SUPPLEMENTARY MATERIALS FOR *"INTEGRATIVE NETWORK LEARNING FOR MULTI-MODALITY BIOMARKER DATA"*

BY SHANGHONG XIE[1,*], DONGLIN ZENG[2] AND YUANJIA WANG[1,†]

[1] *Department of Biostatistics, Mailman School of Public Health, Columbia University,* [*]*sx2168@cumc.columbia.edu;* [†]*yw2016@cumc.columbia.edu*

[2] *Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill,* *dzeng@email.unc.edu*

Supplement A: Reducing Computational Burden. The computation of the posterior expectations of connection indicators $E(B_{ijk}|M_{ij}, D_i)$ and $E(B_{ijk}B_{ijk'}|M_{ij}, D_i)$ increases exponentially with the network size p. There are 2^{p-2} combinations of summations involved in $P(B_{ijk}, M_{ij}|D_i)$ for each subject i and an edge between nodes j and k among $p*(p-1)$ edges in total.

A.1: Pruning Based on Posterior Expectation. Motivated by automatic relevance determination (Wipf and Nagarajan (2008)), we treat the estimated posterior expectation $E(B_{ijk}|M_{ij}, D_i)$ from the previous iteration in EM algorithm as a tuning parameter to further prune b_{ijk} and reduce the combinations of summations in computation. In each EM iteration, we set $b_{ijk} = 1$ if $E(B_{ijk}|M_{ij}, D_i)$ estimated from the previous iteration is greater than an upper threshold, and set $b_{ijk} = 0$ if the estimated $E(B_{ijk}|M_{ij}, D_i)$ is smaller than a lower threshold. The combinations are then restricted only to those b_{ijk}'s with $E(B_{ijk}|M_{ij}, D_i)$ within the interval defined by two thresholds.

A.2: Approximation of the Posterior Expectation. The pruning approach discussed above (Section A.1) still calculates the summations directly. We propose an approximation to bypass direct sum to speed up the computation. Note that the summation involves the form of $\exp(-u^2)$. Our goal is to compute the function $\exp(-u^2)$ with a fast, non-iterative, non Monte-Carlo algorithm. To this end, we approximate $\exp(-u^2)$ by using the mixture of exponential distributions $\exp(-\lambda u)$ uniformly in a range [-A, A]. For u_i in [-A, A], let $y_i = \exp(-u_i^2)$. Let $\breve{\alpha}_i = (\exp(-\lambda_1 u_i), \cdots, \exp(\lambda_K u_i))^T$ be a vector of K basis functions. We use least squares approximation $y_i = \breve{\alpha}_i \boldsymbol{\zeta} + e_i$ to compute basis coefficients $\boldsymbol{\zeta}$. We can then approximate any value of $\exp(-u_0^2)$ by $\breve{\alpha}_{u_0}\widehat{\boldsymbol{\zeta}}$. The mean squared error of such approximation is about 1×10^{-10} when using the mixture exponentials we set in Section 3.1.

Once we have obtained $\widehat{\boldsymbol{\zeta}}$, we can use it to calculate an approximation of

$$T(c_1, ..., c_p, a_1, ..., a_p) \equiv \sum_{(y_1,...,y_p)\in\{0,1\}^p} \exp\left\{-(\sum_{j=1}^p c_j y_j)^2 + \sum_{j=1}^p a_j y_j\right\}$$

$$\approx \sum_{(y_1,...,y_p)\in\{0,1\}^p} \left[\sum_{k=1}^m p_k \exp\left\{-\lambda_k(\sum_{j=1}^p c_j y_j)\right\}\right] \exp\left\{\sum_{j=1}^p a_j y_j\right\}$$

which is equal to

$$\sum_{k=1}^m p_k \prod_{j=1}^p (1 + e^{a_j - \lambda_k c_j}).$$

Therefore we can approximate the posterior expectations by

$$
\begin{aligned}
&P(B_{ijk}=b, M_{ij}|D_i)\\
&=\left(\prod_{k\neq j}(1-p_{ijk})\right)\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{1}{2\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik})^2\right)\exp\left(b\log(\frac{p_{ijk}}{1-p_{ijk}})\right)\times\\
&\sum_{b_{ij1},\ldots,b_{ijp}}\exp\left(-\frac{1}{2\sigma_j^2}(\sum_{l\neq j,k}B_{ijl}\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il})^2\right)\\
&\times\exp\left(\sum_{l\neq j,k}B_{ijl}[\frac{1}{\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik})\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il}+\log(\frac{p_{ijl}}{1-p_{ijl}})]\right)\\
&=\left(\prod_{k\neq j}(1-p_{ijk})\right)\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{1}{2\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik})^2\right)\exp\left(b\log(\frac{p_{ijk}}{1-p_{ijk}})\right)\\
&\times T(c_l^{(1)},a_l^{(1)},b;l\neq j,k),
\end{aligned}
$$

where $c_l^{(1)}=\frac{1}{\sqrt{2}\sigma_j}\beta_{jl}^T\boldsymbol{X}_iM_{il}$ and $a_l^{(1)}=\frac{1}{\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik})\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il}+\log(\frac{p_{ijl}}{1-p_{ijl}})$, $l\neq j,k$, and

$$
\begin{aligned}
&P(B_{ijk}=b, B_{ijk'}=b', M_{ij}|D_i)\\
&=\left(\prod_{k\neq j}(1-p_{ijk})\right)\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{1}{2\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik}-b'\boldsymbol{\beta}_{jk'}^T\boldsymbol{X}_i)M_{ik'}^2\right)\\
&\times\exp\left(b\log(\frac{p_{ijk}}{1-p_{ijk}})+b'\log(\frac{p_{ijk'}}{1-p_{ijk'}})\right)\\
&\sum_{b_{ij1},\ldots,b_{ijp}}\exp\left(-\frac{1}{2\sigma_j^2}(\sum_{l\neq j,k,k'}B_{ijl}\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il})^2\right)\times\\
&\exp\left(\sum_{l\neq j,k,k'}B_{ijl}[\frac{1}{\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik}-b'\boldsymbol{\beta}_{jk'}^T\boldsymbol{X}_iM_{ik'})\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il}+\log(\frac{p_{ijl}}{1-p_{ijl}})]\right)\\
&=\left(\prod_{k\neq j}(1-p_{ijk})\right)\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{1}{2\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik}-b'\boldsymbol{\beta}_{jk'}^T\boldsymbol{X}_iM_{ik'})^2\right)\times\\
&\exp\left(b\log(\frac{p_{ijk}}{1-p_{ijk}})+b'\log(\frac{p_{ijk'}}{1-p_{ijk'}})\right)T(c_l^{(2)},a_l^{(2)},b,b';l\neq j,k,k'),
\end{aligned}
$$

where $c_l^{(2)}=\frac{1}{\sqrt{2}\sigma_j}\beta_{jl}^T\boldsymbol{X}_iM_{il}$ and $a_l^{(2)}=\frac{1}{\sigma_j^2}(M_{ij}-b\boldsymbol{\beta}_{jk}^T\boldsymbol{X}_iM_{ik}-b'\boldsymbol{\beta}_{jk'}^T\boldsymbol{X}_iM_{ik'})$ $\times\,\boldsymbol{\beta}_{jl}^T\boldsymbol{X}_iM_{il}+\log(\frac{p_{ijl}}{1-p_{ijl}})$, $l\neq j,k,k'$.

This approximation does not require the computation of an exponential number of terms in the summation. Instead, compute the posterior expectation using a basis expansion and least squares regression to obtain basis coefficients. The computational burden reduces from exponential in the graph size to polynomial.

Supplement B: Details in Simulation Settings. Setting 1 (Varied by external modality network alone, Weak signal):

- When $p=5$, $\beta_{jk}=1*\sigma_j^2$ for $(j,k)\in\{(1,2),(2,1),(1,4),(4,1)\}$, $\beta_{jk}=-1*\sigma_j^2$ for $(j,k)\in\{(2,3),(3,2),(2,5),(5,2)\}$, and $\sigma^2_{2*j-1}=0.2$, $\sigma^2_{2*j}=0.3$ for $j=1,\dots,p$;
- When $p=10$, $\beta_{jk}=1*\sigma_j^2$ for $(j,k)\in\{(1,3),(3,1),(1,4),(4,1),(4,5),$ $(5,4),(8,9),(9,8),(9,10),(10,9)\}$, $\beta_{jk}=-1*\sigma_j^2$ for $(j,k)\in\{(2,5),$ $(5,2),(2,6),(6,2),(6,7),(7,6)\}$, and $\sigma^2_{2*j-1}=0.3$, $\sigma^2_{2*j}=0.4$ for $j=1,\dots,p$.

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

- When $p=5$, $\beta_{jk}=1.5*\sigma_j^2$ for $(j,k)\in\{(1,2),(2,1),(1,4),(4,1)\}$, $\beta_{jk}=-1.5*\sigma_j^2$ for $(j,k)\in\{(2,3),(3,2),(2,5),(5,2)\}$, and $\sigma^2_{2*j-1}=0.2$, $\sigma^2_{2*j}=0.3$ for $j=1,\dots,p$;
- When $p=10$, $\beta_{jk}=1.5*\sigma_j^2$ for $(j,k)\in\{(1,3),(3,1),(1,4),(4,1),(4,5),$ $(5,4),(8,9),(9,8),(9,10),(10,9)\}$, $\beta_{jk}=-1.5*\sigma_j^2$ for $(j,k)\in\{(2,5),$ $(5,2),(2,6),(6,2),(6,7),(7,6)\}$, and $\sigma^2_{2*j-1}=0.3$, $\sigma^2_{2*j}=0.4$ for $j=1,\dots,p$.

Setting 3 (Covariate-dependent, Weak signal):

- When $p=5$, $\boldsymbol{\beta}_{jk}^T=(0.5,-0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(1,2),(2,1)\}$, $\boldsymbol{\beta}_{jk}^T=(0.5,0.5,-0.5)*\sigma_j^2$ for $(j,k)\in\{(1,4),(4,1)\}$, $\boldsymbol{\beta}_{jk}^T=(0.5,0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(2,3),(3,2)\}$, $\boldsymbol{\beta}_{jk}^T=(-0.5,0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(2,5),(5,2)\}$, and $\sigma^2_{2*j-1}=0.2$, $\sigma^2_{2*j}=0.3$ for $j=1,\dots,p$;
- When $p=10$, $\boldsymbol{\beta}_{jk}^T=(0.5,0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(1,3),(3,1)\}$, $\boldsymbol{\beta}_{jk}^T=(-0.5,0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(1,4),(4,1),(2,6),(6,2)\}$, $\boldsymbol{\beta}_{jk}^T=(-0.5,0.5,-0.5)*\sigma_j^2$ for $(j,k)\in\{(2,5),(5,2),(9,10),(10,9)\}$, $\boldsymbol{\beta}_{jk}^T=(0.5,-0.5,0.5)*\sigma_j^2$ for $(j,k)\in\{(4,5),(5,4)\}$, $\boldsymbol{\beta}_{jk}^T=(0.5,0.5,-0.5)*\sigma_j^2$ for $(j,k)\in\{(6,7),(7,6),(8,9),(9,8)\}$, and $\sigma^2_{2*j-1}=0.3$, $\sigma^2_{2*j}=0.4$ for $j=1,\dots,p$.

Setting 4 (Covariate-dependent, Strong signal):

- When $p=5$, $\boldsymbol{\beta}_{jk}^T=(1,-1,1)*\sigma_j^2$ for $(j,k)\in\{(1,2),(2,1)\}$, $\boldsymbol{\beta}_{jk}^T=(1,1,-1)*\sigma_j^2$ for $(j,k)\in\{(1,4),(4,1)\}$, $\boldsymbol{\beta}_{jk}^T=(1,1,1)*\sigma_j^2$ for $(j,k)\in\{(2,3),(3,2)\}$, $\boldsymbol{\beta}_{jk}^T=(-1,1,1)*\sigma_j^2$ for $(j,k)\in\{(2,5),(5,2)\}$, and $\sigma^2_{2*j-1}=0.2$, $\sigma^2_{2*j}=0.3$ for $j=1,\dots,p$;
- When $p=10$, $\boldsymbol{\beta}_{jk}^T=(1,1,1)*\sigma_j^2$ for $(j,k)\in\{(1,3),(3,1)\}$, $\boldsymbol{\beta}_{jk}^T=(-1,1,1)*\sigma_j^2$ for $(j,k)\in\{(1,4),(4,1),(2,6),(6,2)\}$, $\boldsymbol{\beta}_{jk}^T=(-1,1,-1)*\sigma_j^2$ for $(j,k)\in\{(2,5),(5,2),(9,10),(10,9)\}$, $\boldsymbol{\beta}_{jk}^T=(1,-1,1)*\sigma_j^2$ for $(j,k)\in\{(4,5),(5,4)\}$, $\boldsymbol{\beta}_{jk}^T=(1,1,-1)*\sigma_j^2$ for $(j,k)\in$ $\{(6,7),(7,6),$ $(8,9),(9,8)\}$, and $\sigma^2_{2*j-1}=0.3$, $\sigma^2_{2*j}=0.4$ for $j=1,\dots,p$.

Supplement C: EBIC for Fused Graphical Lasso. We adapted EBIC for EBIC glasso to define EBIC for fused graphical lasso (FGL) as:

$$EBIC=-2\ell+\sum_{k=1}^{K}E_k\log(n_k)+4\delta\log(p)\sum_{k=1}^{K}E_k,$$

where ℓ is the log-likelihood function in FGL (Danaher, Wang and Witten (2014)) , K is the total number of subgroups, E_k is the number of non-zero edges in k-th subgroup network, and δ is the hyperparameter.

Supplement D: Other Details in Real Data Application.

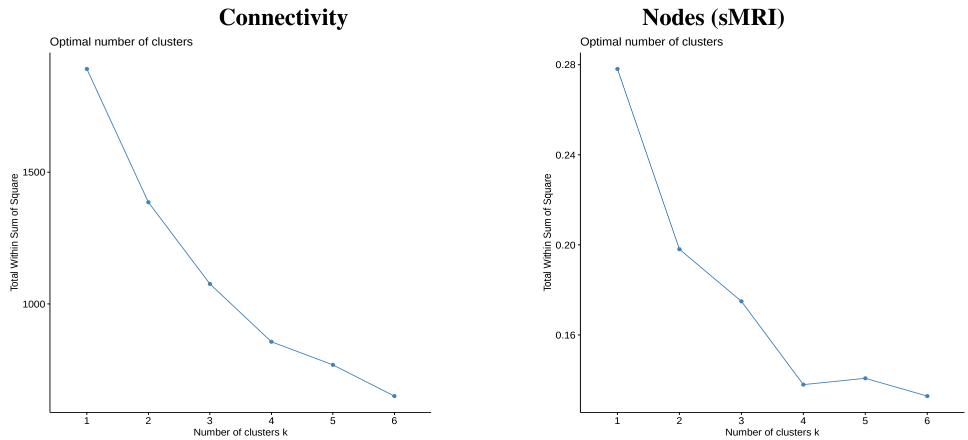

Fig S1: Elbow plots of connectivity and nodes. $K = 4$ is the elbow point in real data application.

REFERENCES

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.

WIPF, D. P. and NAGARAJAN, S. S. (2008). A new view of automatic relevance determination. In *Advances in Neural Information Processing Systems* 1625–1632.

TABLE S1

Identified Cortical Connections in TRACK-ON Study.

Connections	Proposed Method (22 edges)	EBIC glasso (24 edges)	FGL (41 edges)	DTI (8 edges)
R.isthmuscingulate/R.pericalcarine	✓	✓	✓	
L.precuneus/R.paracentral	✓		✓	✓
R.isthmuscingulate/R.precuneus	✓	✓	✓	✓
L.precuneus/R.pericalcarine	✓		✓	✓
R.isthmuscingulate/L.paracentral	✓		✓	✓
L.precuneus/R.precuneus	✓	✓	✓	
L.isthmuscingulate/L.paracentral	✓	✓	✓	
R.cuneus/L.precuneus	✓	✓	✓	
R.precuneus/L.paracentral	✓		✓	
R.lateraloccipital/R.isthmuscingulate	✓		✓	
R.lateraloccipital/R.pericalcarine	✓		✓	
R.supramarginal/R.paracentral	✓	✓	✓	
R.cuneus/R.paracentral	✓		✓	
R.supramarginal/R.pericalcarine	✓			
R.lateraloccipital/L.precuneus	✓		✓	
R.supramarginal/L.isthmuscingulate	✓		✓	
R.supramarginal/L.paracentral	✓			
R.cuneus/R.lateraloccipital	✓		✓	✓
R.lateraloccipital/L.isthmuscingulate	✓		✓	
R.supramarginal/R.precuneus	✓	✓	✓	
R.cuneus/R.supramarginal	✓		✓	
R.lateraloccipital/R.supramarginal	✓	✓	✓	
R.cuneus/R.isthmuscingulate			✓	
R.cuneus/R.precuneus		✓	✓	
R.cuneus/L.isthmuscingulate			✓	
R.cuneus/L.paracentral		✓	✓	
R.cuneus/R.pericalcarine		✓	✓	
R.lateraloccipital/R.paracentral				
R.lateraloccipital/R.precuneus		✓	✓	
R.lateraloccipital/L.paracentral			✓	
L.precuneus/R.supramarginal		✓	✓	✓
L.precuneus/R.isthmuscingulate		✓	✓	
L.precuneus/L.isthmuscingulate		✓	✓	
L.precuneus/L.paracentral		✓	✓	
R.supramarginal/R.isthmuscingulate			✓	
R.paracentral/R.isthmuscingulate		✓	✓	
R.paracentral/R.precuneus		✓	✓	
R.paracentral/L.isthmuscingulate		✓	✓	
R.paracentral/L.paracentral		✓	✓	
R.paracentral/R.pericalcarine		✓	✓	
R.isthmuscingulate/L.isthmuscingulate		✓	✓	
R.precuneus/L.isthmuscingulate		✓	✓	✓
R.precuneus/R.pericalcarine				
L.isthmuscingulate/R.pericalcarine			✓	✓
L.paracentral/R.pericalcarine		✓	✓	