and $n = 1000$. In each panel, the left plot shows estimated coefficients (lower triangular) versus true values (upper triangular); the right plot shows the selected edges in the network, where black edges denote true signal edges and red ones falsely identified edges; the thickness of each edge is proportional to the magnitude of its estimated coefficient.

are also much higher than that of the other methods. The FDR of all the methods excluding TNFA are around or above 0.5 on average, while the mean FDR of glmlasso, glmnet and SBLR models decrease as the sample size increases. The high FDR is probably due to the complex correlations among edge weights in the simulated networks, which is often the case for structural brain networks. The correlations among edges add difficulty to variable selection as false edges correlated with true signal edges are likely to be selected. It is a challenging problem to further control the FDR here, as the MT-FDR, which is supposed to keep the FDR below 0.05, turns out to have a FDR around 0.9. All the methods except for MT-FDR have competitive predictive performance in terms of similar AUC for test data. Gmllasso has the lowest mean FPR but also low mean TPR. SBLR models with either L_1 or elastic-net penalty achieve higher TPR and lower FPR on average than glmnet does under either sample size.

All the numerical experiments are conducted on a machine with six Intel Core i7 3.2 GHz processors and 64 GB of RAM. Algorithm 1 of SBLR method is implemented in Matlab (R2018a). Under the sample size $n = 1000$, SBLR with $K = 10$ takes 99.6 s on average to run a validation instance over the 5×11 grid under 5 initializations, glmnet takes 21 seconds and GC approach takes 130.4 s on average.

3.2. Non-clique signal subgraphs

Clique subgraphs have appealing interpretations but could be a restrictive assumption. We evaluate the performance of SBLR when such assumption is not true by associating the outcome y_i with non-clique signal subgraphs. Specifically, y_i is related to a 4-node ring graph and a 7-node star graph as displayed in Fig. 8. The nodes in the two sub-

graphs correspond to the basis vectors q_3 and q_6 in (6), respectively. In this setting, the total number of signal edges and their coefficients are the same as in Section 3.1. The sample size is set at $n = 1000$ and SBLR is estimated under $K = 10$. The tuning method and other settings are maintained the same for SBLR, MT-FDR and TNFA as in Section 3.1. Of note, we use 10-fold cross validation combined with the “one-standard-error” rule (Hastie et al., 2009) for glmlasso, glmnet and GC to further encourage sparsity for these methods. Under the one-standard-error rule, the optimal tuning parameters are selected corresponding to the most parsimonious model whose mean cross-validated deviance is within one standard-error of the minimum.

For a dataset simulated based on the signal subgraphs in Fig. 8, the estimated results of glmnet and GC under the optimal penalty factors are displayed in Fig. 9. Glnet and GC select fewer edges under the finer tuning approach, but the results can still be difficult to interpret. MT-FDR declares 92.11% of the total number of edges significant in this case, while TNFA selects no signal networks.

SBLR selects 5 subgraphs under the optimal penalty pair as shown in Fig. 10. The first subgraph corresponding to $\lambda_1 \beta_1 \beta_1^T$ in Fig. 10 partially recovers the star graph in Fig. 8, where the hub structure with Node 1 as the central node is evident. In specific, the first entry of β_1 has much larger magnitude than the other nonzero entries corresponding to the other 4 nodes in this subgraph. Therefore the edges from Node 1 to the other 4 nodes have much larger coefficients than other edges do in this subgraph, as shown in the estimated coefficient matrix $\lambda_1 \beta_1 \beta_1^T$ in Fig. 10. The second subgraph corresponding to $\lambda_3 \beta_3 \beta_3^T$ in Fig. 10 contains the ring graph in Fig. 8 with two wrong edges due to the clique constraint. But SBLR is still useful for detecting the set of nodes for this signal subgraph.

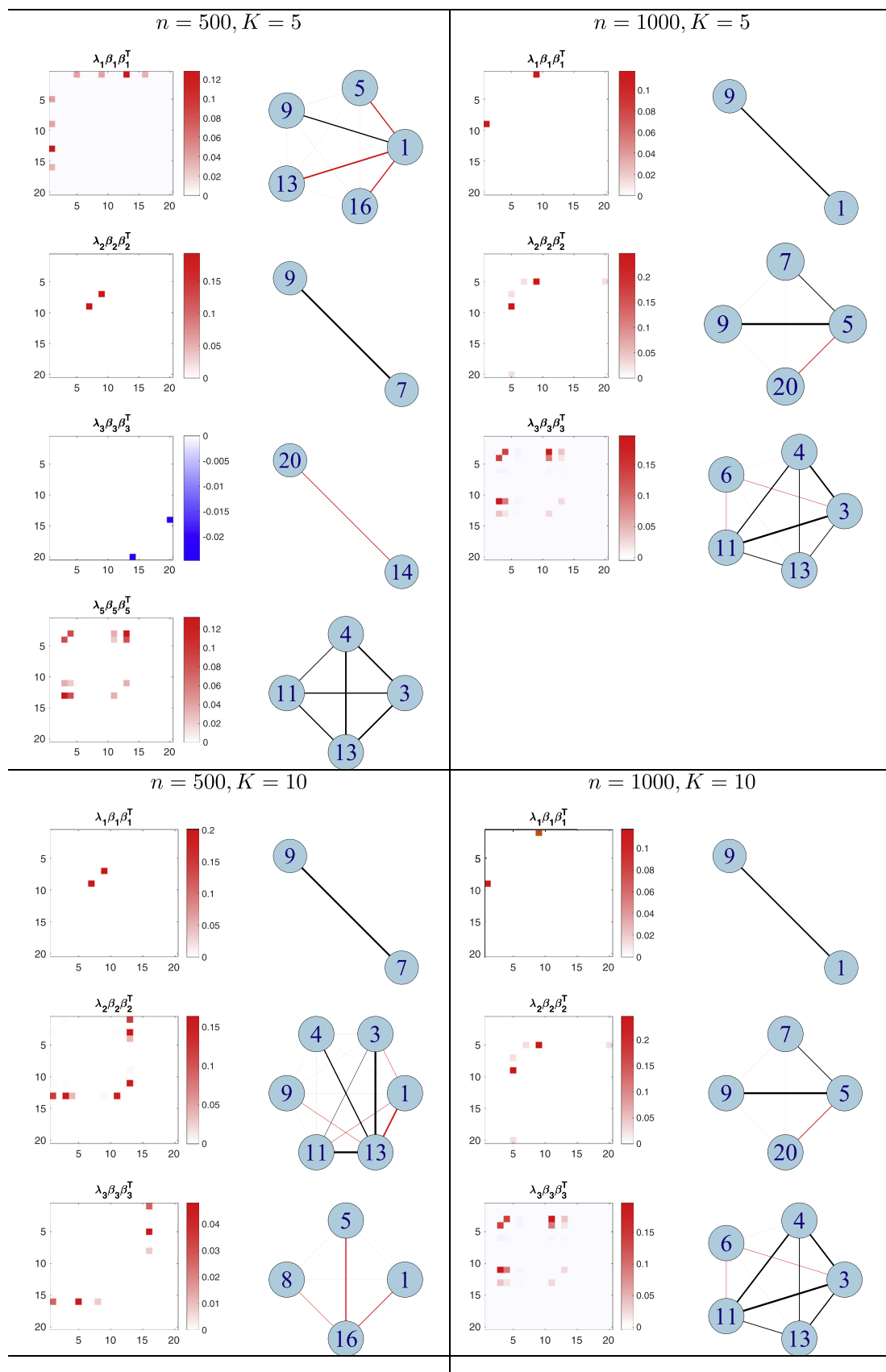


Fig. 6. Estimated nonzero coefficient components $\{\lambda_h \beta_h \beta_h^T\}$ (left) of SBLR and their selected subgraphs (right) under $K=5$ and $K=10$ for different sample sizes $n=500$ and $n=1000$, respectively. Black edges denote true signal edges and red ones falsely identified edges; the thickness of each edge is proportional to the magnitude of its estimated coefficient.

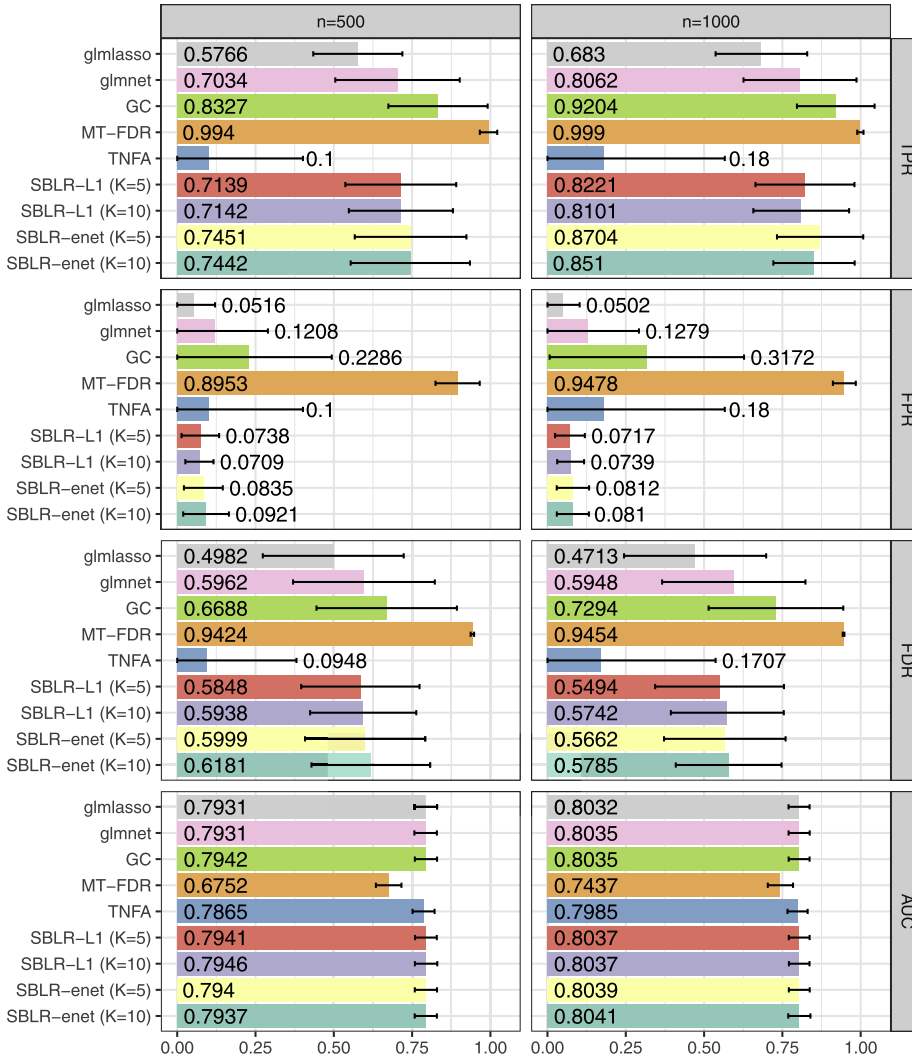


Fig. 7. Mean (bar labels) and standard deviation (error bars) of TPR, FPR, FDR and AUC across 100 simulations under two sample sizes ($n = 500$ and $n = 1000$). *Glmlasso* is logistic regression with L_1 penalty; *SBLR-L1* indicates that only L_1 penalty ($\eta = 1$) is applied for SBLR.

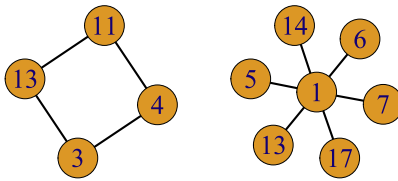


Fig. 8. True signal subgraphs in one simulation: a 4-node ring graph and a 7-node star graph.

Fig. 11 displays the mean and SD of the TPR, FPR, FDR in edge selection and AUC on test data for each method across 100 datasets simulated under $n = 1000$ in this setting. Compared to **Fig. 7**, the performance of SBLR decreases as expected when the clique assumption is not true, in terms of lower TPR, higher FPR and higher FDR on average under the same sample size with either L_1 or elastic-net penalty. The mean FPR and FDR of glmlasso, glmnet and GC decrease considerably under the one-standard-error rule compared to **Fig. 7**, but their mean TPRs also decrease, especially for glmlasso and GC. In addition, the 10-fold cross validation is more time-consuming than the validation tuning in **Section 3.1**. The glmnet now takes 204 seconds on average to run a cross-validation instance and GC takes about 25 minutes on average, while SBLR takes 106.2 seconds on average to run a validation instance in this case.

4. Application: SC subgraphs distinguishing high and low crystallized intelligence of adolescents

We apply our method to the ABCD dataset described in **Section 2.1** to examine the associations between structural brain network and crystallized intelligence for adolescents aged 9–10 years. NIH toolbox measures crystallized intelligence through two core tests: Picture Vocabulary test and Oral Reading Recognition test. They also provide a composite score, Crystallized Cognition Composite, to allow for evaluation of overall crystallized intelligence.

SBLR model is applied to differentiate high from low crystallized intelligence for each composite or domain-specific score, as well as identifying signal subgraphs contributing to the classification. Some input parameters of **Algorithm 1** for SBLR are set below in this section. The L_1 fractional penalty factor η is set to 1 to encourage sparsity in the results, as the AUC on test data is not sensitive to different values of η in this case. We set $K = 20$, which is larger than the number of selected subgraphs. The tolerance $\epsilon = 10^{-5}$ and 5 initializations are used in **Algorithm 1**. SBLR is trained over a sequence of 11 equally spaced δ values on the logarithmic scale with $\delta_{\min} = 0.1\delta_{\max}$. We also compare to glmlasso, penalized graph classification (GC), multiple testing screening (MT-FDR) and tensor factorization analysis (TNFA) on variable selection and predictive performance. Full rank $K = 68$ is set in TNFA, which explains around 75% of the variation in the brain networks.

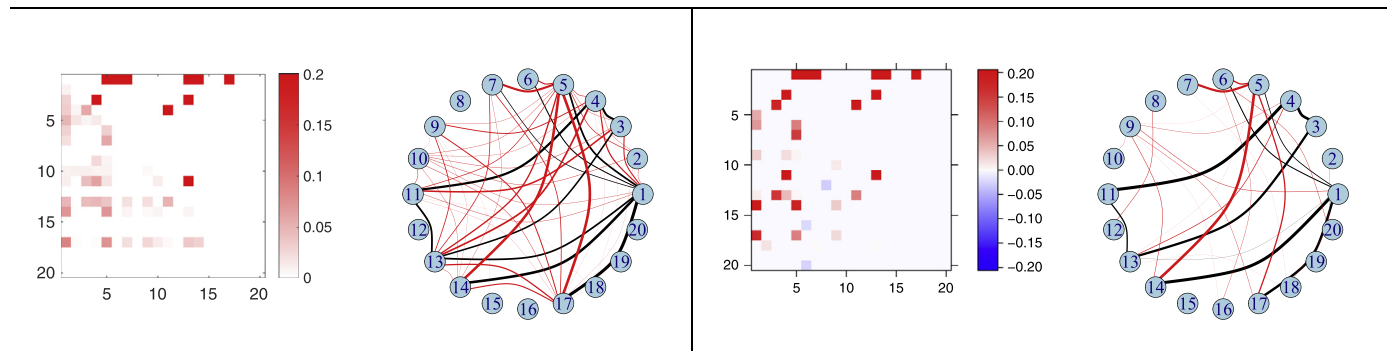


Fig. 9. Estimated results of glmnet (left) and GC (right) under the optimal penalty factors. In each panel, the left plot shows estimated coefficients (lower triangular) versus true values (upper triangular); the right plot shows the selected edges in the network, where black edges denote true signal edges and red ones falsely identified edges; the thickness of each edge is proportional to the magnitude of its estimated coefficient.

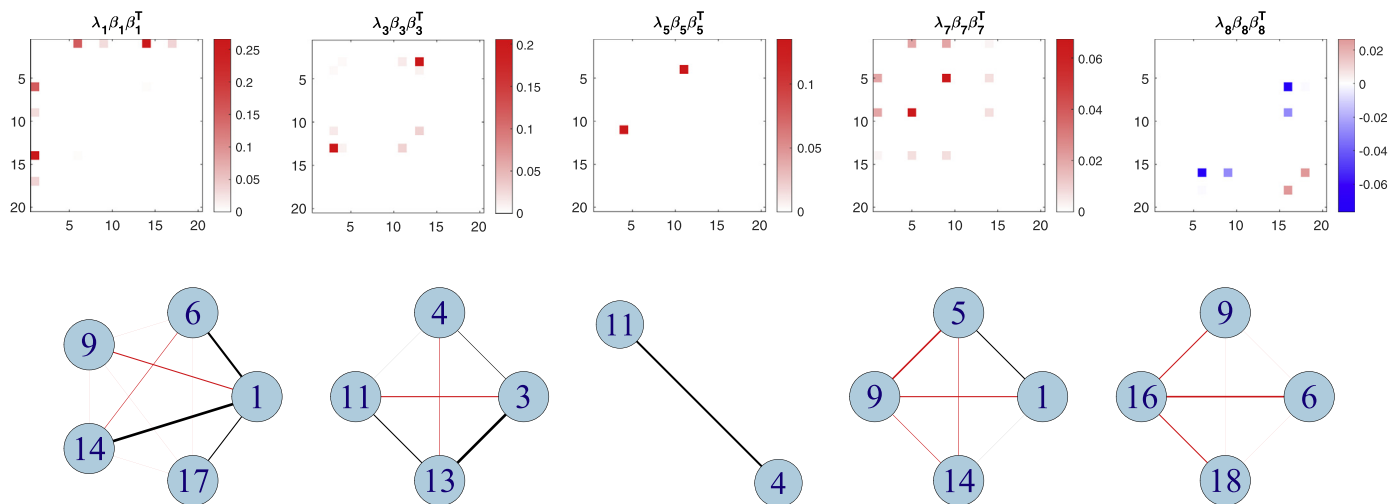


Fig. 10. Estimated nonzero coefficient matrices $\{\lambda_h \beta_h \beta_h^T\}$ (upper) of SBLR and their selected subgraphs (bottom). Black edges denote true signal edges and red ones falsely identified edges; the thickness of each edge is proportional to the magnitude of its estimated coefficient.

We employ stability selection (Meinshausen and Bühlmann, 2010; Shah and Samworth, 2013) to enhance significance of the selected variables (subgraphs or individual edges) for each method, which involves many rounds of random data splitting and keeping the variables with high selection probability. This approach reduces the variations in the results caused by single data split and random initialization. Specifically, the dataset for each cognitive measure is divided into 3 parts: training (60%), validation (20%) and test set (20%). SBLR, glmlasso and GC are fitted with training set and the optimal penalty factor(s) is tuned by validation set. The grid settings for tuning glmlasso and GC are the same as in the simulation study. MT-FDR and TNFA are fitted with training and validation sets. We record the selected variables for each method (under the optimal penalty factor) as well as AUC for test data across 30 rounds of random data splitting, and calculate the probability of each variable being selected. Directly counting the frequency of each identified subgraph of SBLR across multiple data splitting involves massive permutations. We actually took advantage of the special structure in the extracted subgraphs when counting their frequencies in the stability selection. Since the identified subgraphs of SBLR from ABCD data often display hub structure, we labeled each selected subgraph with the central node whose entry has the largest magnitude in the corresponding component vector β_h . Then we aligned the identified subgraphs (component vectors) with the same central node across 30 data splits to find the largest common intersection of these subgraphs that appear frequently, for example more than 50% of the time, which is still a clique.

4.1. Picture vocabulary

The Picture Vocabulary test uses an audio recording of words, presented with four photographic images on the computer screen. The participants are asked to select the picture that best matches the meaning of the word. Using the 1 standard deviation rule proposed in Section 2.1, we obtain a dataset containing 1034 kids with high picture vocabulary scores and 282 low scores.

SBLR on average selects 9.6 nonzero subgraphs $\{\lambda_h \beta_h \beta_h^T\}$ across 30 data splits with a standard deviation of 4.0. Summarizing the results from stability selection, the subgraphs with more than two ROIs and selection probabilities greater than 0.5 are displayed in Fig. 12. Fig. 12 also displays individual connections selected by glmlasso, GC and MT-FDR with probabilities greater than 0.5. As can be seen, the nonzero coefficients estimated by glmlasso and GC are quite similar while GC selects more predictive connections. But the results of these methods generally lack meaningful structure and are difficult to interpret neurologically. 12 components remain significant in TNFA more than 50% of the time, while their corresponding basis networks involve all the connections in the brain network. The average AUCs on test data across 30 splits of the data are all around 0.78 for SBLR, glmlasso and TNFA with a standard deviation around 0.03. GC achieves a bit higher AUC of 0.79 ± 0.03 , while that of MT-FDR is 0.72 ± 0.04 .

The three subgraphs identified by SBLR in Fig. 12 all seem to have hub structure, with the hub nodes being 26r (right rostral middle frontal), 34r (right insula) and 28r (right superior parietal) respectively.

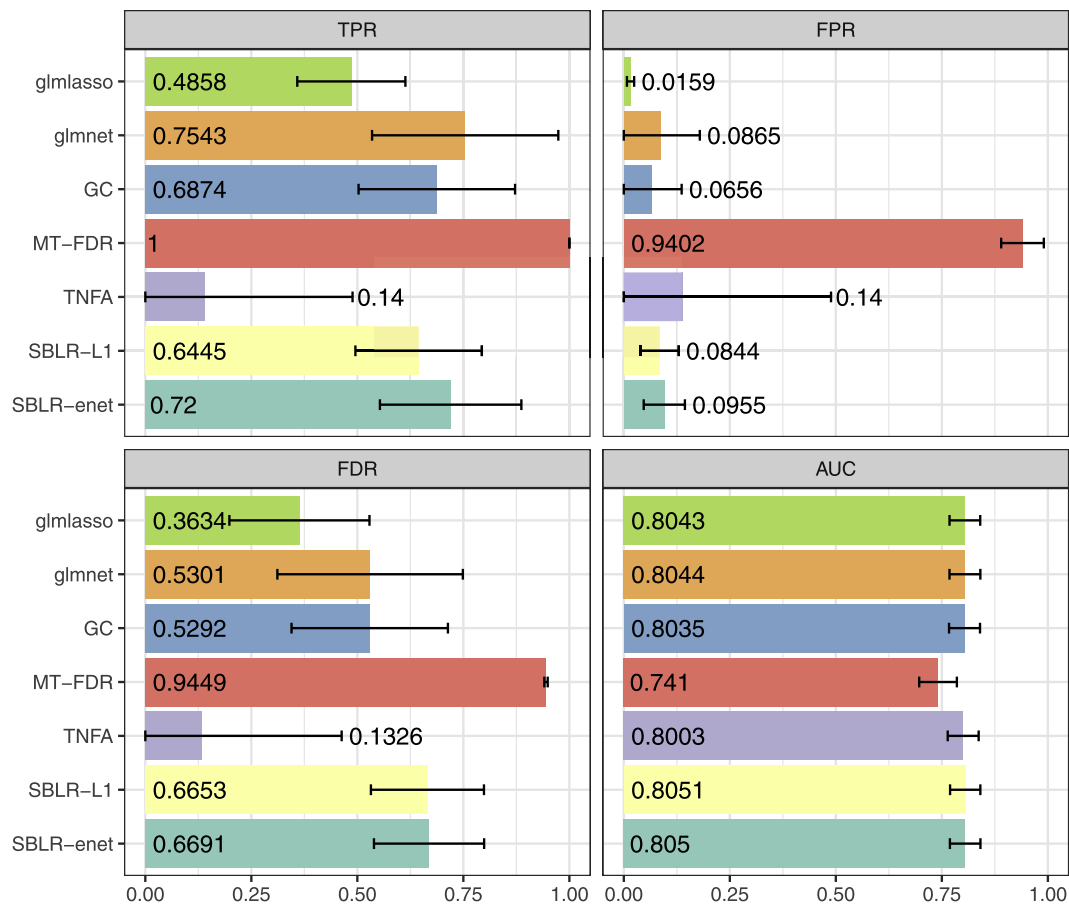


Fig. 11. Mean (bar labels) and standard deviation (error bars) of TPR, FPR, FDR and AUC across 100 simulations under $n = 1000$. *Glmlasso* is logistic regression with L_1 penalty; *SBLR-L1* indicates that only L_1 penalty ($\eta = 1$) is applied for SBLR ($K = 10$).

Plot (d) of Fig. 12 implies that right-handed children with stronger neural connections among 26r (right rostral middle frontal), 27r (right superior frontal) and 2r (right caudal anterior cingulate), and weaker connections among 26r, 29r (right superior temporal) and 30r (right supramarginal) are more likely to get high scores on Picture Vocabulary test. Plot (e) of Fig. 12 implies that right-handed children with stronger neural connections among 34r (right insula), 21r (right postcentral) and 1r (right bankssts), and weaker connections among 34r, 14r (right middle temporal) and 8r (right inferior temporal) are more likely to get high scores in Picture Vocabulary test. Plot (f) of Fig. 12 implies that right-handed children with stronger neural connection between 28r (right superior parietal) and 30r (right supramarginal), and weaker connection between 28r (right superior parietal) and 34l (left insula) are more likely to achieve high ability of this cognition measure. Some of these brain regions agree with the findings in neuroscience literature, for example, right superior parietal gyrus, right supramarginal gyrus and left insula are among the activated regions in auditory language processing tasks for children (Oh et al., 2014; Sugiura et al., 2011; Vogan et al., 2016).

4.2. Reading recognition

Participants on this test are asked to read and pronounce letters and words as accurately as possible. Applying the 1 standard deviation rule proposed in Section 2.1, we construct a dataset containing 918 subjects with high Reading Recognition scores and 477 subjects with low scores.

SBLR on average selects 9.3 ± 4.2 nonzero subgraphs across 30 splits of the dataset. Fig. 13 displays the subgraphs identified by SBLR with selection probabilities > 0.5 , along with connections selected by glmlasso, GC and MT-FDR with probabilities > 0.5 . The estimated nonzero coefficients of glmlasso and GC still look similar, but GC on average selects

almost double the connections as that of glmlasso. Four components remain significant in TNFA more than 50% of the time, which correspond to all the connections in the brain network. GC on average achieves the highest AUC for test data of 0.73 ± 0.02 across 30 splits. The AUCs of the rest methods are all around 0.70 ± 0.03 in this case.

The subgraphs in Fig. 13 located by SBLR display hub structure again and share many similarities to those associated with Picture Vocabulary. Compared to Fig. 12, Reading Recognition measure has two hub nodes in common with Picture Vocabulary: 26r (right rostral middle frontal) and 28r (right superior parietal). And Plot (f) of Fig. 12 is a subgraph of Plot (f) in Fig. 13. The latter has two extra ROIs: 21r (right postcentral) and 27l (left superior frontal), and stronger connections among the two nodes and the hub node (28r) have positive effect on the cognitive functioning of Reading Recognition. Plot (d) of Fig. 13 implies that right-handed children with stronger neural connections among 17r (right pars opercularis), 26r (right rostral middle frontal) and 31r (right frontal pole) are more likely to get high scores on Reading Recognition test, while the connection strengths among 26r, 32r (right temporal pole) and 34r (right insula) may have negative effect on this cognitive functioning. Plot (e) of Fig. 13 implies that stronger neural connections among 17l (left pars opercularis), 27r (right superior frontal) and 2r (right caudal anterior cingulate) have positive effect on this cognitive functioning, while the connection strength between 27r and 9r (right isthmus cingulate) may have a slight negative effect.

4.3. Crystallized cognition composite

Crystallized Cognition Composite can be interpreted as a global assessment of verbal reasoning. This composite score is derived by averaging the normalized scores of Picture Vocabulary and Reading Recogni-

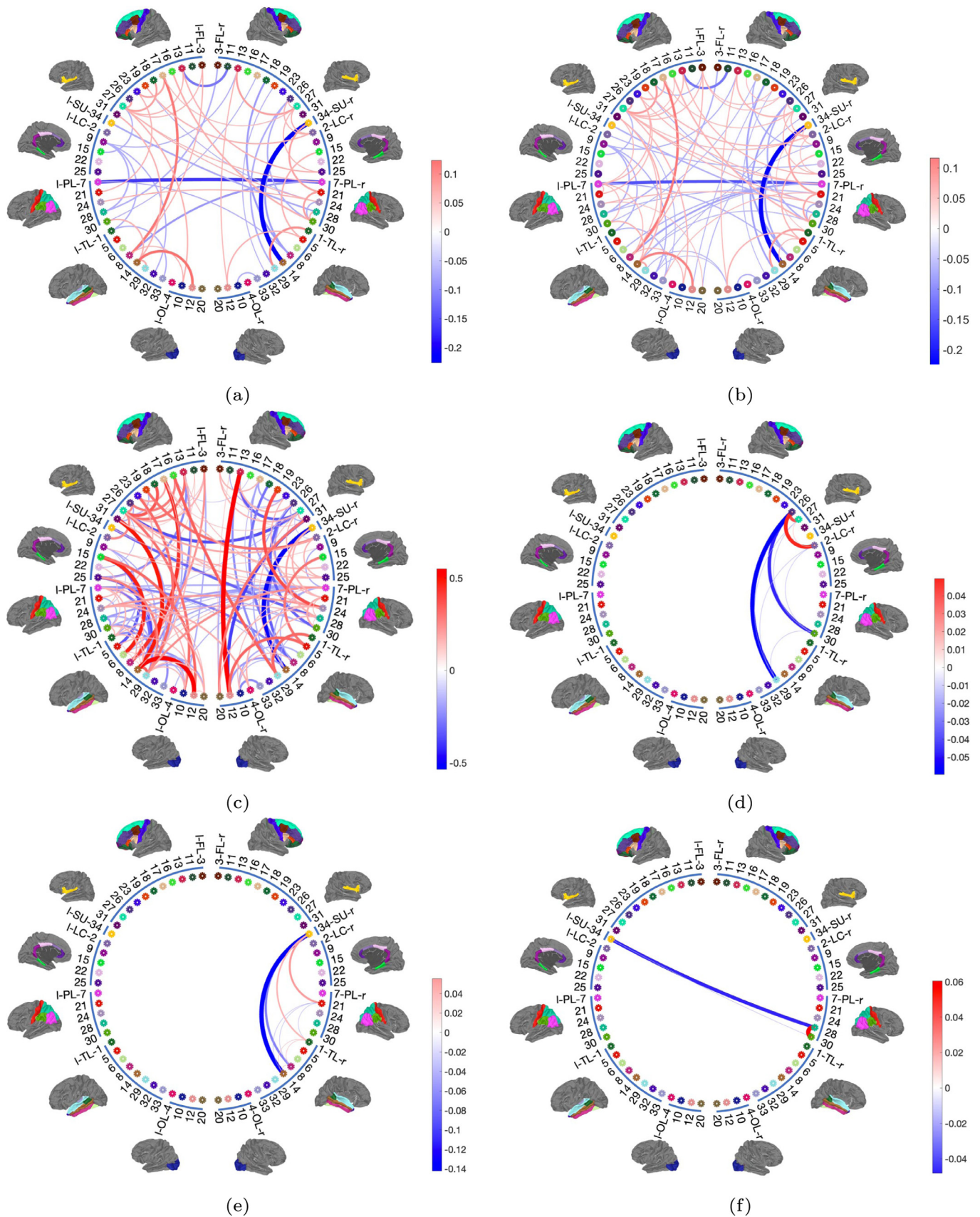


Fig. 12. Results of Picture Vocabulary: (a) connections selected by glmlasso with selection probabilities > 0.5 across 30 rounds of random data splitting; (b) connections selected by GC with selection probabilities > 0.5; (c) connections selected by MT-FDR with selection probabilities > 0.5; (d) - (f) subgraphs (≥ 3 ROIs) selected by SBLR with selection probabilities > 0.5. The subgraphs are arranged in the descending order of selection frequency. The thickness of each edge is proportional to the magnitude of its mean estimated coefficient; the color goes from blue to red as the coefficient goes from negative to positive.

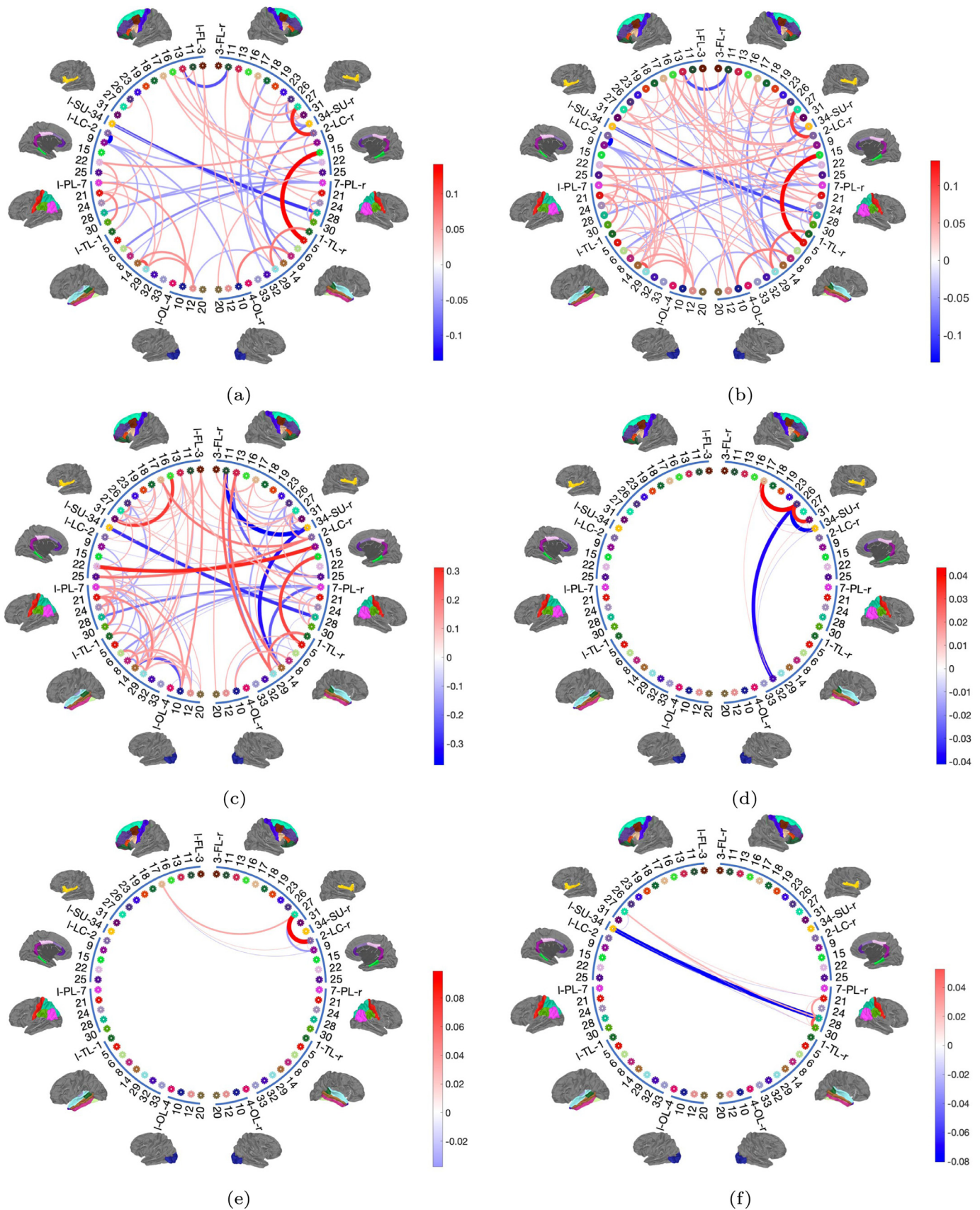


Fig. 13. Results of Reading Recognition: (a) connections selected by glmlasso with selection probabilities > 0.5 across 30 splits of the dataset; (b) connections selected by GC with selection probabilities > 0.5 ; (c) connections selected by MT-FDR with selection probabilities > 0.5 ; (d) - (f) subgraphs (≥ 3 ROIs) selected by SBLR with selection probabilities > 0.5 . The subgraphs are arranged in the descending order of selection frequency. The thickness of each edge is proportional to the magnitude of its mean estimated coefficient; the color goes from blue to red as the coefficient goes from negative to positive.

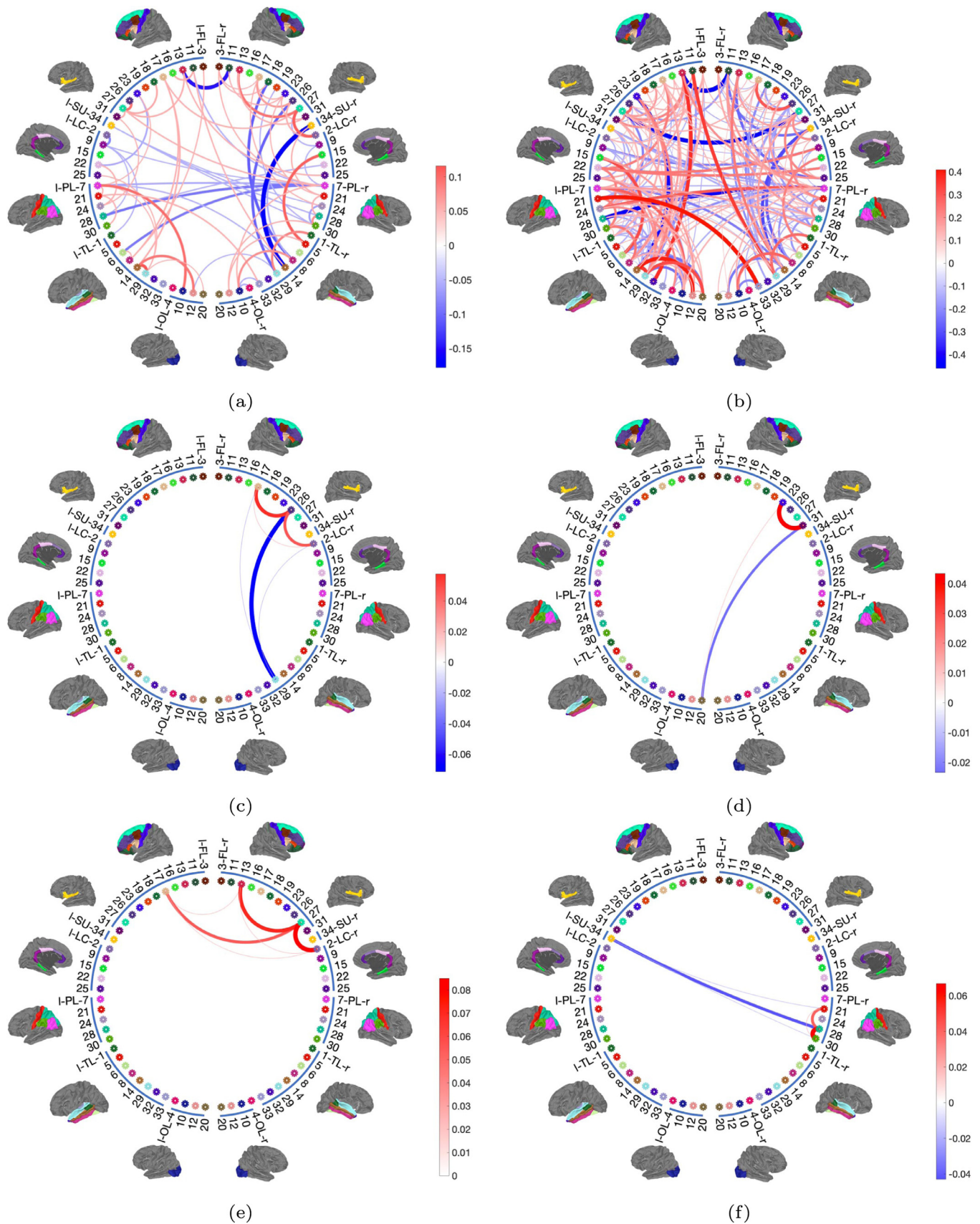


Fig. 14. Results of Crystallized Cognition Composite: (a) connections selected by glmlasso with selection probabilities > 0.5 across 30 splits of the data; (b) connections selected by MT-FDR with probabilities > 0.5; (c) - (f) subgraphs (≥ 3 ROIs) selected by SBLR with probabilities > 0.5. The subgraphs are arranged in the descending order of selection frequency. The thickness of each edge is proportional to the magnitude of its mean estimated coefficient; the color goes from blue to red as the coefficient goes from negative to positive.

tion test, and the age-adjusted scale scores are calculated based on this new distribution (Weintraub et al., 2013). Under the 1 standard deviation rule, we identify 1084 high crystallized cognitive cases and 536 low crystallized cognitive cases.

SBLR on average selects 12.3 ± 3.5 nonzero subgraphs across 30 splits of the data. Fig. 14 displays the subgraphs identified by SBLR with selection probabilities > 0.5 , along with individual connections selected by glmlasso and MT-FDR with probabilities > 0.5 . GC selects far more connections on average than glmlasso or MT-FDR does in this case, with almost two times that of MT-FDR, and hence its results are not displayed. 9 components maintain significant in TNFA more than 50% of the time, which again corresponds to all the connections in the brain network. The AUCs on test data across 30 splits are around 0.76 ± 0.02 for SBLR, glmlasso and TNFA, 0.78 ± 0.02 for GC, and 0.71 ± 0.03 for MT-FDR.

The subgraphs identified by SBLR in Fig. 14 predictive of high and low crystallized cognitive ability have many overlaps with the selected subgraphs associated with Picture Vocabulary and Reading Recognition. The subgraph in Plot (c) of Fig. 14 seems like a “composite” of the subgraphs in Plot (d)’s of both Figs. 12 and 13. Plot (f) of Fig. 14 echoes Plot (f)’s in both Figs. 12 and 13. The subgraph in Plot (e) of Fig. 14 overlaps with that in Plot (e) of Fig. 13. Plot (e) of Fig. 14 implies that right-handed children with stronger neural connections among 17l (left pars opercularis), 13r (right medial orbitofrontal), 27r (right superior frontal) and 2r (right caudal anterior cingulate) are more likely to have high level of functioning on crystallized cognition. Plot (d) of Fig. 14 implies that stronger connection between 31r (right frontal pole) and 23r (right pre-central) has positive effect on crystallized cognitive functioning, while the connection strength between 31r and 20l (left pericalcarine) may have negative effect.

5. Discussion

We have presented a useful tool for studying differences in brain connectivity patterns between groups, which produces more interpretable results than unstructured classifiers do, while maintaining competitive predictive performance. Simulation studies show that SBLR can successfully identify true signal subgraphs with high TPR and low FPR when the model assumption is true, and its selection accuracy could generally be improved by increasing the number of subjects in the data. In applications to crystallized cognition data, although penalized graph classification method (Arroyo Reli6n et al., 2019) obtains a bit higher predictive performance than SBLR, our method contributes to a more insightful understanding of the sub-structure of brain connectome related to crystallized intelligence, with discovery of a sequence of sub-networks with hub structure as well as the leading brain regions in these subgraphs.

SBLR is theoretically able to represent any coefficients in a generalized linear regression with sufficient number of components (one edge per component). But the regularization terms encourage small and fewer cliques. Simulation study shows that SBLR tends to introduce redundant edges when true signal subgraphs have sparse structure. But SBLR is still useful for capturing the hub structure of a star signal subgraph and detecting the sets of nodes for predictive subgraphs. In practice, we can evaluate the out-of-sample predictive performance of SBLR to determine whether the model assumption is reasonable. High prediction accuracy could justify the model assumption to the extent that putting the structured constraints on the coefficients makes the unstructured model less overfitting. We suggest to use the proposed model when the number of selected subgraphs is small and the out-of-sample predictive performance is comparable to or better than glmnet or GC. Ideally, we could have statistical tests for each coefficient in the subgraphs, but testing the significance of the coefficients in SBLR is a challenging open problem due to the bilinear framework and inclusion of the penalty terms (Xia et al., 2020).

There are other ways to define subgraphs, for example, Vogelstein et al. (2012) defined a subgraph as a minimum set of vertices and edges distinguishing groups; Khambhati et al. (2018) constructed

subgraphs based on an unsupervised non-negative matrix factorization, which is similar to Zhang et al. (2019); Chen et al. (2020) detected and tested altered brain connectivity networks with k-partite graph topological structure; mutually exclusive subgraphs may capture the modular structure of brain connectome. More flexible ways of defining subgraphs with more interesting network topological properties can be explored in the future.

Although SBLR is motivated from structural connectivity analysis, the method can certainly be used in functional brain connectivity analysis to identify signal subgraphs in the functional brain network related to a binary outcome, after constructing the functional connectivity matrices based on correlations of BOLD time series data of pairwise brain regions. In addition, it is an interesting future direction to consider how to aggregate community-level or lobe-level information into SBLR to improve the selection accuracy of signal subgraphs. For example, Xia et al. (2020) exploited both edge- and community-level information in modeling brain networks. Xie et al. (2013) compared different algorithms in overlapping community detection.

Very few studies have focused on WM’s contribution to cognition. Specific to crystallized cognition, there have only been a handful of studies examining major WM tracts’ roles. For example, higher FA in superior longitudinal fasciculus was related to higher crystallized cognition in children (Simpson-Kent et al., 2020), higher FA in forceps minor was related to higher crystallized cognition in adults (G6ngora et al., 2020). Both tracts contain numerous inter-hemisphere connections across several brain regions (e.g., superior longitudinal fasciculus) or within frontal region (e.g., forceps minor). Our study represents the very first effort in the literature specifying clique subgraphs of SC related to crystallized cognition. Importantly, across composite score and individual domains of crystallized cognition, we identified consistent brain regions and subgraphs (refer to Figure 12, 13 and 14) predominantly in right-hemisphere frontal-parietal regions. The finding is consistent to cumulative functional subgraph literature on frontal-parietal driven executive network during neuro-development (Chai et al., 2017), and flexible periphery of the language network (Fedorenko and Thompson-Schill, 2014).

We also noticed that the classification accuracy with AUC around 0.78 for crystallized cognition is very high. From our previous study, SC is robust and reproducible and more predictive of cognition compared with functional connectivity (FC) derived from functional MRI (Zhang et al., 2019; 2018). We hypothesize that SC is a better biomarker for understanding the cognition development in adolescents. To verify this hypothesis, analyses and comparisons with FC seem to be a natural next step.

The frontal-parietal driven subgraphs among those aged 9-10 years represent a set of critical biomarkers for overall neuro-development. However, these subgraphs’ physiological meaning can be transient across ages. Comparisons across a broader age range should be conducted to further confirm the role of these subgraphs in crystallized cognition and its development. With more data being recorded in the ABCD study, we hope to further analyze SC and cognition development.

Data and code availability statement

The ABCD dataset analyzed in the article is downloaded from NIH Data Archive (<https://nda.nih.gov>). The code developed in this study is publicly available at <https://github.com/wangronglu/SBLR>

CRediT authorship contribution statement

Lu Wang: Methodology, Formal analysis, Writing - original draft. **Feng Vankee Lin:** Validation, Writing - review & editing. **Martin Cole:** Software, Data curation. **Zhengwu Zhang:** Conceptualization, Writing - review & editing, Project administration.

Acknowledgment

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from NIMH Data Archive Digital Object Identifier (10.15154/1520158). DOIs can be found at <https://dx.doi.org/10.15154/1520158>.

The research of L. Wang is supported by National Natural Science Foundation of China [grant number 11901583]. Zhang wants to acknowledge support from the National Institutes of Health (NIH) of the United States under award number MH118927 and MH118020.

Appendix A. Coordinate descent algorithm for model estimation

A1. Updates for entries in $\{\beta_h\}_{h=1}^K$

Minimizing the loss function (4) with respect to β_{hu} , the u th entry of β_h , given all the other parameters becomes:

$$\min_{\beta_{hu}} L(\beta_{hu}) = -\frac{1}{n} \sum_{i=1}^n l_i(\beta_{hu}) + d_{hu} |\beta_{hu}| + e_{hu} \beta_{hu}^2 / 2 \quad (\text{A.1})$$

where

$$d_{hu} = \delta \eta |\lambda_h| \sum_{v \neq u} |\beta_{hv}| \quad (\text{A.2})$$

$$e_{hu} = \delta(1 - \eta) \lambda_h^2 \sum_{v \neq u} \beta_{hv}^2. \quad (\text{A.3})$$

Eqs. (A.2), (A.3) imply that the penalty factors d_{hu} and e_{hu} in (A.1) for $|\beta_{hu}|$ are related to the nonzero entries in β_h excluding β_{hu} . Hence β_{hu} is more likely to be shrunk to zero if the current number of nonzero entries in β_h is large. This adaptive penalty will lead to a set of sparse vectors $\{\beta_h\}_{h=1}^K$ and hence a set of small signal subgraphs.

Note that

$$\frac{\partial l_i}{\partial \beta_{hu}} = 2(y_i - p_i) \lambda_h \bar{W}_{i[u-]} \beta_h \quad (\text{A.4})$$

$$\frac{\partial^2 l_i}{\partial \beta_{hu}^2} = -p_i(1 - p_i) (2 \lambda_h \bar{W}_{i[u-]} \beta_h)^2 \leq 0. \quad (\text{A.5})$$

Therefore each log-likelihood l_i is a concave function of β_{hu} when fixing the others, and (A.1) is a convex optimization for β_{hu} .

Suppose the current estimate for β_{hu} at iteration t is $\beta_{hu}^{(t)}$ and fix the other parameters at their current estimates. The Newton algorithm for minimizing $-\frac{1}{n} \sum_{i=1}^n l_i(\beta_{hu})$ in (A.1) is equivalent to minimizing the following second-order Taylor expansion at the current estimate $\beta_{hu}^{(t)}$:

$$L_1^{(t)}(\beta_{hu}) \triangleq c_{hu}^{(t)} + b_{hu}^{(t)} (\beta_{hu} - \beta_{hu}^{(t)}) + \frac{a_{hu}^{(t)}}{2} (\beta_{hu} - \beta_{hu}^{(t)})^2 \quad (\text{A.6})$$

where

$$c_{hu}^{(t)} = -\frac{1}{n} \sum_{i=1}^n l_i(\beta_{hu}^{(t)}) \quad (\text{A.7})$$

$$b_{hu}^{(t)} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial l_i(\beta_{hu}^{(t)})}{\partial \beta_{hu}} \quad (\text{A.8})$$

$$a_{hu}^{(t)} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 l_i(\beta_{hu}^{(t)})}{\partial \beta_{hu}^2}. \quad (\text{A.9})$$

Similar to Friedman et al. (2010), β_{hu} is then updated by minimizing

$$L_1^{(t)}(\beta_{hu}) + d_{hu}^{(t)} |\beta_{hu}| + e_{hu}^{(t)} \beta_{hu}^2 / 2 \quad (\text{A.10})$$

which has a closed form solution

$$\beta_{hu}^{(t+1)} = \frac{\text{sign}(a_{hu}^{(t)} \beta_{hu}^{(t)} - b_{hu}^{(t)})}{a_{hu}^{(t)} + e_{hu}^{(t)}} \cdot \left(|a_{hu}^{(t)} \beta_{hu}^{(t)} - b_{hu}^{(t)}| - d_{hu}^{(t)} \right)_+ \quad (\text{A.11})$$

if $a_{hu}^{(t)} + e_{hu}^{(t)} > 0$. From (A.3), (A.5) and (A.9) we know that $a_{hu}^{(t)} \geq 0$ and $e_{hu}^{(t)} \geq 0$.

The case $a_{hu}^{(t)} + e_{hu}^{(t)} = 0$ implies that (i) $\lambda_h^{(t)} = 0$ (ii) $\beta_{hv}^{(t)} = 0$ for $v \neq u$ or (iii) $\eta = 1$ and $p_i^{(t)} = 0$ or 1 , $\forall i$. For the former two cases, the lower triangular part of the component matrix $\lambda_h^{(t)} \beta_h^{(t)} \beta_h^{(t)\top}$ becomes zero no matter what value β_{hu} takes. So we set $\beta_{hu}^{(t+1)} = 0$ in these cases. Regarding (iii), if $p_i^{(t)} \equiv y_i$, $\forall i$, (A.10) is minimized at $\beta_{hu} = 0$ because $b_{hu}^{(t)} = 0$ at this time according to (A.4) and (A.8) together with $a_{hu}^{(t)} = e_{hu}^{(t)} = 0$. Then the first derivative of (A.10) becomes $d_{hu}^{(t)} (> 0)$ when $\beta_{hu} > 0$ and $-d_{hu}^{(t)} (< 0)$ when $\beta_{hu} < 0$. Otherwise, $p_i^{(t)} = 0$ or $1 \forall i$ may be due to bad initialization. For example, a large magnitude of initial values of the parameters could easily make p_i become 1 or 0, $\forall i$ through the logit link in (1). In this case, setting $\beta_{hu} = 0$ could prevent the divergence of the solution. In practice, we always normalize the entries in $\{W_i\}$ before applying SBLR. We also recommend to initialize each parameter from $U(-0.1, 0.1)$ to avoid the explosion in the logit scale.

The computational complexity of updating each entry β_{hu} is $O(nV)$ per iteration and thus that of updating $\{\beta_h\}_{h=1}^K$ is $O(nKV^2)$.

A2. Updates for $\{\lambda_h : h = 1, \dots, K\}$

Minimizing the loss function (4) with respect to λ_h when fixing the others amounts to:

$$\min_{\lambda_h} L(\lambda_h) = -\frac{1}{n} \sum_{i=1}^n l_i(\lambda_h) + d_h |\lambda_h| + e_h \frac{\lambda_h^2}{2} \quad (\text{A.12})$$

where $d_h = \delta \eta \sum_{u=1}^V \sum_{v < u} |\beta_{hu}| |\beta_{hv}|$ and $e_h = \delta(1 - \eta) \sum_{u=1}^V \sum_{v < u} \beta_{hu}^2 \beta_{hv}^2$.

Since λ_h is linear in the logit link (1) given the others, the log-likelihood is a concave function of λ_h . Suppose the current estimate for λ_h is $\lambda_h^{(t)}$. Similar to Section A.1, λ_h is updated by minimizing the following quadratic approximation to (A.12):

$$c_{\lambda,h}^{(t)} + b_{\lambda,h}^{(t)} (\lambda_h - \lambda_h^{(t)}) + \frac{a_{\lambda,h}^{(t)}}{2} (\lambda_h - \lambda_h^{(t)})^2 + d_h^{(t)} |\lambda_h| + e_h^{(t)} \frac{\lambda_h^2}{2} \quad (\text{A.13})$$

where

$$\begin{aligned} c_{\lambda,h}^{(t)} &= -\frac{1}{n} \sum_{i=1}^n l_i(\lambda_h^{(t)}) \\ b_{\lambda,h}^{(t)} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial l_i(\lambda_h^{(t)})}{\partial \lambda_h} = -\frac{1}{n} \sum_{i=1}^n (y_i - p_i^{(t)}) \beta_h^{(t)\top} \bar{W}_i \beta_h^{(t)} \\ a_{\lambda,h}^{(t)} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 l_i(\lambda_h^{(t)})}{\partial \lambda_h^2} = \frac{1}{n} \sum_{i=1}^n p_i^{(t)} (1 - p_i^{(t)}) (\beta_h^{(t)\top} \bar{W}_i \beta_h^{(t)})^2. \end{aligned} \quad (\text{A.14})$$

Note that $a_{\lambda,h}^{(t)} \geq 0$ and $e_h^{(t)} \geq 0$. If $a_{\lambda,h}^{(t)} + e_h^{(t)} > 0$, λ_h is updated to the argmin of (A.13):

$$\lambda_h^{(t+1)} = \frac{\text{sign}(a_{\lambda,h}^{(t)} \lambda_h^{(t)} - b_{\lambda,h}^{(t)})}{a_{\lambda,h}^{(t)} + e_h^{(t)}} \cdot \left(|a_{\lambda,h}^{(t)} \lambda_h^{(t)} - b_{\lambda,h}^{(t)}| - d_h^{(t)} \right)_+. \quad (\text{A.15})$$

If $a_{\lambda,h}^{(t)} + e_h^{(t)} = 0$, then either the component matrix $\beta_h^{(t)} \beta_h^{(t)\top}$ is a zero matrix or $p_i^{(t)} = 0$ or 1, $\forall i$. In either case, $\lambda_h^{(t+1)}$ is set to 0 following a similar discussion in Section A.1.

The computational complexity for updating each λ_h is $O(nV^2)$ per iteration and hence that of updating $\{\lambda_h\}_{h=1}^K$ is $O(nKV^2)$.

A.3. Update for α

$\alpha \in \mathbb{R}^m$ is also updated by minimizing the quadratic approximation to $-\frac{1}{n} \sum_{i=1}^n l_i$ at the current estimate $\alpha^{(t)}$ with the updating rule

$$\alpha^{(t+1)} = \begin{cases} \alpha^{(t)} - [A^{(t)}]^{-1} b^{(t)} & \text{if } A^{(t)} \text{ is nonsingular} \\ \mathbf{0} & \text{otherwise} \end{cases} \quad (\text{A.16})$$

where

$$\begin{aligned} b^{(t)} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial l_i(\alpha^{(t)})}{\partial \alpha} = -\frac{1}{n} \sum_{i=1}^n (y_i - p_i^{(t)}) \mathbf{x}_i \\ A^{(t)} &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 l_i}{\partial \alpha \partial \alpha^\top} = \frac{1}{n} \sum_{i=1}^n p_i^{(t)} (1 - p_i^{(t)}) \mathbf{x}_i \mathbf{x}_i^\top. \end{aligned} \quad (\text{A.17})$$

The computational complexity of this step is $O(m^2 n)$.

References

- Arroyo Reli3n, J.D., Kessler, D., Levina, E., Taylor, S.F., et al., 2019. Network classification with applications to brain connectomics. *Ann. Appl. Stat.* 13 (3), 1648–1677.
- Auchter, A.M., Mejia, M.H., Heyser, C.J., Shilling, P.D., Jernigan, T.L., Brown, S.A., Taupert, S.F., Dowling, G.J., 2018. A description of the ABCD organizational structure and communication framework. *Dev. Cogn. Neurosci.* 32, 8–15.
- Bassett, D.S., Sporns, O., 2017. Network neuroscience. *Nat. Neurosci.* 20 (3), 353–364.
- Benjamini, Y., Yekutieli, D., et al., 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29 (4), 1165–1188.
- Casey, B., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Cornejo, M.D., Scat, C.S., Harms, M.P., Dosenbach, N.U., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* 32, 43–54.
- Chai, L.R., Khambhati, A.N., Ciric, R., Moore, T.M., Gur, R.C., Gur, R.E., Satterthwaite, T.D., Bassett, D.S., 2017. Evolution of brain network dynamics in neurodevelopment. *Netw. Neurosci.* 1 (1), 14–30.
- Chen, S., Bowman, F.D., King, Y., 2020. Detecting and testing altered brain connectivity networks with k-partite network topology. *Comput. Stat. Data Anal.* 141, 109–122.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44 (3), 837–845.
- 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.
- Durante, D., Dunson, D.B., et al., 2018. Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.* 13 (1), 29–58.
- Fedorenko, E., Thompson-Schill, S.L., 2014. Reworking the language network. *Trends Cogn. Sci.* 18 (3), 120–126.
- Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., S3gonne, F., Quinn, B.T., Dale, A.M., 2004. Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23 (Supplement 1), S69–S84.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22.
- Girard, G., Whittingstall, K., Deriche, R., Descoteaux, M., 2014. Towards quantitative connectivity analysis: reducing tractography biases. *NeuroImage* 98, 266–278.
- Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L., Auerbach, E.J., Behrens, T.E., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., et al., 2016. The human connectome project's neuroimaging approach. *Nat. Neurosci.* 19 (9), 1175–1187.
- G3ngora, D., Vega-Hern3ndez, M., Jahanshahi, M., Vald3s-Sosa, P.A., Bringas-Vega, M.L., CHBMP, 2020. Crystallized and fluid intelligence are predicted by microstructure of specific white-matter tracts. *Hum. Brain Mapp.* 41 (4), 906–916.
- Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
- Khambhati, A.N., Medaglia, J.D., Karuzu, E.A., Thompson-Schill, S.L., Bassett, D.S., 2018. Subgraphs of functional brain networks identify dynamical constraints of cognitive control. *PLoS Comput. Biol.* 14 (7), e1006234.
- Maier-Hein, K.H., Neher, P.F., Houde, J.-C., C3t3e, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., et al., 2017. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* 8 (1), 1349.
- Meinshausen, N., B3hlmann, P., 2010. Stability selection. *J. Royal Stat. Soc. Ser. B* 72, 417–473.
- Miller, K.L., Alfaro-Almagro, F., Bangarter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19 (11), 1523.
- Oh, A., Duerden, E.G., Pang, E.W., 2014. The role of the insula in speech and language processing. *Brain Lang.* 135, 96–103.
- Park, H.-J., Friston, K., 2013. Structural and functional brain networks: from connections to cognition. *Science* 342 (6158), 1238411.
- Penke, L., Maniega, S.M., Bastin, M., Hern3ndez, M.V., Murray, C., Royle, N., Starr, J., Wardlaw, J., Deary, I., 2012. Brain white matter tract integrity as a neural foundation for general intelligence. *Mol. Psychiatry* 17 (10), 1026–1030.
- Seidman, S., Foster, B., 1978. A graph-theoretic generalization of the clique concept. *J. Math. Sociol.* 6 (1), 139–154.
- Shah, R.D., Samworth, R.J., 2013. Variable selection with error control: another look at stability selection. *J. Royal Stat. Soc. Ser. B (Stat. Methodol.)* 75 (1), 55–80.
- Simpson, S.L., Hayasaka, S., Laurienti, P.J., 2011. Exponential random graph modeling for complex brain networks. *PLoS One* 6 (5), 1–11.
- Simpson-Kent, L.L., Fuhrmann, D., Bathelt, J., Achterberg, J., Borgeest, G.S., Kievit, R.A., CALM Team, et al., 2020. Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Dev. Cogn. Neurosci.* 41, 100743.
- Sugiura, L., Ojima, S., Matsuba-Kurita, H., Dan, I., Tsuzuki, D., Katura, T., Hagiwara, H., 2011. Sound to language: different cortical processing for first and second languages in elementary school children as revealed by a large-scale study using fNIRS. *Cereb. Cortex* 21 (10), 2374–2393.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Wu-Minn HCP Consortium, et al., 2013. The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79.
- Vogan, V., Morgan, B., Powell, T., Smith, M., Taylor, M., 2016. The neurodevelopmental differences of increasing verbal working memory demand in children and adults. *Dev. Cogn. Neurosci.* 17, 19–27. Special Section: The Developmental Neuroscience of Adolescence: Revisiting, Refining, and Extending Seminal Models.
- Vogelstein, J.T., Roncal, W.G., Vogelstein, R.J., Priebe, C.E., 2012. Graph classification using signal-subgraphs: applications in statistical connectomics. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (7), 1539–1551.
- Wang, L., Zhang, Z., Dunson, D.B., 2019. Symmetric bilinear regression for signal subgraph estimation. *IEEE Trans. Signal Process.* 67, 1929–1940.
- Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulsky, D.S., Zelazo, P.D., Bauer, P.J., Carulloz, N.E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N.A., Beaumont, J.L., Mungas, D., Nowinski, C.J., Richler, J., Deocampo, J.A., Anderson, J.E., Manly, J.J., Borosh, B., Havlik, R., Conway, K., Edwards, E., Freund, L., King, J.W., Moy, C., Witt, E., Gershon, R.C., 2013. Cognition assessment using the NIH toolbox. *Neurology* 80 (Issue 11, Supplement 3), S54–S64.
- Xia, C.H., Ma, Z., Cui, Z., Bzdok, D., Bassett, D.S., Satterthwaite, T.D., Shinohara, R.T., Witten, D., 2020. Multi-scale network regression for brain-phenotype associations. *Hum. Brain Mapp.* 41, 2553–2566.
- Xie, J., Kelley, S., Szymanski, B.K., 2013. Overlapping community detection in networks: the state-of-the-art and comparative study. *ACM Comput. Surv.* 45 (4), 43.
- Yeh, C.-H., Smith, R.E., Liang, X., Calamante, F., Connelly, A., 2016. Correction for diffusion MRI fibre tracking biases: the consequences for structural connectomic metrics. *Neuroimage* 142, 150–162.
- Zhang, Z., Allen, G.I., Zhu, H., Dunson, D., 2019. Tensor network factorizations: relationships between brain structural connectomes and traits. *Neuroimage* 197, 330–343.
- 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.
- Zhao, T., Duan, F., Liao, X., Dai, Z., Cao, M., He, Y., Shu, N., 2015. Test-retest reliability of white matter structural brain networks: a multiband diffusion MRI study. *Front. Hum. Neurosci.* 9, 59.
- Zhou, H., Li, L., Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. *J. Am. Stat. Assoc.* 108 (502), 540–552.
- Zhu, J., Rosset, S., Hastie, T., Tibshirani, R., 2004. 1-norm support vector machines. *Adv. Neural Inf. Process. Syst.* 49–56.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. Royal Stat. Soc. Ser. B (Stat. Methodol.)* 67 (2), 301–320.