

# INTEGRATIVE NETWORK LEARNING FOR MULTI-MODALITY BIOMARKER DATA

BY SHANGHONG XIE<sup>1,\*</sup>, DONGLIN ZENG<sup>2</sup> AND YUANJIA WANG<sup>1,†</sup>

<sup>1</sup> Department of Biostatistics, Mailman School of Public Health, Columbia University, \*sx2168@cumc.columbia.edu;  
†yw2016@cumc.columbia.edu

<sup>2</sup> Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill,  
dzeng@email.unc.edu

The biomarker networks measured by different modalities of data (e.g., structural magnetic resonance imaging (sMRI), diffusion tensor imaging (DTI)) may share the same true underlying biological model. In this work, we propose a node-wise biomarker graphical model to leverage the shared mechanism between multi-modality data to provide a more reliable estimation of the target modality network and account for the heterogeneity in networks due to differences between subjects and networks of external modality. Latent variables are introduced to represent the shared unobserved biological network and the information from the external modality is incorporated to model the distribution of the underlying biological network. We propose an efficient approximation to the posterior expectation of the latent variables that reduces computational cost by at least 50%. The performance of the proposed method is demonstrated by extensive simulation studies and an application to construct gray matter brain atrophy network of Huntington's disease by using sMRI data and DTI data. The identified network connections are more consistent with clinical literature and better improve prediction in follow-up clinical outcomes and separate subjects into clinically meaningful subgroups with different prognosis than alternative methods.

**1. Introduction.** Network analysis is often used to learn interrelationships between biological measures (e.g., brain measures). A shared underlying biological network  $\mathcal{G}_0$  may give rise to multiple networks that are measured by different modalities of technology. Leveraging the shared mechanism, the estimation of a target network  $\mathcal{G}_*$  which may not be the same as the biological network, may be improved by borrowing information from an external and observable network  $\mathcal{G}_1$ . Our goal is to learn a covariate-dependent target network  $\mathcal{G}_*$  by incorporating features from external modality network  $\mathcal{G}_1$  that are observed at the individual level.

In the context of brain networks,  $\mathcal{G}_0$  can be structural connectivity which is the inter-regional anatomical associations between neural elements (e.g., axons). Studies have revealed the importance of structural connectivity on disease etiology and progression (Fornero, Zalesky and Bullmore (2016)). One approach to study structural connectivity is to construct structural co-variance network ( $\mathcal{G}_*$ ) which represents the co-variation patterns of the morphometric characteristics of grey matter regions (e.g., structural co-variance network of regional brain cortical thickness) from structural magnetic resonance imaging (sMRI) data (Alexander-Bloch, Giedd and Bullmore (2013)). Given that cortical thickness is regarded as an important early biomarker for disease progression, especially for Huntington's disease (HD; Rosas et al. (2005, 2008)), we regard structural co-variance network as the target network. However, another brain network used to describe structural connectivity is the white matter connectivity ( $\mathcal{G}_1$ ), which is obtained by reconstructing the trajectories of axonal fibers

---

*Keywords and phrases:* Graphical models, Network analysis, Multi-modality data, Structural co-variance network, White matter connectivity, Huntington's disease.

from white matter diffusion tensor imaging (DTI) data using tractography which is a 3D reconstruction technique used to visualize neural tracts for each individual.

Both approaches are exploited to understand the biologically meaningful characteristics of structural connectivity. The regions physically connected through white matter fibers may show stronger co-variation in their morphology. It has been documented that 30-40% of cortical thickness co-variance occurs between regions that manifest white matter connections ([Alexander-Bloch, Giedd and Bullmore \(2013\)](#)). White matter hyperintensities have also been shown to be associated with cortical thickness ([Rizvi et al. \(2018\)](#)). Borrowing information from white matter connectivity network  $\mathcal{G}_1$  may yield a more reliable estimate of structural co-variance network  $\mathcal{G}_*$ .

We propose a novel integrative learning approach for network estimation from multi-modality biomarker data. Specifically, we will borrow information from white matter connectivity  $\mathcal{G}_1$  to estimate grey matter structural co-variance network  $\mathcal{G}_*$  because they share the same underlying biological network  $\mathcal{G}_0$ . Our method makes two contributions: (a) incorporating information from an external modality in a biologically meaningful way and providing biological interpretation; and (b) accounting for between-subject heterogeneity in networks. We elaborate our contribution in the context of current literature below.

(a) First, several recent works have been proposed to estimate the functional connectivity from functional magnetic resonance imaging (fMRI) data by incorporating information from white matter connectivity. [Hinne et al. \(2014\)](#) modeled the posterior distribution of precision matrix for functional connectivity, assuming that the sparseness structure of the precision matrix is given by white matter connectivity. This assumption is restrictive since it assumes that functional connectivity exists only between regions that show white matter connections. Adaptive Gaussian graphical lasso models ([Ng et al. \(2012\)](#); [Pineda-Pardo et al. \(2014\)](#)) have been proposed to provide some room to allow the discordance between functional connectivity and white matter connectivity, by using connection-specific shrinkage parameter which is a function of fiber count between regions obtained from DTI data. However, such type of shrinkage parameters is arbitrary, sensitive to the functional form, and also lacks biological interpretation and meaning. In addition, these approaches do not account for the heterogeneity in functional connectivity resulting from other sources of variations beyond white matter connectivity. [Gong et al. \(2012\)](#) estimated the structural co-variance network and white matter connectivity separately, and compare their patterns. To the best of our knowledge, no method is available to estimate structural co-variance network by integrating information from white matter connectivity.

(b) Second, structural co-variance networks are strongly influenced by genetic factors ([Schmitt et al. \(2008\)](#)) and change with age ([He, Chen and Evans \(2008\)](#)) and disease status ([Seeley et al. \(2009\)](#)). However, current methods of constructing structural co-variance networks are at the population level, by calculating pairwise correlations between cortical thickness of brain regions, each measured once from each individual in a population ([Lerch et al. \(2006\)](#); [Alexander-Bloch et al. \(2013\)](#)). Gaussian graphical models have been studied extensively to construct high-dimensional complex networks ([Friedman, Hastie and Tibshirani \(2008\)](#); [Epskamp and Fried \(2018\)](#)), where the connection strength between two nodes is represented by the partial correlation between these nodes conditioning on the remaining nodes. However, Gaussian graphical models assume a homogeneous network for subjects in a population. When a population separates into known subgroups, fused graphical lasso (FGL; [Danaher, Wang and Witten \(2014\)](#)) was proposed to jointly estimate subgroup networks. However, when population heterogeneity and subgroups are unknown, FGL may not be applicable. [Hao et al. \(2018\)](#) and [Städler et al. \(2017\)](#) have extended the methods for unknown subgroups cases but assumed a small number of discrete subgroups.

To fill above gaps, we propose an integrative network learning approach to integrate multiple networks and account for between-subject heterogeneity. Our approach builds on an

unknown shared underlying biological network  $\mathcal{G}_0$  measured by multiple modalities. First, latent variables are introduced to represent the connections in the unobserved  $\mathcal{G}_0$ . The probability of connection in  $\mathcal{G}_0$  is modeled as a function of connections measured in the external network  $\mathcal{G}_1$  based on a network growth model (Vértes et al. (2012)) accounting for distance between nodes and degree of nodes for brain networks. Second, the connections between nodes in the target network  $\mathcal{G}_*$  is absent if there is no connection in  $\mathcal{G}_0$ . However, when there exists a connection in  $\mathcal{G}_0$ , the connection in  $\mathcal{G}_*$  is still allowed to be absent. This modeling framework captures the substantial but incomplete overlap between  $\mathcal{G}_*$  and  $\mathcal{G}_1$  of different modalities and also provides biologically meaningful interpretation that reflects  $\mathcal{G}_0$ . To address population heterogeneity, our approach provides subgroup- or subject-specific networks to capture the heterogeneity due to both subject's covariates  $X$  (e.g., genetic variants and age) and the networks of external modality  $\mathcal{G}_1$  (i.e., white matter connectivity) which are measured at the individual level. By including these individual networks to inform estimation of  $\mathcal{G}_*$ , the resolution of  $\mathcal{G}_*$  is expected to improve substantially.

For estimation, we propose an EM algorithm to infer latent connection statuses in  $\mathcal{G}_0$  from the observed data in two modalities. Through these latent connections, the estimation of  $\mathcal{G}_*$  will be improved by communicating with  $\mathcal{G}_1$  through  $\mathcal{G}_0$ . To use  $\mathcal{G}_0$  and  $\mathcal{G}_1$  to improve  $\mathcal{G}_*$ , the posterior means of  $\mathcal{G}_0$  given observed nodes  $M$  for target modality and  $\mathcal{G}_1$  need to be computed. However, this is a highly challenging problems due to the extensive number of potential connection patterns. For example, for a network with  $p$  nodes, the potential number of different combinations of latent connection statuses increases exponentially with the network size  $p$  and linearly with the sample size  $n$  (i.e., on the order of  $np(p-1)2^{p-2}$ ). When the number of nodes increases, it is prohibitive to compute these posterior probabilities directly. We propose a novel approximation to reduce the computational burden.

We conduct simulation studies to examine the performance of proposed method with varying number of nodes and sample sizes and compare with Gaussian graphical model and FGL. We apply the method to an observational study of HD, where both sMRI data and DTI data are available. We evaluate clinical utility of the identified connections in terms of prediction and stratification of patients' follow-up clinical outcomes. Incorporating connections measured at baseline in linear regression improves the prediction of the follow up motor outcomes, comparing to standard model and non-connection model which includes nodes but not their connections. The clusters identified by connections better distinguish between cluster motor and cognitive outcomes than clusters identified by regional nodes. We conclude the paper with some discussions and future directions.

**2. Methodology.** Our goal is to learn a covariate-dependent target network  $\mathcal{G}_*$  by incorporating features from external modality network  $\mathcal{G}_1$  that are available for each individual. We assume that  $\mathcal{G}_*$  and  $\mathcal{G}_1$  arise from the same underlying biological network  $\mathcal{G}_0$ . For the  $i$ th subject, let  $\mathbf{M}_i = (M_{i1}, M_{i2}, \dots, M_{ip})'$  be a  $p$ -dimensional vector that denotes measurements (centered at the mean) of  $p$  nodes in  $\mathcal{G}_*$  (e.g., sMRI cortical thickness of  $p$  brain regions). Let  $\mathbf{X}_i = (1, X_{i1}, X_{i2}, \dots, X_{iq})'$  denote a  $q + 1$ -dimensional vector including a constant of one and  $q$  exogenous covariates (e.g., age, genetic variants, and baseline clinical biomarkers) that may modify the edges (i.e., connections) in  $\mathcal{G}_*$  when the corresponding edges exist in  $\mathcal{G}_0$ .

**2.1. Model.** We consider a graphical model for estimating target network  $\mathcal{G}_*$ , in which the edge between two nodes depends on both observed covariates  $\mathbf{X}_i$  and a latent indicator of being connected in the shared biological network  $\mathcal{G}_0$  for each individual. We introduce latent variables  $B_{ijk}$  to indicate the presence of an edge between nodes  $j$  and  $k$  ( $j \neq k$ ) in  $\mathcal{G}_0$  for subject  $i$ , i.e.,  $B_{ijk} = 1$  indicates that there is an edge. A subject-specific graphical model

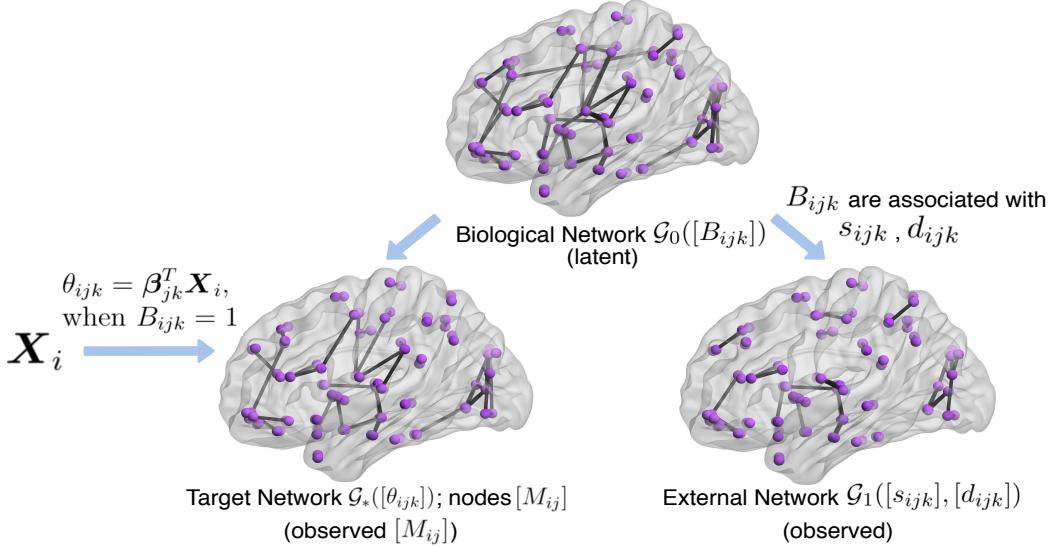


Fig 1: Schematics of the proposed method.  $\mathcal{G}_*$  and  $\mathcal{G}_1$  arise from an unknown shared network  $\mathcal{G}_0$ . We use  $\mathcal{G}_1$  to improve estimation of  $\mathcal{G}_*$  through their relationships to  $\mathcal{G}_0$ . An edge absent in  $\mathcal{G}_0$  is also absent in  $\mathcal{G}_*$ . An edge present in  $\mathcal{G}_0$  can still be absent in  $\mathcal{G}_*$  or have connection strength  $\theta_{ijk}$  depending on  $\mathbf{X}_i$  and  $\mathcal{G}_1$  through  $B_{ijk}$ , where  $B_{ijk}$  represents the presence or absence of an edge between nodes  $j$  and  $k$  in  $\mathcal{G}_0$  (see equation (1)).

(see Figure 1) with connection strength depends on both  $\mathbf{X}_i$  and  $B_{ijk}$  is that for node  $j$ ,

$$(1) \quad M_{ij} = \sum_{k \neq j} \theta_{ijk} M_{ik} + \varepsilon_{ij},$$

where  $\theta_{ijk}$  is modeled as

$$\theta_{ijk} = \begin{cases} 0, & B_{ijk} = 0, \\ \beta_{jk}^T \mathbf{X}_i, & B_{ijk} = 1. \end{cases}$$

Here  $\varepsilon_{ij}$  are independent of  $M_{ik}$  and follow  $N(0, \sigma_j^2)$ . Note that  $\beta_{jk}$  can still be zero if  $B_{ijk} = 1$ . An edge between nodes  $j$  and  $k$  is defined as present when  $\|\beta_{jk}\|_2 \neq 0$  or  $\|\beta_{kj}\|_2 \neq 0$  and the proportion of  $B_{ijk} = 1$  over all subjects is larger than a pre-specified value  $\alpha$ . To guarantee the identifiability of  $\beta_{jk}$ , we require that for any pair of nodes  $(j, k)$ , the covariate matrix for all subjects with biological connectivities between the nodes ( $B_{ijk} = 1$ ) is a full rank matrix. When  $B_{ijk}$  are observed, the model reduces to a covariate-dependent Gaussian graphical model.

Since the probability of observing an edge in  $\mathcal{G}_0$  is unknown, we estimate it by conditioning on the external network  $\mathcal{G}_1$ . Human brain networks exhibit small-worldness, organizes in a modular fashion, and the hub nodes have a fat-tailed degree distributions (Bullmore and Sporns (2009); He, Chen and Evans (2007)). To capture these observations, network growth models were recently used to model the structural co-variation network using cortical thickness (He, Chen and Evans (2007)), fMRI data on healthy subjects and Schizophrenia patients (Vértes et al. (2012)), and DTI network (Betzel et al. (2016)). Furthermore, Vértes et al. (2012) showed that modeling network growth as a function of anatomical distance alone is insufficient to capture the topological properties of functional brain networks and proposed to include an additional topological term (e.g., composite node degree or the number of nearest neighbors in common between two nodes). In addition, the degree distribution of the nodes

in white matter connectivity from DTI data in our real data application (Figure 2) is fat-tailed and shows similar trend as the one generated by network growth model depending on a topological term, see Vértes et al. (2012), Figure 1B (orange and red lines).

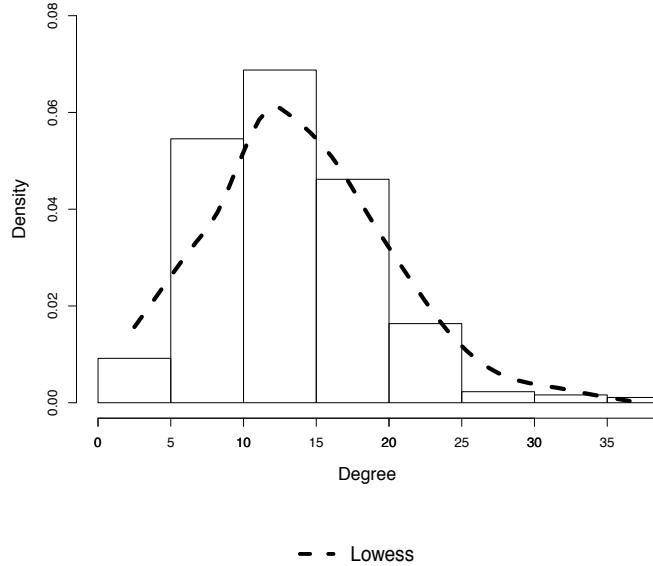


Fig 2: Degree distribution of the nodes with Lowess smoothing density curve in white matter connectivity network of our real data application.

Motivated from these prior studies, we propose a partially known network growth model which depends on the composite node degree (or the number of nearest neighbors in common between two nodes) and distances between nodes estimated from  $\mathcal{G}_1$  to model the pseudo probability of connection in  $\mathcal{G}_0$ . Specifically, we obtain a pseudo probability of edge  $B_{ijk}$  for node pairs  $(j, k)$  given  $\mathcal{G}_1$  as

$$(2) \quad p_{ijk} = P(B_{ijk} = 1 | s_{ijk}, d_{ijk}) = \frac{(1 + s_{ijk})^\gamma d_{ijk}^{-\eta}}{1 + (1 + s_{ijk})^\gamma d_{ijk}^{-\eta}},$$

where  $s_{ijk}$  is the composite node degree (i.e., the product of node  $j$ 's degree and node  $k$ 's degree) or the number of common neighbors of nodes  $j$  and  $k$ , and  $d_{ijk}$  is the anatomical distance between nodes  $j$  and  $k$  in external network  $\mathcal{G}_1$  of subject  $i$ . In this model,  $s_{ijk}$  and  $d_{ijk}$  are observed from  $\mathcal{G}_1$  and  $\gamma$  and  $\eta$  are non-negative parameters to be estimated from data. The probability of connecting two nodes in  $\mathcal{G}_0$  is increased when they belong to a large cluster (with a high degree) in  $\mathcal{G}_1$ , or when they are anatomically close to each other (within a short distance).

**2.2. Estimation.** For inference, we consider a node-wise conditional likelihood of the observed data at node  $j$  given the other nodes, treating  $f(\{B_{ijk}, k \neq j\} | D_i)$  as a prior distribution. Define  $D_i = (\{s_{ijk}\}_{j,k}, \{d_{ijk}\}_{j,k})$ , observed composite node degree and distance in  $\mathcal{G}_1$ . We assume that  $B_{ijk}$  are independent for nodes  $k \neq j$  given  $D_i$  and we incorporate  $\mathcal{G}_1$  through these prior distributions. The pseudo-likelihood for node  $j$  is

$$L_j = \prod_{i=1}^n \left\{ \sum_{B_{ijk} \in \{0,1\}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left( -\frac{(M_{ij} - \sum_{k \neq j} B_{ijk} \beta_{jk}^T \mathbf{X}_i M_{ik})^2}{2\sigma_j^2} \right) f(\{B_{ijk}, k \neq j\} | D_i) \right\}$$

$$= \prod_{i=1}^n \left\{ \sum_{B_{ijk} \in \{0,1\}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left( -\frac{(M_{ij} - \sum_{k \neq j} B_{ijk} \beta_{jk}^T \mathbf{X}_i M_{ik})^2}{2\sigma_j^2} \right) \prod_{k \neq j} p_{ijk}^{B_{ijk}} (1 - p_{ijk})^{1 - B_{ijk}} \right\},$$

where  $n$  is the sample size.

Treating  $B_{ijk}$  as missing data, EM algorithm is used to estimate the parameters  $(\beta, \sigma_j^2, \gamma, \eta)$ . An advantage of incorporating external networks as prior information by introducing  $B_{ijk}$  is to facilitate the EM algorithm development: since the 'complete data' distribution gives a regular likelihood so the EM ensures that the objective function increases over iterations. Next, we describe details of the EM algorithm.

**2.2.1. E-Step.** In the E-step, we compute the expectation of the log pseudo complete data likelihood given observed data  $M_{ij}$  at node  $j$  as

$$\begin{aligned} Q_j &= \sum_{i=1}^n E \left( \left\{ -\log(\sqrt{2\pi}\sigma_j) - \frac{1}{2\sigma_j^2} (M_{ij} - \sum_{k \neq j} B_{ijk} \beta_{jk}^T \mathbf{X}_i M_{ik})^2 \right\} \middle| M_{ij}, D_i \right) \\ &\quad + \sum_{i=1}^n \sum_{k \neq j} E \left( B_{ijk} \log(p_{ijk}) + (1 - B_{ijk}) \log(1 - p_{ijk}) \middle| M_{ij}, D_i \right) \\ &= -n \log(\sqrt{2\pi}\sigma_j) - \frac{1}{2\sigma_j^2} \sum_{i=1}^n M_{ij}^2 + \sum_{i=1}^n \sum_{k \neq j} \log(1 - p_{ijk}) \\ &\quad + \sum_{i=1}^n \sum_{k \neq j} \left( \frac{1}{\sigma_j^2} M_{ij} M_{ik} \beta_{jk}^T \mathbf{X}_i + \log \frac{p_{ijk}}{1 - p_{ijk}} \right) E(B_{ijk}|M_{ij}, D_i) \\ &\quad - \frac{1}{2\sigma_j^2} \sum_{i=1}^n \sum_{k \neq j} (\beta_{jk}^T \mathbf{X}_i) (\beta_{jk'}^T \mathbf{X}_i) M_{ik} M_{ik'} E(B_{ijk} B_{ijk'}|M_{ij}, D_i). \end{aligned}$$

To compute the expected likelihood given observed data, we need to obtain the posterior expectations of connectivity indicators  $E[B_{ijk}|M_{ij}, D_i]$  and  $E[B_{ijk} B_{ijk'}|M_{ij}, D_i]$ , which can be expressed as

$$\begin{aligned} E(B_{ijk}|M_{ij}, D_i) &= P(B_{ijk} = 1|M_{ij}, D_i) \\ &= \frac{P(B_{ijk} = 1, M_{ij}|D_i)}{P(B_{ijk} = 1, M_{ij}|D_i) + P(B_{ijk} = 0, M_{ij}|D_i)}, \end{aligned}$$

$$E(B_{ijk}^2|M_{ij}, D_i) = E(B_{ijk}|M_{ij}, D_i),$$

and for  $k' \neq k$ ,

$$\begin{aligned} E(B_{ijk} B_{ijk'}|M_{ij}, D_i) &= P(B_{ijk} = 1, B_{ijk'} = 1|M_{ij}, D_i) \\ &= \frac{P(B_{ijk} = 1, B_{ijk'} = 1, M_{ij}|D_i)}{\sum_{b,b' \in \{0,1\}} P(B_{ijk} = b, B_{ijk'} = b', M_{ij}|D_i)}. \end{aligned}$$

Then we obtain

$$\begin{aligned}
& P(B_{ijk} = b, M_{ij}|D_i) \\
&= \sum_{b_{ij1}, \dots, b_{ijp}} P(B_{ij1} = b_{ij1}, \dots, B_{ijk} = b, \dots, B_{ijp} = b_{ijp}, M_{ij}|D_i) \\
&= \left( \prod_{k \neq j} (1 - p_{ijk}) \right) \times \exp \left( b \log \left( \frac{p_{ijk}}{1 - p_{ijk}} \right) + \sum_{l \neq j, k} b_{ijl} \log \left( \frac{p_{ijl}}{1 - p_{ijl}} \right) \right) \times \\
&\quad \sum_{b_{ij1}, \dots, b_{ijp}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left( -\frac{(M_{ij} - b\beta_{jk}^T \mathbf{X}_i M_{ik} - \sum_{l \neq j, k} b_{ijl} \beta_{jl}^T \mathbf{X}_i M_{il})^2}{2\sigma_j^2} \right),
\end{aligned}$$

where  $b \in \{1, 0\}$  and  $b_{ijl} \in \{1, 0\}$  for  $l \neq j, k$ . In addition,

$$\begin{aligned}
& P(B_{ijk} = b, B_{ijk'} = b', M_{ij}|D_i) \\
&= \sum_{b_{ij1}, \dots, b_{ijp}} P(B_{ij1} = b_{ij1}, \dots, B_{ijk} = b, B_{ijk'} = b', \dots, B_{ijp} = b_{ijp}, M_{ij}|M_{ij}, D_i) \\
&= \left( \prod_{k \neq j} (1 - p_{ijk}) \right) \times \exp \left( b \log \left( \frac{p_{ijk}}{1 - p_{ijk}} \right) + b' \log \left( \frac{p_{ijk'}}{1 - p_{ijk'}} \right) + \sum_{l \neq j, k, k'} b_{ijl} \log \left( \frac{p_{ijl}}{1 - p_{ijl}} \right) \right) \times \\
&\quad \sum_{b_{ij1}, \dots, b_{ijp}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left( -\frac{(M_{ij} - b\beta_{jk}^T \mathbf{X}_i M_{ik} - b'\beta_{jk'}^T \mathbf{X}_i M_{ik'} - \sum_{l \neq j, k, k'} b_{ijl} \beta_{jl}^T \mathbf{X}_i M_{il})^2}{2\sigma_j^2} \right),
\end{aligned}$$

where  $b \in \{1, 0\}$ ,  $b' \in \{1, 0\}$  and  $b_{ijl} \in \{1, 0\}$  for  $l \neq j, k, k'$ .

The computational complexity of the posterior expectations  $E(B_{ijk}|M_{ij}, D_i)$  and  $E(B_{ijk}B_{ijk'}|M_{ij}, D_i)$  increases exponentially with the network size  $p$  due to exponential number of terms in the summation of probabilities. The summation in  $P(B_{ijk} = b, M_{ij}|D_i)$  and  $P(B_{ijk} = b, B_{ijk'} = b', M_{ij}|D_i)$  can be calculated by (1) exhaustive enumeration (Direct); (2) Markov Chain Monte Carlo (MCMC); or (3) approximation. The first two methods are computationally expensive and may not be feasible for large networks. In this case, we propose an approximation algorithm. Details of the approximation are given in the Supplement A.2 of the Supplementary Material ([Xie, Zeng and Wang \(2020\)](#)).

**2.2.2. M-Step.** In the M-step, we maximize the expected likelihood  $Q_j$  replacing latent connection statuses by their conditional expectations,  $E(B_{ijk}|M_{ij}, D_i)$  and  $E(B_{ijk}B_{ijk'}|M_{ij}, D_i)$ , computed from the E-step. To estimate parameters  $\gamma$  and  $\eta$  under model (2), we perform a grid search to find the optimal combination that results in the largest log pseudo complete data likelihood for node  $j$  and then take the average over all the nodes. For each combination of  $\gamma$  and  $\eta$ , we maximize  $Q_j$  and solve the following equations for each node  $j$  separately to estimate parameters  $\beta_{jk}$  and  $\sigma_j^2$

$$\frac{\partial Q_j}{\partial \beta_{jk}} = \frac{1}{n\sigma_j^2} \sum_{i=1}^n \left[ M_{ij} M_{ik} E(B_{ijk}|M_{ij}, D_i) \mathbf{X}_i^T - \sum_{k' \neq j} \beta_{jk'}^T M_{ik'} M_{ik} E(B_{ijk}B_{ijk'}|M_{ij}, D_i) \mathbf{X}_i \mathbf{X}_i^T \right],$$

and

$$\begin{aligned} \frac{\partial Q_j}{\partial \sigma_j^2} = & -\frac{1}{2\sigma_j^2} + \frac{1}{2n(\sigma_j^2)^2} \sum_{i=1}^n M_{ij}^2 - \frac{1}{(n\sigma_j^2)^2} \sum_{i=1}^n \sum_{k \neq j} E(B_{ijk}|M_{ij}, D_i) M_{ij} M_{ik} \beta_{jk}^T \mathbf{X}_i \\ & + \frac{1}{2n(\sigma_j^2)^2} \sum_{i=1}^n \sum_{k \neq j} (\beta_{jk}^T \mathbf{X}_i) M_{ik} \sum_{k' \neq k, j} (\beta_{jk'}^T \mathbf{X}_i) M_{ik'} E(B_{ijk} B_{ijk'}|M_{ij}, D_i) \\ & + \frac{1}{2n(\sigma_j^2)^2} \sum_{i=1}^n \sum_{k \neq j} (\beta_{jk}^T \mathbf{X}_i)^2 M_{ik}^2 E(B_{ijk}|M_{ij}, D_i), \end{aligned}$$

and compute

$$\begin{aligned} \hat{\sigma}_j^2 = & \frac{1}{n} \sum_{i=1}^n M_{ij}^2 - \\ & \frac{1}{n} \sum_{i=1}^n \sum_{k \neq j} \left[ 2E(B_{ijk}|M_{ij}, D_i) M_{ij} - \sum_{k' \neq k, j} (\hat{\beta}_{jk'}^T \mathbf{X}_i) M_{ik'} E(B_{ijk} B_{ijk'}|M_{ij}, D_i) \right. \\ & \left. - E(B_{ijk}|M_{ij}, D_i) (\hat{\beta}_{jk}^T \mathbf{X}_i) M_{ik} \right] M_{ik} (\hat{\beta}_{jk}^T \mathbf{X}_i). \end{aligned}$$

Iterations in the EM algorithm will terminate when the estimates of  $\beta_{jk}$ 's and  $\sigma_j^2$  converge.

To remove spurious edges, hard-thresholding is performed on  $\beta_{jk}$  based on EBIC criterion (Foygel and Drton (2011); Chen and Chen (2008)). EBIC for node  $j$  is defined as:

$$EBIC_j = -2 \log(L_j(\beta_{jk}, \sigma_j^2, \hat{\gamma}, \hat{\eta})) + E \times \log(n) + 2\delta E \times \log(p-1),$$

where  $E$  is the number of non-zero  $\|\beta_{jk}\|_2$ ,  $\delta$  is a hyperparameter, and  $\hat{\gamma}$  and  $\hat{\eta}$  are estimated from EM algorithm. The estimates  $\hat{\beta}_{jk}$  and the corresponding  $\hat{\sigma}_j^2$  that minimize  $EBIC_j$  are selected. An edge between nodes  $j$  and  $k$  is defined as present when  $\|\hat{\beta}_{jk}\|_2$  or  $\|\hat{\beta}_{kj}\|_2$  is non-zero and the average of  $\hat{E}(B_{ijk}|M_{ij}, D_i)$  over all subjects is larger than a pre-specified threshold value  $\alpha$ .

### 3. Simulation Studies.

**3.1. Simulation Design.** We conducted extensive simulations to evaluate the proposed method. We varied the number of biomarker nodes,  $p = 5, 10$  and the sample size,  $n = 200, 400$  with the number of covariates  $q = 3$ . Four settings were considered. In Settings 1-2, target network varied only by subject-specific external network. In Settings 3-4, target network varied by both covariates and subject-specific external network. In all settings, four out of ten  $\beta_{jk}$ 's ( $\beta_{jk}$  and  $\beta_{kj}$  are considered as a pair,  $\beta_{jk}/\sigma_j^2 = \beta_{kj}/\sigma_k^2$  and counted only once) are non-zeros when  $p = 5$  and eight out of 45  $\beta_{jk}$ 's are non-zeros when  $p = 10$  (Figure 3a). We refer  $\beta_{jk}$  with a large  $L_2$  norm to as 'strong signal' and that with a small  $L_2$  norm to as 'weak signal'. The values of  $\beta_{jk}$  and  $\sigma_j$  are presented in Supplement B of the Supplementary Material (Xie, Zeng and Wang (2020)). We simulated covariates  $\mathbf{X}_i$  independently from  $N(0, 1)$  truncated between -1 and 1.

In Settings 1-2,  $\beta_{jk}$  is a scalar instead of a vector, representing population average edge strength between nodes  $j$  and  $k$  among subjects with  $B_{ijk} = 1$ .

Setting 1 (Varied by external modality network alone, Weak signal):

- When  $p = 5$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Left panel) have  $\|\beta_{jk}\|_2 = 0.45$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.55$  when  $j$  is even;

- When  $p = 10$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Left panel) have  $\|\beta_{jk}\|_2 = 0.55$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.63$  when  $j$  is even.

Setting 2 (Varied by external modality network alone, Strong signal):

- When  $p = 5$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Left panel) have  $\|\beta_{jk}\|_2 = 0.67$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.82$  when  $j$  is even;
- When  $p = 10$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Left panel) have  $\|\beta_{jk}\|_2 = 0.82$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.95$  when  $j$  is even.

In the external network  $\mathcal{G}_1$ , we let five edges be connected when  $p = 5$  and ten edges be connected when  $p = 10$  for each subject. Subjects with  $X_{i1} > 0$  (52% of subjects) had an additional edge between nodes 1 and 2 when  $p = 5$  and edge between nodes 4 and 5 when  $p = 10$ . Subjects with  $X_{i3} > 0.8$  (8.55% of subjects) had an additional edge between nodes 3 and 5 when  $p = 5$  and edge between nodes 8 and 9 when  $p = 10$ . The frequency of edges appearing in  $\mathcal{G}_1$  among subjects is shown in Figure 3b, Left panel.

Setting 3 (Covariate-dependent, Weak signal):

- When  $p = 5$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Right panel) have  $\|\beta_{jk}\|_2 = 0.17$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.26$  when  $j$  is even;
- When  $p = 10$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Right panel) have  $\|\beta_{jk}\|_2 = 0.26$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.35$  when  $j$  is even.

Setting 4 (Covariate-dependent, Strong signal):

- When  $p = 5$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Right panel) have  $\|\beta_{jk}\|_2 = 0.35$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.52$  when  $j$  is even;
- When  $p = 10$ , non-zero  $\beta_{jk}$ 's (Figure 3a, Right panel) have  $\|\beta_{jk}\|_2 = 0.52$  when  $j$  is odd and have  $\|\beta_{jk}\|_2 = 0.69$  when  $j$  is even.

In the external modality network  $\mathcal{G}_1$ , the edges existed in all subjects are as the same as those in Settings 1-2. To introduce between-subject variation, there exists an extra edge between nodes 4 and 5 if subjects have  $X_{i1} > 0$  (52% of subjects) and an edge between nodes 3 and 5 if subjects have  $X_{i3} > 0$  (52% of subjects). The frequency of edges appearing in  $\mathcal{G}_1$  among subjects is shown in Figure 3b, Right panel. The distance between two nodes ranges from 0.1 to 2.  $s_{ijk}$  is the number of common neighbors between two nodes. We set the degree parameter  $\gamma = 2$  and the distance parameter  $\eta = 1$ .

We assessed the edge selection performance by AUC of varying both hard-thresholding values and the pre-specified proportion values  $\alpha$ , the number of true positive (TP) edges, false positive (FP) edges, true negative (TN) edges, false negative (FN) edges, sensitivity, specificity, and the performance of parameter estimations by mean squared error (MSE). The lower and upper thresholds of  $E(B_{ijk}|M_{ij}, D_i)$  introduced in Supplement A.1 of the Supplementary Material (Xie, Zeng and Wang (2020)) were set to be 0.1 and 0.9, respectively. We fixed  $\eta = 1$  to save computation time and selected  $\Gamma = \{1.8, 1.9, 2.0, 2.1, 2.2\}$  for grid search of  $\gamma$ . The number of hard-thresholding cutoff values is 50 and the values are based on the magnitude of  $\hat{\beta}_{jk}$ 's. The proportion values  $\alpha$ 's are  $\{0, 0.1, 0.2, 0.3, 0.4\}$ . In the approximation version of the proposed method, we used the exponential basis functions with  $(\lambda_1, \dots, \lambda_{18}) = (0.001, 0.002, 0.005, 0.02, 0.1, 0.2, 0.5, 1, 2, 3, 4, 5, 8, 10, 16, 25, 32, 64)$ , generated  $u_i$  uniformly in  $[-1.5, 1.5]$  and obtained the least square estimates of the basis function coefficients  $\zeta$ .

We compared our method with EBIC glasso which is the Gaussian graphical model with lasso penalty and the tuning parameter is chosen by EBIC criterion (Foygel and Drton (2010)) and FGL using EBIC (defined in Supplement C of the Supplementary Material (Xie, Zeng and Wang (2020))) criterion. The length of the searching path of the tuning parameter for

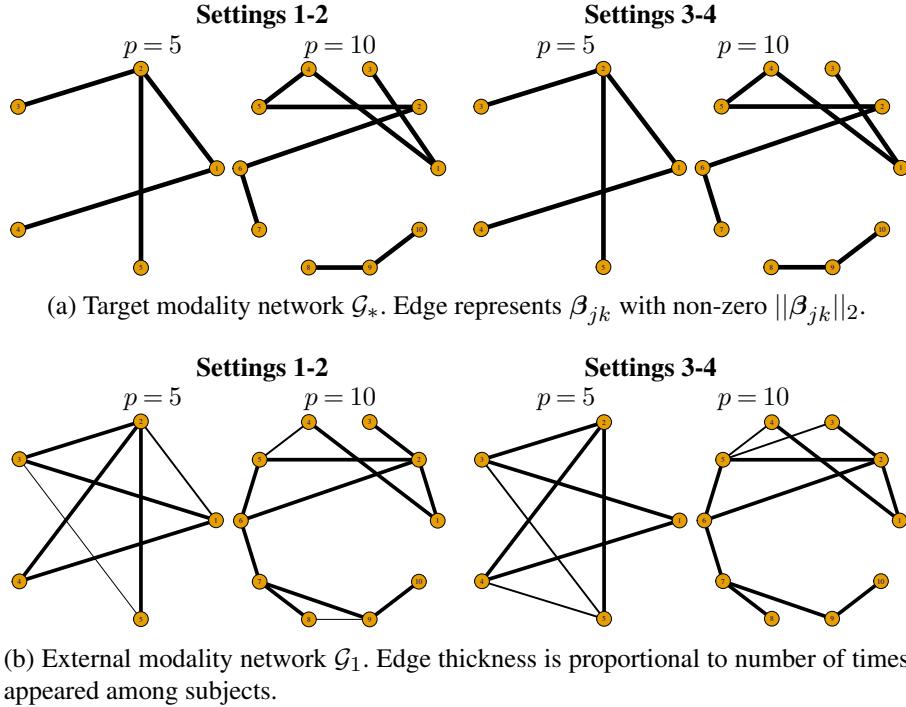


Fig 3: Networks in the simulation settings.

EBIC glasso is 50. FGL requires the subgroups are known. We stratified sample into two subgroups based on the median value of the first covariate  $X_{i1}$ , which is consistent with the true simulation model. The tuning parameter for Lasso penalty was searched from 0 to 0.2 and that for penalty on encouraging similarity between subgroups was searched from 0 to 0.1 as suggested in [Danaher, Wang and Witten \(2014\)](#). We set the hyperparameter  $\delta$  in EBIC formula to be 0.5 for all methods, as suggested in [Epskamp and Fried \(2018\)](#). The simulations were repeated 100 times.

**3.2. Simulation Results.** In Settings 1-2, we demonstrated the improvement of performance by incorporating external network alone. Our method selected all the true edges with high frequency across 100 simulations, while the frequency of true edges selected by EBIC glasso was lower and with non-negligible false positives and FGL selected true edges more frequently than EBIC glasso but also along with more null-edges (Figure 4). When  $p = 5$  (Table 1), sensitivity increased from 18.5% to 51.5% by incorporating external modality information in Setting 1 when  $n = 200$  and increased from 65.3% to 80.5% when  $n = 400$ . AUC of our method was 0.9 when  $n = 200$  and increased to 0.98 when  $n = 400$ , while that of EBIC glasso was 0.86 when  $n = 200$  and 0.96 when  $n = 400$ , and that of FGL was 0.88 when  $n = 200$  and 0.97 when  $n = 400$ . When the signal was strong (Setting 2), our method identified almost all the true edges and selected fewer null edges than EBIC glasso and FGL. AUCs of the three methods were all larger than 0.95 but our method still outperformed EBIC glasso and FGL, achieving AUC=1 when  $n = 400$ . When  $p = 10$  (Table 2), our method achieved 78.1% sensitivity comparing to 51.6% sensitivity by EBIC glasso and 46.6% sensitivity for FGL in Setting 1 with  $n = 200$ . In Setting 1 with  $n = 400$  and Setting 2, all the three methods achieved high sensitivity (almost 100%) while our method selected much fewer null edges than EBIC glasso and FGL. These results suggest that external network improves learning in terms of correctly identifying more non-null edges and pruning more null edges.

TABLE 1

*Selection Performance of Simulations for  $p = 5$  in Settings 1-2. The best performance for each row is bolded.*

	Proposed Method		EBIC glasso	FGL
	$\alpha \leq 0.4$		$\alpha \leq 0.4$	$\alpha \leq 0.4$
	Direct	Approx.		
<b>Setting 1: External Modality Alone, Weak Signal</b>				
<i>n = 200</i>				
TP <sup>1</sup>	<b>2.06</b>	1.98	0.74	1.76
FP <sup>2</sup>	0.14	0.13	<b>0.07</b>	0.92
TN <sup>3</sup>	5.86	5.87	<b>5.93</b>	5.08
FN <sup>4</sup>	<b>1.94</b>	2.02	3.26	2.24
Sensitivity <sup>6</sup>	<b>0.515</b>	0.495	0.185	0.440
Specificity <sup>7</sup>	0.977	0.978	<b>0.988</b>	0.847
<i>n = 400</i>				
TP <sup>1</sup>	<b>3.22</b>	3.20	2.61	2.83
FP <sup>2</sup>	<b>0.09</b>	<b>0.09</b>	0.42	1.20
TN <sup>3</sup>	<b>5.91</b>	<b>5.91</b>	5.58	4.80
FN <sup>4</sup>	<b>0.78</b>	0.80	1.39	1.17
Sensitivity <sup>6</sup>	<b>0.805</b>	0.800	0.653	0.708
Specificity <sup>7</sup>	<b>0.985</b>	<b>0.985</b>	0.930	0.800
<b>Setting 2: External Modality Alone, Strong Signal</b>				
<i>n = 200</i>				
TP <sup>1</sup>	<b>3.66</b>	3.63	3.49	2.68
FP <sup>2</sup>	<b>0.18</b>	0.20	0.88	0.78
TN <sup>3</sup>	<b>5.82</b>	5.80	5.12	5.22
FN <sup>4</sup>	<b>0.34</b>	0.37	0.51	1.32
Sensitivity <sup>6</sup>	<b>0.915</b>	0.908	0.873	0.670
Specificity <sup>7</sup>	<b>0.970</b>	0.967	0.853	0.870
<i>n = 400</i>				
TP <sup>1</sup>	3.99	3.99	<b>4.00</b>	3.62
FP <sup>2</sup>	<b>0.07</b>	0.08	1.12	0.95
TN <sup>3</sup>	<b>5.93</b>	5.92	4.88	5.05
FN <sup>4</sup>	0.01	0.01	<b>0.00</b>	0.38
Sensitivity <sup>6</sup>	0.998	0.998	<b>1.000</b>	0.905
Specificity <sup>7</sup>	<b>0.988</b>	0.987	0.813	0.842

<sup>1</sup>TP: Average number of true positive edges across 100 simulations; <sup>2</sup>FP: Average number of false positive edges across 100 simulations; <sup>3</sup> TN: Average number of true negative edges across 100 simulations; <sup>4</sup> FN: Average number of false negative edges across 100 simulations; <sup>5</sup>: Average false discovery rate across 100 simulations; <sup>6</sup>: Average sensitivity across 100 simulations; <sup>7</sup>: Average specificity across 100 simulations.

When the target network was also covariate-dependent (Settings 3-4), AUC of our method was much higher than EBIC glasso and FGL. In Setting 3 with  $p = 5$  (Figure 5, Top row, Left panel), AUC of our method using either direct calculation or approximation was 0.62 when  $n = 200$  and achieved 0.68 when  $n = 400$ . In Setting 4 with strong signal (Figure 5, Top row, Right panel), AUC of our method was 0.78 when  $n = 200$  and 0.91 when  $n = 400$ . AUCs of EBIC glasso and FGL for all settings ranged from 0.5 to 0.6.

Table 3 shows the selection performance of the simulation studies when  $p = 5$ . In Setting 3 (weak signal), the sensitivity of our method was about 30% while that of EBIC glasso was

TABLE 2

*Selection Performance of Simulations for  $p = 10$  in Settings 1-2. The best performance for each row is bolded.*

	Proposed Method				EBIC glasso		FGL	
	$\alpha \leq 0.3$		$\alpha = 0.4$		$\alpha \leq 0.3$	$\alpha = 0.4$	$\alpha \leq 0.3$	$\alpha = 0.4$
	Direct	Approx.	Direct	Approx.				
<b>Setting 1: External Modality Alone, Weak Signal</b>								
<i>n = 200</i>								
TP <sup>1</sup>	<b>6.25</b>	6.15	6.10	6.00	4.13	4.08	3.73	3.66
FP <sup>2</sup>	0.97	1.23	0.72	0.89	<b>0.39</b>	0.44	3.82	3.89
TN <sup>3</sup>	36.03	35.77	36.57	36.40	36.61	<b>36.85</b>	33.18	33.40
FN <sup>4</sup>	1.75	1.85	<b>1.61</b>	1.71	3.87	3.63	4.27	4.05
Sensitivity <sup>6</sup>	0.781	0.769	<b>0.791</b>	0.778	0.516	0.527	0.466	0.472
Specificity <sup>7</sup>	0.974	0.967	0.981	0.976	<b>0.989</b>	0.988	0.897	0.896
<i>n = 400</i>								
TP <sup>1</sup>	7.51	7.51	7.34	7.34	7.67	7.50	<b>7.76</b>	7.57
FP <sup>2</sup>	<b>0.44</b>	0.65	<b>0.44</b>	0.59	1.40	1.57	8.98	9.17
TN <sup>3</sup>	36.56	36.35	<b>36.78</b>	36.63	35.60	35.65	28.02	28.05
FN <sup>4</sup>	0.49	0.49	0.44	0.44	0.33	0.28	0.24	<b>0.21</b>
Sensitivity <sup>6</sup>	0.939	0.939	0.944	0.944	0.959	0.964	0.970	<b>0.972</b>
Specificity <sup>7</sup>	<b>0.988</b>	0.982	<b>0.988</b>	0.984	0.962	0.958	0.757	0.754
<b>Setting 2: External Modality Alone, Strong Signal</b>								
<i>n = 200</i>								
TP <sup>1</sup>	<b>8.00</b>	<b>8.00</b>	7.71	7.69	<b>8.00</b>	7.71	<b>8.00</b>	7.71
FP <sup>2</sup>	<b>0.79</b>	2.75	0.87	2.75	8.17	8.46	14.23	14.52
TN <sup>3</sup>	36.21	34.25	<b>36.42</b>	34.56	28.83	28.83	22.77	22.77
FN <sup>4</sup>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sensitivity <sup>6</sup>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Specificity <sup>7</sup>	<b>0.979</b>	0.926	0.977	0.926	0.779	0.773	0.615	0.611
<i>n = 400</i>								
TP <sup>1</sup>	<b>8.00</b>	<b>8.00</b>	7.78	7.76	<b>8.00</b>	7.78	<b>8.00</b>	7.78
FP <sup>2</sup>	<b>0.73</b>	3.32	0.83	3.17	9.06	9.28	16.68	16.90
TN <sup>3</sup>	36.27	33.68	<b>36.39</b>	34.07	27.94	27.94	20.32	20.32
FN <sup>4</sup>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sensitivity <sup>6</sup>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Specificity <sup>7</sup>	<b>0.980</b>	0.910	0.978	0.915	0.755	0.751	0.549	0.546

<sup>1</sup>TP: Average number of true positive edges across 100 simulations; <sup>2</sup>FP: Average number of false positive edges across 100 simulations; <sup>3</sup> TN: Average number of true negative edges across 100 simulations; <sup>4</sup> FN: Average number of false negative edges across 100 simulations; <sup>5</sup>: Average false discovery rate across 100 simulations; <sup>6</sup>: Average sensitivity across 100 simulations; <sup>7</sup>: Average specificity across 100 simulations.

only about 3% and that of FGL was smaller than 25%. In Setting 4 (strong signal), our method achieved above 70% sensitivity comparing to EBIC glasso with 4% sensitivity and FGL with 18.5% sensitivity when  $n = 200$ . When  $n$  increases to 400, our method almost identified all the true edges with 93.3% sensitivity and reasonable numbers of null edges with about 73% specificity, whereas the sensitivity of EBIC glasso was only 4.5% and that of FGL was 25%.

Similar results are seen when  $p = 10$  (Table 4). The sensitivity of our method was about 30% while that of EBIC glasso was just 0.6% and that of FGL was 3.4% when  $n = 400$  in Setting 3. In Setting 4, our method almost identified all the true edges, whereas both EBIC glasso and FGL detected fewer than 1 true edges. Our method achieved 87.8% sensitivity, meanwhile attained 77.4% specificity when  $n = 200$  and further improved them when  $n =$

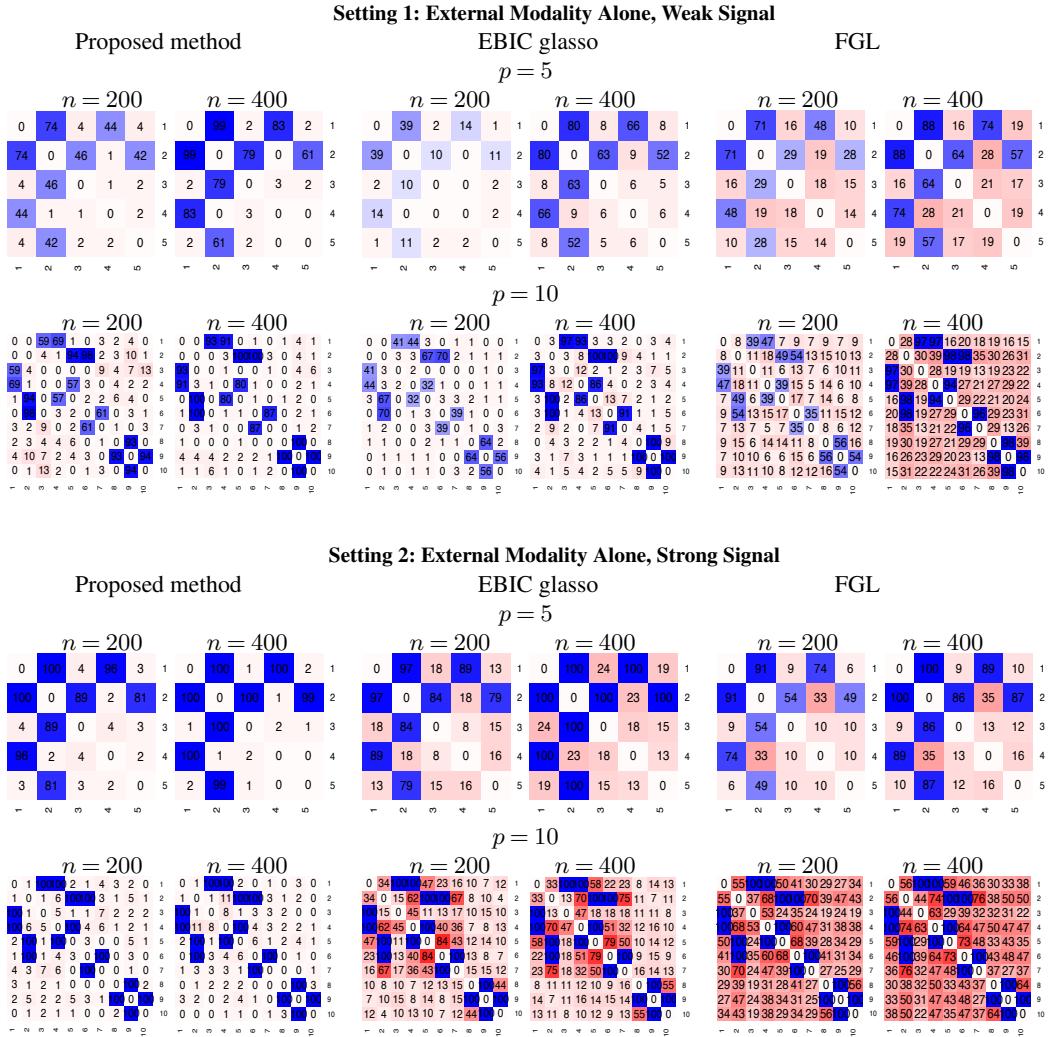


Fig 4: The frequency of edges selected in estimated target networks in Settings 1-2. Each cell represents the number of times an edge between two nodes was selected across 100 simulations. Blue cell: true positive edges; Red cell: false positive edges. Darker color means higher frequency of being selected. Our method outperformed EBIC glasso and FGL in terms of high frequency in selecting true edges and low frequency of null-edges.

400. AUCs were 0.65 and 0.75 for  $n = 200$  and  $n = 400$  respectively (Figure 5, Bottom row, Left panel). Improvement of performance is seen when the signal was strong (Figure 5, Bottom row, Right panel). AUC of our method was above 0.9 when  $n = 200$ , and increased to 0.98 when  $n = 400$ . AUCs of EBIC glasso and FGL for all settings ranged from 0.5 to 0.6. EBIC glasso correctly selected true edges less than 10 times across 100 simulations even when the signal was strong and  $n = 400$ . FGL correctly selected true edges more frequently than EBIC glasso but selected more null-edges. In contrast, our method selected all true edges with a high frequency (greater than 95 times out of 100 simulations; Figure 6). We also notice that the performance is not sensitive to  $\alpha$ . The results of  $\alpha = 0, 0.1, 0.2, 0.3$  were the same and those of  $\alpha = 0.4$  were slightly different.

In terms of parameter estimation, MSE of  $\beta_{jk}$  ranged from 0.0014 to 0.0173 in Settings 1 and 2 and ranged from 0.025 to 0.09 in Settings 3 and 4 (Table 5).

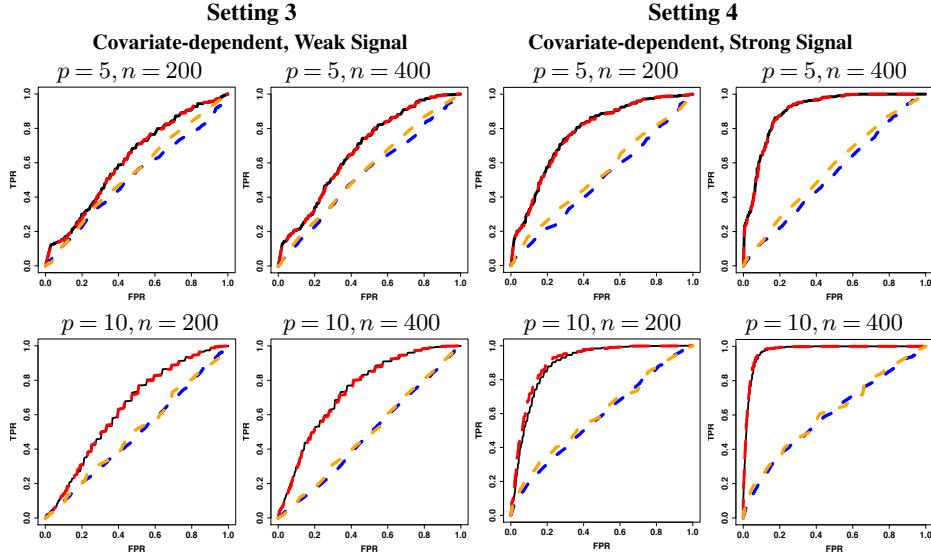


Fig 5: ROC curves of the simulations in Settings 3-4. Black solid line: Proposed Method (Direct); Red long-dash line: Proposed Method (Approximation); Blue dash line: EBIC glasso; Orange dash line: FGL. Our method provided higher AUC than EBIC glasso and FGL.

The major difference between the proposed method and EBIC glasso is that EBIC glasso estimates the edge strength between two nodes at the population level and does not account for covariate-dependent or subject-specific connection status. When  $B_{ijk} = 1$ ,  $\theta_{ijk} = \beta_{jk}^T \mathbf{X}_i$  can be either positive or negative depending on covariates  $\mathbf{X}_i$ . When  $B_{ijk} = 0$ ,  $\theta_{ijk} = 0$  directly. Therefore, the average edge signal at the population is low for many edges. This may explain the poor performance of EBIC glasso. FGL performed better than EBIC glasso because it incorporated covariate information when we stratified sample into subgroups based on covariates. However, it requires that the subgroups are pre-specified which is not flexible and one may not know how to stratify samples in practice.

**3.3. Computational Advantage of the Approximation.** The results of our method with approximation were comparable to those of direct calculation when  $n = 200$ , and they were very similar when  $n = 400$ , in terms of both selection performance in the final model (Tables 1, 2, 3, and 4) and ROC curves (Figure 5). Although losing some accuracy with low sample size, computational efficiency was gained when we used approximation to calculate the posterior expectations (Table 6) in all cases. All calculations were carried out on an Intel Core i7 2.8 GHz processor. The average running time per simulation of using approximation reduced by about 50% when  $n = 200$  comparing to the direct calculation. The approximation approach cost similar time for  $n = 400$  and  $n = 200$ , thus the running time reduced by more than 66% comparing to direct calculation when  $n = 400$ .

**4. Application to the Grey Matter Atrophy Network of HD.** Huntington's Disease (HD), a progressive genetic neurodegenerative disease, is caused by an inherited CAG repeat expansion in the huntingtin gene and is characterized by cognitive, motor and psychiatric symptoms (Paulsen et al. (2008, 2006); Klöppel et al. (2015)). Studies have shown that cortical thinning begins a decade before disease onset (Rosas et al. (2005); Tabrizi et al. (2009)). The grey matter structural co-variance networks are reported to differ among healthy controls, preHD, and HD patients, and might be an early biomarker for HD (Coppen et al. (2016)). Our

TABLE 3

*Selection Performance of Simulations for  $p = 5$  in Settings 3-4. The best performance for each row is bolded.*

	Proposed Method		EBIC glasso		FGL	
	$\alpha \leq 0.3$		$\alpha = 0.4$		$\alpha \leq 0.4$	
	Direct	Approx.	Direct	Approx.		
<b>Setting 3: Covariate-dependent, Weak Signal</b>						
<i>n = 200</i>						
TP <sup>1</sup>	<b>1.19</b>	1.03	<b>1.19</b>	1.03	0.11	0.50
FP <sup>2</sup>	1.07	0.98	0.64	0.62	<b>0.12</b>	0.76
TN <sup>3</sup>	4.93	5.02	5.36	5.38	<b>5.88</b>	5.24
FN <sup>4</sup>	<b>2.81</b>	2.97	<b>2.81</b>	2.97	3.89	3.50
Sensitivity <sup>6</sup>	<b>0.298</b>	0.258	<b>0.298</b>	0.258	0.028	0.125
Specificity <sup>7</sup>	0.822	0.837	0.893	0.897	<b>0.980</b>	0.873
<i>n = 400</i>						
TP <sup>1</sup>	<b>1.35</b>	1.28	<b>1.35</b>	1.28	0.16	0.96
FP <sup>2</sup>	0.84	0.80	0.45	0.43	<b>0.16</b>	1.28
TN <sup>3</sup>	5.16	5.20	5.55	5.57	<b>5.84</b>	4.72
FN <sup>4</sup>	<b>2.65</b>	2.72	<b>2.65</b>	2.72	3.84	3.04
Sensitivity <sup>6</sup>	<b>0.338</b>	0.320	<b>0.338</b>	0.320	0.040	0.240
Specificity <sup>7</sup>	0.860	0.867	0.925	0.928	<b>0.973</b>	0.787
<b>Setting 4: Covariate-dependent, Strong Signal</b>						
<i>n = 200</i>						
TP <sup>1</sup>	<b>3.03</b>	2.84	<b>3.03</b>	2.84	0.16	0.74
FP <sup>2</sup>	1.81	1.72	1.29	1.17	<b>0.11</b>	0.90
TN <sup>3</sup>	4.19	4.28	4.71	4.83	<b>5.89</b>	5.10
FN <sup>4</sup>	<b>0.97</b>	1.16	<b>0.97</b>	1.16	3.84	3.26
Sensitivity <sup>6</sup>	<b>0.758</b>	0.710	<b>0.758</b>	0.710	0.040	0.185
Specificity <sup>7</sup>	0.698	0.713	0.785	0.805	<b>0.982</b>	0.850
<i>n = 400</i>						
TP <sup>1</sup>	<b>3.73</b>	<b>3.73</b>	<b>3.73</b>	<b>3.73</b>	0.18	1.00
FP <sup>2</sup>	1.65	1.61	1.07	1.00	<b>0.17</b>	1.31
TN <sup>3</sup>	4.35	4.39	4.93	5.00	<b>5.83</b>	4.69
FN <sup>4</sup>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	3.82	3.00
Sensitivity <sup>6</sup>	<b>0.933</b>	<b>0.933</b>	<b>0.933</b>	<b>0.933</b>	0.045	0.250
Specificity <sup>7</sup>	0.725	0.732	0.822	0.833	<b>0.972</b>	0.782

<sup>1</sup>TP: Average number of true positive edges across 100 simulations; <sup>2</sup>FP: Average number of false positive edges across 100 simulations; <sup>3</sup> TN: Average number of true negative edges across 100 simulations; <sup>4</sup> FN: Average number of false negative edges across 100 simulations; <sup>5</sup>: Average false discovery rate across 100 simulations; <sup>6</sup>: Average sensitivity across 100 simulations; <sup>7</sup>: Average specificity across 100 simulations.

goal is to enhance estimation of the grey matter structural co-variance network of the rate of change in cortical thickness using information from white matter connectivity in HD patients.

We analyzed data collected from an international longitudinal natural history study of HD, TRACK-ON ([Klöppel et al. \(2015\)](#)). The cohort in the analyses includes 87 premanifest HD (preHD) patients who carried mutant gene but not yet diagnosed, and was followed up at 3 time points (year 2012, 2013, and 2014). The grey matter cortical thickness regions of interest (ROIs) were obtained from sMRI data and generated by segmenting a T1-weighted image by Freesurfer ([Desikan et al. \(2006\)](#)). The white matter weighted streamline counts between ROIs for each individual resulting from probabilistic tractography were available from DTI

TABLE 4

*Selection Performance of Simulations for  $p = 10$  in Settings 3-4. The best performance for each row is bolded.*

	Proposed Method				EBIC glasso		FGL	
	$\alpha \leq 0.3$		$\alpha = 0.4$		$\alpha \leq 0.3$	$\alpha = 0.4$	$\alpha \leq 0.3$	$\alpha = 0.4$
	Direct	Approx.	Direct	Approx.				
<b>Setting 3: Covariate-dependent, Weak Signal</b>								
<i>n = 200</i>								
TP <sup>1</sup>	<b>1.72</b>	1.65	1.64	1.57	0.04	0.03	0.07	0.07
FP <sup>2</sup>	4.13	3.61	2.93	2.55	<b>0.12</b>	<b>0.12</b>	0.43	0.43
TN <sup>3</sup>	32.87	33.39	34.36	34.74	36.88	<b>37.17</b>	36.57	36.86
FN <sup>4</sup>	6.28	6.35	<b>6.07</b>	6.14	7.96	7.67	7.93	7.64
Sensitivity <sup>6</sup>	<b>0.215</b>	0.206	0.211	0.203	0.005	0.005	0.009	0.009
Specificity <sup>7</sup>	0.888	0.902	0.921	0.932	<b>0.997</b>	<b>0.997</b>	0.988	0.988
<i>n = 400</i>								
TP <sup>1</sup>	<b>2.84</b>	2.35	2.76	2.28	0.05	0.05	0.27	0.26
FP <sup>2</sup>	3.23	2.72	2.12	1.84	<b>0.23</b>	<b>0.23</b>	1.19	1.20
TN <sup>3</sup>	33.77	34.28	35.10	33.38	36.77	<b>36.99</b>	35.81	36.02
FN <sup>4</sup>	5.16	5.65	<b>5.02</b>	5.50	7.95	7.73	7.73	7.52
Sensitivity <sup>6</sup>	0.355	0.294	<b>0.356</b>	0.294	0.006	0.006	0.034	0.034
Specificity <sup>7</sup>	0.913	0.926	0.943	0.951	<b>0.994</b>	<b>0.994</b>	0.968	0.968
<b>Setting 4: Covariate-dependent, Strong Signal</b>								
<i>n = 200</i>								
TP <sup>1</sup>	<b>7.02</b>	5.92	6.76	5.72	0.27	0.22	0.55	0.54
FP <sup>2</sup>	8.36	5.91	6.47	4.67	<b>0.33</b>	0.38	0.79	0.80
TN <sup>3</sup>	28.64	31.09	30.84	32.64	36.67	<b>36.93</b>	36.21	36.51
FN <sup>4</sup>	0.98	2.08	<b>0.93</b>	1.97	7.73	7.47	7.45	7.15
Sensitivity <sup>6</sup>	<b>0.878</b>	0.740	<b>0.878</b>	0.746	0.034	0.030	0.069	0.070
Specificity <sup>7</sup>	0.774	0.840	0.827	0.875	<b>0.991</b>	0.990	0.979	0.979
<i>n = 400</i>								
TP <sup>1</sup>	<b>7.91</b>	7.69	7.58	7.36	0.44	0.42	0.87	0.82
FP <sup>2</sup>	5.91	5.31	4.28	3.86	<b>0.52</b>	0.54	1.33	1.38
TN <sup>3</sup>	31.09	31.69	33.05	33.46	36.48	<b>36.79</b>	35.67	35.95
FN <sup>4</sup>	<b>0.09</b>	0.31	<b>0.09</b>	0.31	7.56	7.25	7.13	6.85
Sensitivity <sup>6</sup>	<b>0.989</b>	0.961	0.988	0.960	0.055	0.055	0.109	0.108
Specificity <sup>7</sup>	0.840	0.856	0.886	0.897	<b>0.986</b>	<b>0.986</b>	0.964	0.963

<sup>1</sup>TP: Average number of true positive edges across 100 simulations; <sup>2</sup>FP: Average number of false positive edges across 100 simulations; <sup>3</sup> TN: Average number of true negative edges across 100 simulations; <sup>4</sup> FN: Average number of false negative edges across 100 simulations; <sup>5</sup>: Average false discovery rate across 100 simulations; <sup>6</sup>: Average sensitivity across 100 simulations; <sup>7</sup>: Average specificity across 100 simulations.

data. Details of the study design and MRI data acquisition can be found in [McColgan et al. \(2017a\)](#).

A previous white matter connectivity study ([McColgan et al. \(2017a\)](#)) using TRACK-ON has demonstrated that preHD subjects have significantly more vulnerable connections compared with controls in the left and right posterior motor-occipital parietal modules. To construct the white matter network, a linear mixed-effects model to exploit the longitudinal DTI measurements was used to compute the rate of change in connections between ROIs and their  $p$ -values, after adjusting for the baseline connection, CAG, age, gender, baseline total motor score (TMS), baseline symbol digit modalities test (SDMT) score which reflects cognitive ability, and baseline total functional capacity (TFC) which measures progressive

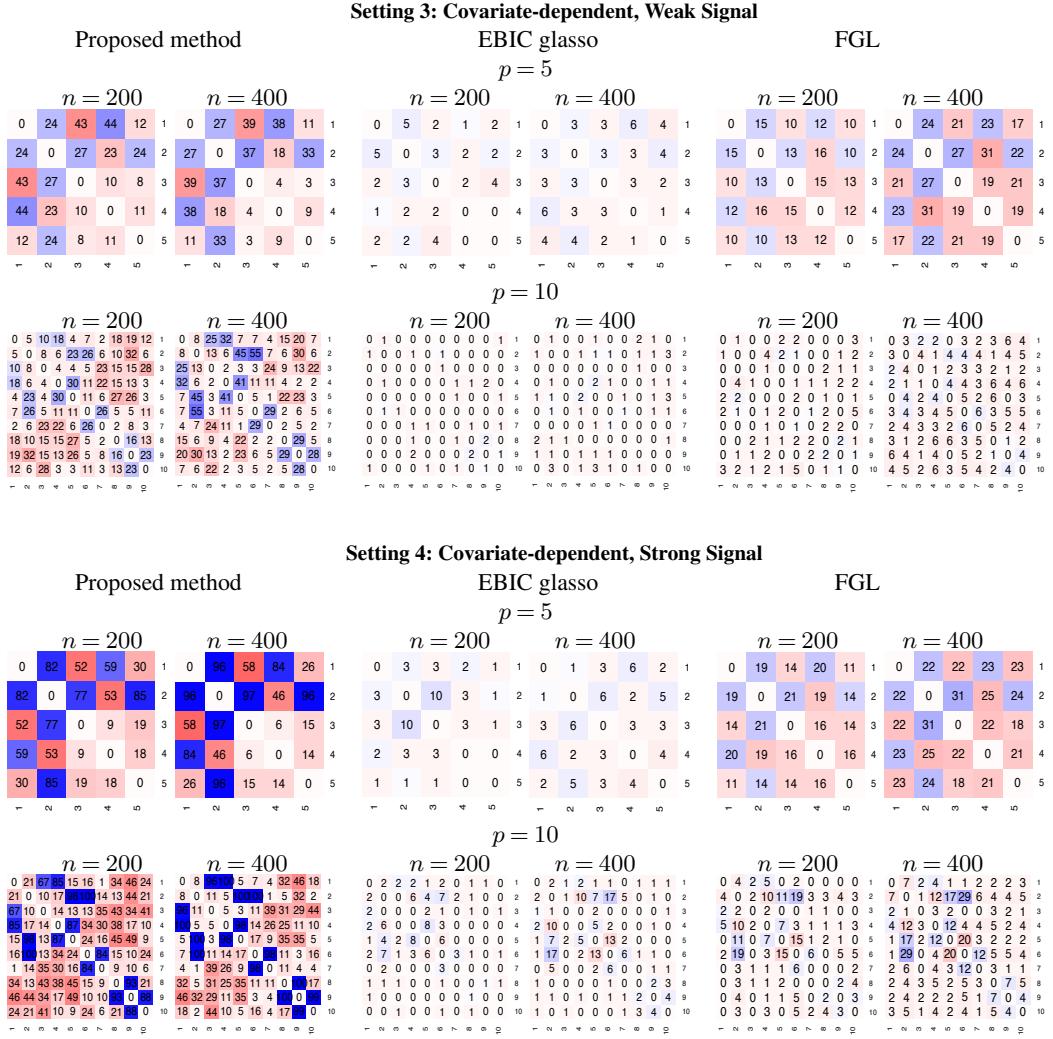


Fig 6: The frequency of edges selected in estimated target networks in Settings 3-4. Each cell represents the number of times an edge between two nodes was selected across 100 simulations. Blue cell: true positive edges; Red cell: false positive edges. Darker color means higher frequency of being selected. Our method outperformed EBIC glasso and FGL in terms of high frequency in selecting true edges while low frequency in null-edges.

functional decline. In the white matter connection network, the nodes were the ROIs which had at least one connection with the false discovery rate (FDR) correction  $q < 0.1$  and within the two modules. If the connection strength between nodes  $j$  and  $k$  for subject  $i$  was also non-zero at the baseline visit, the edge between nodes  $j$  and  $k$  was defined as present in subject  $i$ 's white matter connection network. For each individual, the white matter connection network ( $\mathcal{G}_1$ ) consist of 10 nodes (ROIs) and at most 8 edges (Supplement D of the Supplementary Material (Xie, Zeng and Wang (2020)), Table S1). The anatomical distances between ROIs were obtained from a template and were the euclidean distances. The distances ranged from 14 to 88 and we scaled the distance by dividing by 150. As a preliminary analysis to obtain good initial values, we fitted a logistic regression model using product of degrees and distance as covariates on white matter network and obtained the estimated parameters  $\hat{\gamma} = 0.44$  and

TABLE 5  
*Average Mean Squared Error of  $\beta_{jk}$  Estimation across 100 Simulations.*

	Weak Signal		Strong Signal	
	Direct	Approx.	Direct	Approx.
Settings 1-2: External Modality Alone				
$p = 5, n = 200$	0.0169	0.0173	0.0124	0.0131
$p = 5, n = 400$	0.0080	0.0083	0.0035	0.0037
$p = 10, n = 200$	0.0086	0.0090	0.0025	0.0024
$p = 10, n = 400$	0.0031	0.0033	0.0014	0.0016
Settings 3-4: Covariate-dependent				
$p = 5, n = 200$	0.0636	0.0616	0.0905	0.0898
$p = 5, n = 400$	0.0363	0.0362	0.0487	0.0478
$p = 10, n = 200$	0.0449	0.0413	0.0725	0.0543
$p = 10, n = 400$	0.0276	0.0258	0.0306	0.0253

TABLE 6  
*Average Running Time for One Simulation Using Direct Calculation and Approximation.*

	Direct		Approx.	
	weak signal	strong signal	weak signal	strong signal
$p = 5, n = 200$	2.838 secs	2.730 secs	0.923 secs	1.202 secs
$p = 5, n = 400$	5.517 secs	5.035 secs	1.546 secs	1.618 secs
$p = 10, n = 200$	4.767 mins	5.269 mins	2.079 mins	2.907 mins
$p = 10, n = 400$	7.516 mins	7.798 mins	1.683 mins	2.608 mins

$\hat{\eta} = 0.31$ . Thus, we set the searching path for  $\gamma$  to be  $\{0.2, 0.3, 0.4, 0.5, 0.6\}$  and  $\eta$  to be  $\{0.15, 0.20, 0.25, 0.30, 0.35\}$ .

The same 10 ROIs were used in the cortical grey matter structural co-variance network ( $\mathcal{G}_*$ ). The nodes were the rates of change in cortical thickness ROIs estimated by a linear mixed-effects model adjusting for the baseline cortical thickness. CAP score (age  $\times$  (CAG-35.5); [Zhang et al. \(2011\)](#)), baseline TMS and baseline SDMT were covariates used in our method to estimate the network.

We set the length of the path of hard-thresholding values in our method or tuning parameters in EBIC glasso to be 50. For FGL, we stratified subjects into four groups based on CAP score and searched tuning parameter for lasso penalty from 0 to 0.2 and that for subgroup similarity penalty from 0 to 1 for FGL as suggested in [Danaher, Wang and Witten \(2014\)](#). We set proportion value  $\alpha = 0$  and hyperparameter  $\delta = 0.5$  for all methods.

Our method identified 22 connections, while EBIC glasso identified 24 connections and FGL identified a much denser network with 41 connections (Supplement D of the Supplementary Material ([Xie, Zeng and Wang \(2020\)](#))), Table S1). Eight connections were identified by all three methods. Five connections identified by the proposed method in the grey matter network were in common with the white matter connection network, whereas only three connections identified by EBIC glasso in the grey matter network overlapped with the white matter connection network. Our results are more consistent with previous literature on the characteristics of overlapped connections between structural co-variance network and white matter connectivity ([Alexander-Bloch, Giedd and Bullmore \(2013\)](#)).

Because the estimated connections from our method are subgroup- and subject-specific, our estimated connections can be treated as additional variables for predicting clinical outcomes and patient stratification. In contrast, connections estimated from EBIC glasso resulted in the same connection strengths for the whole population, so cannot be used as patient-specific covariates.

**4.1. Clinical Utility of the Connections.** To assess the clinical utility of the connections identified by integrative network learning, we evaluated their associations with TMS at the last visit, which is a measure of motor symptoms, the hallmark of HD. We compared the R-squares of the linear regression in a standard covariate model (with covariates CAP, baseline SDMT, baseline TMS), non-connection imaging model (with covariates and 10 regional cortical thickness ROIs), and the connection imaging model (with covariates, 10 regional cortical thickness, and identified connections by the proposed method). The estimates  $\hat{\beta}_{jk}^T \mathbf{X}_i / \hat{\sigma}_j^2 * \hat{E}(B_{ijk} | M_{ij}, D_i)$  was treated as the connection strength between ROIs  $j$  and  $k$ . The R-squared value of the standard covariate model was 0.376, increased to 0.494 of the non-connection imaging model, and was 0.631 of the connection imaging model. Thus, an additional 13.7% of variance was explained by the identified connections. When using connections estimated from FGL, the R-squared value of the connection imaging model was 0.519, which was worse than our method. Note that the network was constructed without using the motor symptoms. These results suggest that the identified connections from our method are highly predictive of the future motor symptom outcomes after including standard measures and regional cortical thickness.

The connections identified by the proposed method are comparable to those reported in previous literature. The connection between precuneus and isthmus cingulate regions, identified in both the structural co-variation network and white matter connectivity, was also shown to have greater functional connectivity in preHD subjects compared to control in a recent fMRI study in TRACK-ON ([McColgan et al. \(2017b\)](#)). The thickness of occipital regions, which are important visual processing regions, including the cuneus and lateral occipital, has been found to impact cognition ([Rosas et al. \(2008\)](#); [Johnson et al. \(2015\)](#)). The precuneus region responds to a range of cognitive processes and the paracentral region correlates with cognitive performance ([Rosas et al. \(2005, 2008\)](#)). These regions with similar functions might co-vary as well.

**4.2. Clustering Analysis Based on Connectivity Measures.** We show another utility of the identified connections in stratifying patients into groups with distinct clinical prognosis. Figure 7 shows that the connections have greater between-subject discriminative power and tend to cluster patients into homogeneous groups, whereas no clear cluster is seen in the heatmap of the regional nodes. Connections estimated from our method show stronger discriminative power than FGL. Furthermore, we conducted two  $k$ -means clustering analyses based on either the connectivity identified by our method or connectivity identified by FGL or the ROIs and compared the results. We chose the number of clusters to be 4 since the elbow points for both connectivity and regional nodes are at cluster=4 (Supplement D of the Supplementary Material ([Xie, Zeng and Wang \(2020\)](#)), Figure S1). ANOVA  $F$ -test was conducted to test the differences between the TMS at the last visit and SDMT at the last visit of the identified four subgroups based on connectivity and regional nodes. There is a significant group difference for connectivity-based clusters from our method (Figure 8, Top panel), while no significant difference for connectivity-based clusters from FGL or regional nodes based clusters (Figure 8, Middle and Bottom panels). The connectivities from our method can better stratify subjects into meaningful clinical groups with different motor scores at follow-up than regional nodes.

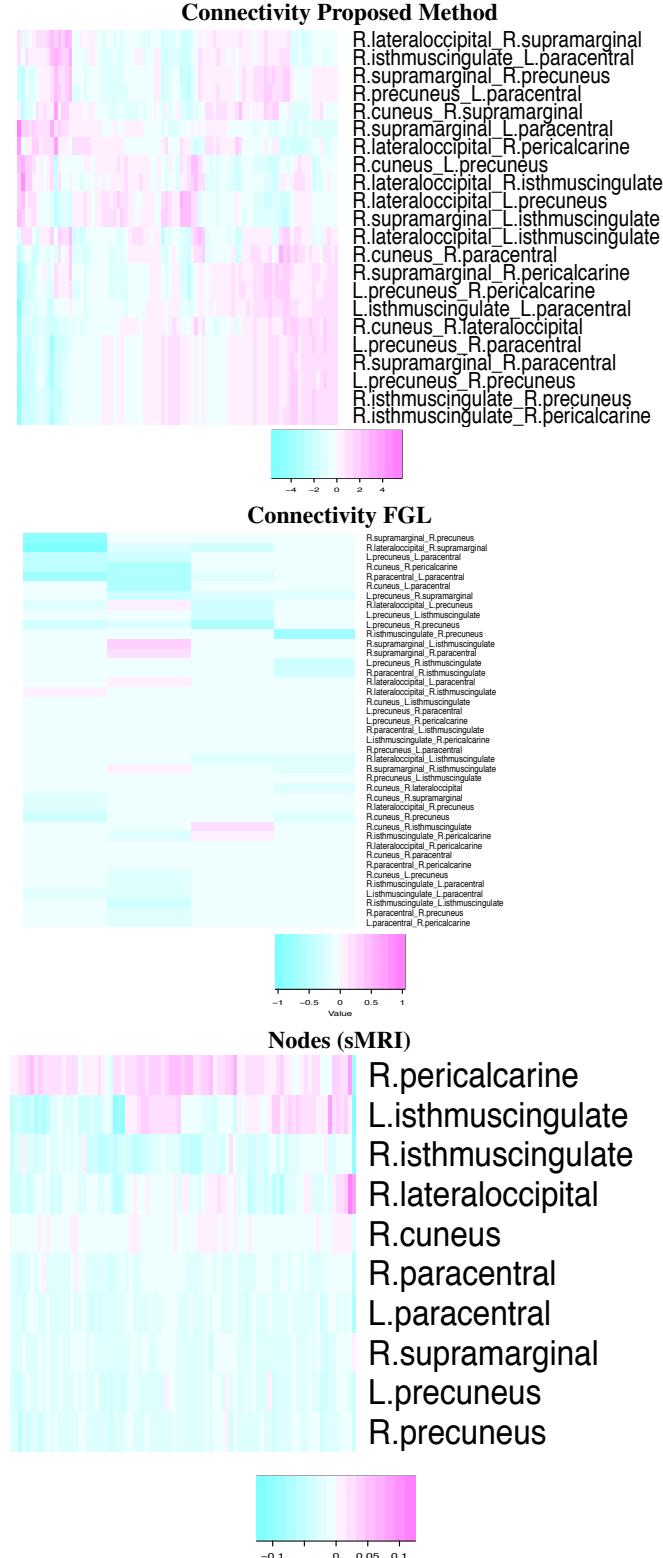


Fig 7: Heatmaps of connectivity and nodes. Each column represents one subject and each row represents one connection or node. Connections tend to cluster patients into groups, whereas no clear cluster is seen for regional nodes. Connectivity estimated from our method show greater discriminative power than connectivity estimated from FGL.

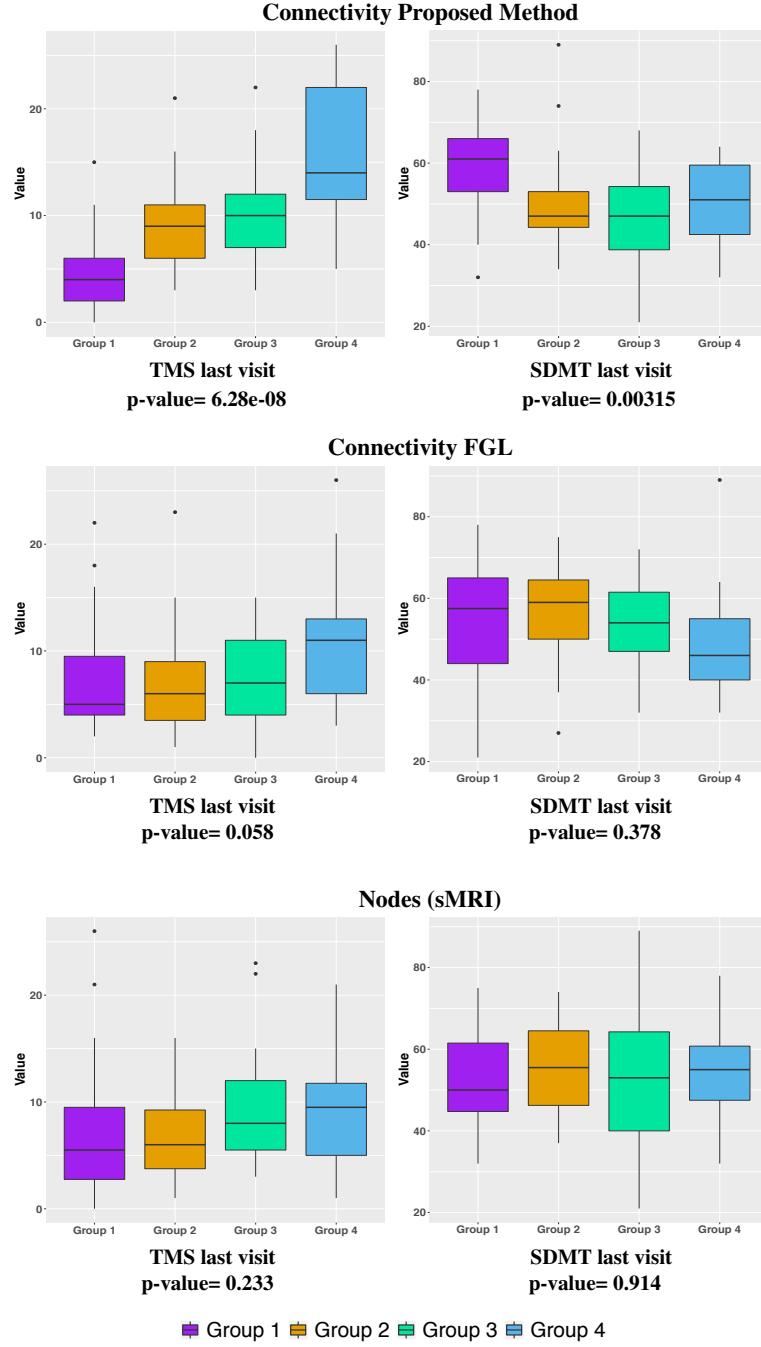


Fig 8: Clustering analysis based on connectivity and nodes. Connectivity-based clusters from proposed method show significant group difference in follow-up clinical outcomes, while no significant difference is observed for connectivity-based clusters from FGL or regional nodes-based clusters.

**5. Discussion.** In this work, we propose an integrative network learning method under a pseudo-likelihood graphical model to improve the estimation of a target covariate-dependent network. Our method exploits a shared latent network between multiple modalities of biomarker measurements. The external network is treated as prior of the shared network

and a network growth model that captures the small-worldness and fat-tailed degree distribution properties in human brain networks has been used to generate the shared network. The shared network then constraints the estimation of the target network.

There are several advantages of our modeling framework in (1) and (2). First, since model (2) has biological underpinning, it improves biological relevance for the estimation of target network  $\mathcal{G}_*$ . The estimated connections in  $\mathcal{G}_*$  will reflect the anatomical distances and common neighbors in  $\mathcal{G}_1$ . Second, by inferring  $B_{ijk}$  as unknown latent connection statuses from observed data  $M_i$ ,  $s_{ijk}$ , and  $d_{ijk}$ , we borrow information from networks measured by two neuroimaging modalities and the efficiency is improved. Third, regular analysis of covariate-dependent structural co-variance networks is conducted at the subgroup level (depending on the resolution of covariates). In our approach, since white matter connectivities are measured for each individual, incorporating its information into the estimation of  $\theta_{ijk}$  of structural covariance network improves its resolution to the subject-level. In addition, our method creates new useful network measures (connectivity) to improve the prediction of clinical outcomes and better stratify subjects into homogeneous groups. Thus, the connectivity measures might be considered to be included as new biomarker features to recruit subjects for future clinical trials.

Several extensions can be considered. We can consider generalizing to use the power prior ([Ibrahim et al. \(2000\)](#); [Spiegelhalter, Abrams and Myles \(2004\)](#)) in the pseudo likelihood  $L_j$ , i.e.,  $f(\{B_{ijk}, k \neq j\} | D_i)$  is raised to some power, where the power is tuned to determine how much information can be borrowed from the external modality network. Since our model is a node-wise model that estimates the edges separately for each node, we can extend current models to high-dimensional case when a large number of nodes exists by adding a sparse penalty on  $\beta_{jk}$  in the M-step. However, it will be computationally expensive in calculating the posterior expectations in the E-step in a high-dimensional EM optimization ([Wang et al. \(2015\)](#)). The current approximation approach may not be directly applicable to large networks and other faster computational techniques such as variational Bayes might be considered to further improve computation.

R code implementing this method is available in the Supplementary Material ([Xie, Zeng and Wang \(2020\)](#)) and online at <http://github.com/shanghongxie/INL>.

**Acknowledgements.** This research is supported by U.S. NIH grants NS073671, GM124104, and MH117458.

## SUPPLEMENTARY MATERIAL

**Supplement A.** This section describes the approaches to reduce computational burden in computing the posterior expectations in E-step.

0.

**Supplement B.** This section contains the details of simulation settings.

0.

**Supplement C.** In this section, we describe the EBIC criteria used for fused graphical lasso.

0.

**Supplement D.** This section contains Table S1 and Figure S1.

0.

**Code.** R code implementing this method.

0.

## REFERENCES

- ALEXANDER-BLOCH, A., GIEDD, J. N. and BULLMORE, E. (2013). Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience* **14** 322–336.
- ALEXANDER-BLOCH, A., RAZNAHAN, A., BULLMORE, E. and GIEDD, J. (2013). The convergence of maturational change and structural covariance in human cortical networks. *Journal of Neuroscience* **33** 2889–2899.
- BETZEL, R. F., AVENA-KOENIGSBERGER, A., GOÑI, J., HE, Y., DE REUS, M. A., GRIFFA, A., VÉRTES, P. E., MIŠIC, B., THIRAN, J.-P., HAGMANN, P. et al. (2016). Generative models of the human connectome. *Neuroimage* **124** 1054–1064.
- BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* **10** 186.
- CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759–771.
- COPPEN, E. M., VAN DER GROND, J., HAFKEMEIJER, A., ROMBOUTS, S. A. and ROOS, R. A. (2016). Early grey matter changes in structural covariance networks in Huntington's disease. *NeuroImage: Clinical* **12** 806–814.
- DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 373–397.
- DESIKAN, R. S., SÉGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P., HYMAN, B. T. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31** 968–980.
- EPSKAMP, S. and FRIED, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*.
- FORNITO, A., ZALESKY, A. and BULLMORE, E. (2016). *Fundamentals of brain network analysis*. Academic Press.
- FOYgel, R. and DRTON, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in Neural Information Processing Systems* 604–612.
- FOYgel, R. and DRTON, M. (2011). Bayesian model choice and information criteria in sparse generalized linear models. *arXiv preprint arXiv:1112.5635*.
- FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- GONG, G., HE, Y., CHEN, Z. J. and EVANS, A. C. (2012). Convergence and divergence of thickness correlations with diffusion connections across the human cerebral cortex. *Neuroimage* **59** 1239–1248.
- HAO, B., SUN, W. W., LIU, Y. and CHENG, G. (2018). Simultaneous Clustering and Estimation of Heterogeneous Graphical Models. *Journal of Machine Learning Research* **18** 1–58.
- HE, Y., CHEN, Z. J. and EVANS, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cerebral Cortex* **17** 2407–2419.
- HE, Y., CHEN, Z. and EVANS, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *Journal of Neuroscience* **28** 4756–4766.
- HINNE, M., AMBROGIONI, L., JANSEN, R. J., HESKES, T. and VAN GERVEN, M. A. (2014). Structurally-informed Bayesian functional connectivity analysis. *Neuroimage* **86** 294–305.
- IBRAHIM, J. G., CHEN, M.-H. et al. (2000). Power prior distributions for regression models. *Statistical Science* **15** 46–60.
- JOHNSON, E. B., REES, E. M., LABUSCHAGNE, I., DURR, A., LEAVITT, B. R., ROOS, R. A., REILMANN, R., JOHNSON, H., HOBBS, N. Z., LANGBEHN, D. R. et al. (2015). The impact of occipital lobe cortical thickness on cognitive task performance: an investigation in Huntington's Disease. *Neuropsychologia* **79** 138–146.
- KLÖPPEL, S., GREGORY, S., SCHELLER, E., MINKOVA, L., RAZI, A., DURR, A., ROOS, R. A., LEAVITT, B. R., PAPOUTSI, M., LANDWEHRMEYER, G. B. et al. (2015). Compensation in preclinical Huntington's disease: evidence from the track-on HD study. *EBioMedicine* **2** 1420–1429.
- LERCH, J. P., WORSLEY, K., SHAW, W. P., GREENSTEIN, D. K., LENROOT, R. K., GIEDD, J. and EVANS, A. C. (2006). Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI. *Neuroimage* **31** 993–1003.
- MCCOLGAN, P., SEUNARINE, K. K., GREGORY, S., RAZI, A., PAPOUTSI, M., LONG, J. D., MILLS, J. A., JOHNSON, E., DURR, A., ROOS, R. A. et al. (2017a). Topological length of white matter connections predicts their rate of atrophy in premanifest Huntington's disease. *JCI Insight* **2**.
- MCCOLGAN, P., RAZI, A., GREGORY, S., SEUNARINE, K. K., DURR, A., AC ROOS, R., LEAVITT, B. R., SCAHILL, R. I., CLARK, C. A., LANGBEHN, D. R. et al. (2017b). Structural and functional brain network

- correlates of depressive symptoms in premanifest Huntington's disease. *Human Brain Mapping* **38** 2819–2829.
- NG, B., VAROQUAUX, G., POLINE, J.-B. and THIRION, B. (2012). A novel sparse graphical approach for multimodal brain connectivity inference. *International Conference on Medical Image Computing and Computer-Assisted Intervention* 707–714.
- PAULSEN, J. S., HAYDEN, M., STOUT, J. C., LANGBEHN, D. R., AYLWARD, E., ROSS, C. A., GUTTMAN, M., NANCE, M., KIEBURTZ, K., OAKES, D. et al. (2006). Preparing for preventive clinical trials: the PREDICT-HD study. *Archives of Neurology* **63** 883–890.
- PAULSEN, J., LANGBEHN, D., STOUT, J., AYLWARD, E., ROSS, C., NANCE, M., GUTTMAN, M., JOHNSON, S., MACDONALD, M., BEGLINGER, L. et al. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery & Psychiatry* **79** 874–880.
- PINEDA-PARDO, J. A., BRUÑA, R., WOOLRICH, M., MARCOS, A., NOBRE, A. C., MAESTÚ, F. and VIDAU-RRE, D. (2014). Guiding functional connectivity estimation by structural connectivity in MEG: an application to discrimination of conditions of mild cognitive impairment. *Neuroimage* **101** 765–777.
- RIZVI, B., NARKHEDE, A., LAST, B. S., BUDGE, M., TOSTO, G., MANLY, J. J., SCHUPF, N., MAYEUX, R. and BRICKMAN, A. M. (2018). The effect of white matter hyperintensities on cognition is mediated by cortical atrophy. *Neurobiology of Aging* **64** 25–32.
- ROSAS, H., HEVELONE, N., ZALETA, A., GREVE, D., SALAT, D. and FISCHL, B. (2005). Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology* **65** 745–747.
- ROSAS, H. D., SALAT, D. H., LEE, S. Y., ZALETA, A. K., PAPPU, V., FISCHL, B., GREVE, D., HEVELONE, N. and HERSCH, S. M. (2008). Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain* **131** 1057–1068.
- SCHMITT, J., LENROOT, R., WALLACE, G., ORDAZ, S., TAYLOR, K., KABANI, N., GREENSTEIN, D., LERCH, J., KENDLER, K., NEALE, M. et al. (2008). Identification of genetically mediated cortical networks: a multivariate study of pediatric twins and siblings. *Cerebral Cortex* **18** 1737–1747.
- SEELEY, W. W., CRAWFORD, R. K., ZHOU, J., MILLER, B. L. and GREICIUS, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62** 42–52.
- SPIEGELHALTER, D. J., ABRAMS, K. R. and MYLES, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation* **13**. John Wiley & Sons.
- STÄDLER, N., DONDELINGER, F., HILL, S. M., AKBANI, R., LU, Y., MILLS, G. B. and MUKHERJEE, S. (2017). Molecular heterogeneity at the network level: high-dimensional testing, clustering and a TCGA case study. *Bioinformatics* **33** 2890–2896.
- TABRIZI, S. J., LANGBEHN, D. R., LEAVITT, B. R., ROOS, R. A., DURR, A., CRAUFURD, D., KENNARD, C., HICKS, S. L., FOX, N. C., SCAHILL, R. I. et al. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *The Lancet Neurology* **8** 791–801.
- VÉRTES, P. E., ALEXANDER-BLOCH, A. F., GOOTAY, N., GIEDD, J. N., RAPOPORT, J. L. and BULLMORE, E. T. (2012). Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences* **109** 5868–5873.
- WANG, Z., GU, Q., NING, Y. and LIU, H. (2015). High dimensional EM algorithm: Statistical optimization and asymptotic normality. In *Advances in Neural Information Processing Systems* 2521–2529.
- XIE, S., ZENG, D. and WANG, Y. (2020). Supplement to “Integrative network learning for multi-modality biomarker data.”
- ZHANG, Y., LONG, J. D., MILLS, J. A., WARNER, J. H., LU, W. and PAULSEN, J. S. (2011). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **156** 751–763.