Appendix A. Mediation Bi-luster and Effect Estimation

Appendix A.1. ***Dense*** *Bi-Cluster*

We now formally define the dense bi-cluster. First of all, the bi-cluster refers to the fact that we are clustering both the rows (mediators) and columns (outcomes) of a matrix (weight matrix $\mathbf{W}$). Using the notation of bi-cluster, $U_c \otimes V_d$, we give the definition of dense bi-cluster as:

Definition 1. *A bi-cluster $U \otimes V$ is said to be dense if:*

$$\frac{\sum_{\{i,j\} \in U \otimes V} I[w_{ij} > r]}{|U||V|} \geq \zeta,$$

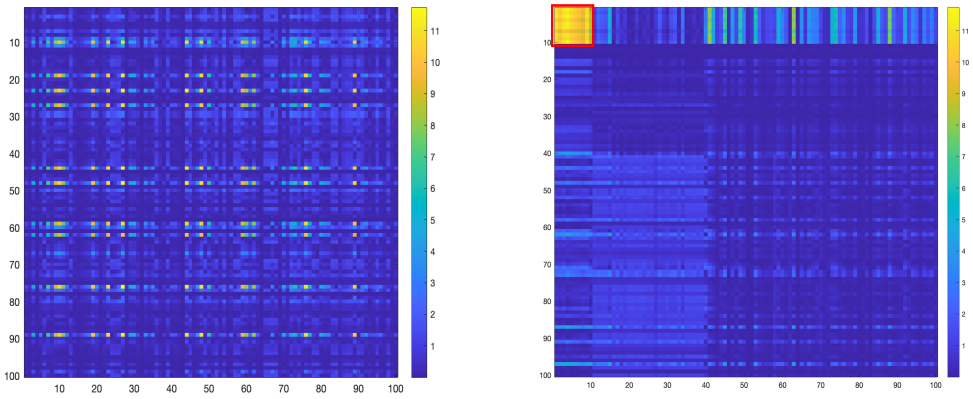
where r is some predefined edge-wise threshold and ζ is bi-cluster density threshold for $U \otimes V \subset G$ in a bipartite graph $G = (U, V, E)$.

Here, the threshold r can be any value predefined, for example, the integration method described in Section 2.2.1. ζ represents a bi-cluster *density* threshold, which ensures that $U \otimes V \subset G$ cannot be a subgraph of a random graph $G = (U, V, E)$ and thus is a latent dense bi-cluster (Wu et al., 2021a). In practice, $\zeta = 0.5$–1 can effectively prohibit false positively selecting a dense bi-cluster. We let $\zeta = 0.8$ in our application. Figure A.1 illustrated how a dense bi-cluster after extraction.

Appendix A.2. Proof of Theorem 1

Proof: We consider that the optimal solution is $\{\mathcal{M}_c, \mathcal{Y}_d\}$, which leads to a mixture distribution of δ_{ij}:

$$P(\delta_{ij}) = \begin{cases} \text{Bern}(\pi_1) & \text{if } \{i,j\} \in \{U_c \otimes V_d\}, \\ \text{Bern}(\pi_0) & \text{otherwise.} \end{cases}$$



(a) (b)

Figure A.1: Demonstration of a *dense* bi-cluster. The left figure is a $-\log(p)$-matrix (weight matrix) before extraction, while the right figure is a p-matrix with re-ordered region indices to highlight the extracted *dense* bi-cluster. A *dense* pattern refers to that within an extracted bi-cluster the proportion of edges with suprathreshold p-values is much higher than the edges in the rest of the graph.

Let $\{\mathcal{M}'_c, \mathcal{Y}'_d\}$ be a non-optimal distribution. Correspondingly, the distribution of δ_{ij} is

$$P'(\delta_{ij}) = \begin{cases} \text{Bern}(\pi'_1) & \text{if } \{i,j\} \in \{U'_c \otimes V'_d\}, \\ \text{Bern}(\pi'_0) & \text{otherwise.} \end{cases}$$

We first calculate the KL functions in 2.5 as $D(P||Q)$ and $D(P'||Q)$ based on $\{\mathcal{M}_c, \mathcal{Y}_d\}$ and $\{\mathcal{M}'_c, \mathcal{Y}'_d\}$ respectively. Define the probability μ to be the probability of being in the $\{U_c \otimes V_d\}$, the estimate of μ can be calculated by $\hat{\mu} = \frac{N_1}{N}$, N_1 is the size of $\{U_c \otimes V_d\}$ and N is the total size.

Define $N_1(1) = \sum_{i,j \in \{U_c \otimes V_d\}} I[\delta_{ij} = 1], N_1(0) = \sum_{i,j \in \{U_c \otimes V_d\}} I[\delta_{ij} = 0]$ as the number of $\delta_{ij} = 1$ and $\delta_{ij} = 0$ in the $\{U_c \otimes V_d\}$ sets. Similarly, $N_0(1) = \sum_{i,j \notin \{U_c \otimes V_d\}} I[\delta_{ij} = 1], N_0(0) = \sum_{i,j \notin \{U_c \otimes V_d\}} I[\delta_{ij} = 0]$ be the number of

$\delta_{ij} = 1$ and $\delta_{ij} = 0$ in the complement of $\{U_c \otimes V_d\}$. For $\{\mathcal{M}'_c, \mathcal{Y}'_d\}$, define $N'_1(1), N'_1(0), N'_0(1), N'_0(0)$ correspondingly. We have

$$\begin{aligned} D(P||Q) =& \frac{1}{N}\left[N_1(1)(\log(\pi_1) - \log(\pi)) + N_1(0)(\log(1-\pi_1) - \log(1-\pi))\right. \\ &\left. + N_0(1)(\log(\pi_0) + \log(\pi)) + N_0(0)(\log(1-\pi_0) + \log(1-\pi))\right], \end{aligned}$$

approximate equation 2.8 when $N \to \infty$.

Using the fact

$$\begin{aligned} N_1(1) + N_0(1) &= N'_1(1) + N'_0(1), \\ N_1(0) + N_0(0) &= N'_1(0) + N'_0(0), \end{aligned}$$

and

$$\begin{aligned} D(P||Q) - D(P'||Q) =& \frac{1}{N}\left[N_1(1)\log(\pi_1) - N'_1(1)\log(\pi'_1)\right. \\ &+ N_1(0)\log((1-\pi_1) - N'_1(0)\log(1-\pi'_1) \\ &+ N_0(1)\log(\pi_0) - N'_0(1)\log(\pi'_0) \\ &\left. + N_0(0)\log(1-\pi_0) - N'_0(0)\log(1-\pi'_0)\right], \end{aligned}$$

is equivalent to comparing the difference between the likelihoods for those two distributions. By Bickel et al. (2013) and Wang and Bickel (2017), the later model has less likelihood than the ture one.

Therefore, the KL criterion $D(P||Q)$ of the optimal solution is greater than KL criterion $D(P'||Q)$ of the non optimal solution. □

Appendix A.3. Choice of Loss Functions

To complete our algorithm by tuning the unknown parameter λ in objective function 2.4, we use a proposed Kullback–Leibler (KL) divergence function 2.5 and maximize the divergence to obtain the best λ. Under the distribution in equation 2.6, the estimate of the unknown probabilities in this model is given by the maximum likelihood estimates:

$$\begin{aligned}\hat{\pi}_1^{MLE} =& \frac{\sum_{c,d}\sum_{i\in U_c, j\in V_d}\delta_{ij}}{\sum_{c,d}|U_c||V_d|},\\ \hat{\pi}_0^{MLE} =& \frac{\sum_{i\in U, j\in V}\delta_{ij} - \sum_{c,d}\sum_{i\in U_c, j\in V_d}\delta_{ij}}{|U||V| - \sum_{c,d}|U_c||V_d|}.\end{aligned} \tag{A.1}$$

U_c, V_d are the selected sets consists the *bi-cluster*. The calculation of MLE is similar to Wu et al. (2020) and details discussions the relative theories can be also found in Wu et al. (2021b).

Another choice of the loss function is to use the entropy function based on the binary entropy:

$$\begin{aligned}H =& \arg\max_{\lambda} -\sum_{c,d}|U_c||V_d|[\hat{\pi}_1\log(\hat{\pi}_1) + (1-\hat{\pi}_1)\log(1-\hat{\pi}_1)]\\ &+ |U\setminus\{U_c\}||V\setminus\{V_d\}|[\hat{\pi}_0\log(\hat{\pi}_0) + (1-\hat{\pi}_0)\log(\hat{\pi}_0)],\end{aligned}$$

for λ selection similar as in Kenley and Cho (2011). Another consideration is to maximum a log-likelihood function:

$$\begin{aligned}\log L =& \arg\max_{\lambda}\sum_{c,d}[\sum_{i\in U_c, j\in V_d}\delta_{ij}\log(\hat{\pi}_1) + \sum_{i\in U_c, j\in V_d}(1-\delta_{ij})\log(1-\hat{\pi}_1)\\ &+ \sum_{i\in U\setminus\{U_c\}, j\in V\setminus\{V_d\}}\delta_{ij}\log(\hat{\pi}_0) + \sum_{i\in U\setminus\{U_c\}, j\in V\setminus\{V_d\}}(1-\delta_{ij})\log(1-\hat{\pi}_0)].\end{aligned}$$

There is numerical evidence showing that KL is superior to the likelihood function and the entropy function, as it is more sensitive to the change of density in the subgraph (Table A.1). In this table, we simulate a dense graph with 10×10 signal area and see the performance of those three functions under different selected areas containing the actual area. Density in the original 10×10 differs, and to compare them consistently, we take the negative entropy (-H), and we tend to choose the subgraph that maximum the function values. In the sparse subgraph situation, we can see that the log-likelihood function has a higher value under a misselected region (the red box). Overall, Kullback–Leibler divergence is more sensitive to signal density as the values diverge more than the other two functions. The null distribution Q eliminates the influence of the π_0 in P as $\pi \approx \pi_0$. This approximate equal is plausible for the assumption that systematic patterns require the distribution outside the $U_c \otimes V_d$ to be random.

80% Density Subgraph Size	LogLike	-H	Kullback–Leibler
10×10	-499.12	-428.46	315.32
20×20	-566.32	-515.34	28.18
40×40	-556.86	-532.43	-29.95
5×5	-591.99	-716.86	42.42
60% Density Subgraph Size	LogLike	-H	Kullback–Leibler
10×10	-520.49	-449.84	169.74
20×20	-536.94	-485.96	10.24
40×40	-498.70	-474.27	-23.97
5×5	-554.31	-639.77	43.77

Table A.1: Performance of Different Function in order to select λ

Appendix A.4. Causal Mediation Estimation

The causal mediation analysis is often used for experimental datasets with manipulable exposure and mediators. Here, we provide the assumptions for causal mediation analysis. Using the potential outcome notation (Rubin, 2005), let $\mathbf{Z}$ as the observed confounders, we adopted the sequential ignorability assumption (Imai et al., 2010a; Chén et al., 2018) as:

$$\begin{aligned} \tilde{\mathbf{M}}(\mathbf{x}) &\perp\!\!\!\perp \mathbf{X}|\mathbf{Z}, \\ \mathbf{Y}(\mathbf{x}, \tilde{\mathbf{m}}) &\perp\!\!\!\perp \tilde{\mathbf{M}}|\mathbf{X}, \mathbf{Z}, \\ \mathbf{Y}(\mathbf{x}, \tilde{\mathbf{M}}(\mathbf{x})) &\perp\!\!\!\perp \mathbf{X}|\mathbf{Z}, \\ \mathbf{Y}(\mathbf{x}, \tilde{\mathbf{m}}) &\perp\!\!\!\perp \tilde{\mathbf{M}}(\mathbf{x}^*)|\mathbf{Z}, \end{aligned} \tag{A.2}$$

which indicates that i) no unmeasured confounders between exposure and mediating factors, ii) no unmeasured confounders between mediating factors and outcomes, iii) no unmeasured confounders between exposure and outcomes, and iv) no confounders for the relationships between mediating factors and outcomes which are affected by the exposure. In addition to sequential ignorability, the identifiability in the causal mediation analysis further requires that the mediators (or mediator factors) are conditionally independent given the exposure (i.e., causally independent mediators, Imai and Yamamoto, 2013; Huang and Pan, 2016).

Under these strong assumptions, for each $\mathbf{Y}_d^{(j)} \in \mathcal{Y}_d$ we calculate the average natural direct effect (NDE), average natural indirect effect (NIE),

and average total effect (ATE) on the orthogonal mediators:

$$\begin{aligned}
\mathrm{NDE}(\mathbf{x})_j &= \mathrm{E}(\mathbf{Y}_d^{(j)}(\mathbf{x}, \tilde{\mathbf{M}}_c(\mathbf{x}^*)) - \mathbf{Y}_d^{(j)}(\mathbf{x}^*, \tilde{\mathbf{M}}_c(\mathbf{x}^*))) = \theta_j(\mathbf{x} - \mathbf{x}^*), \\
\mathrm{NIE}(\mathbf{x})_j &= \mathrm{E}(\mathbf{Y}_d^{(j)}(\mathbf{x}, \tilde{\mathbf{M}}_c(\mathbf{x})) - \mathbf{Y}_d^{(j)}(\mathbf{x}, \tilde{\mathbf{M}}_c(\mathbf{x}^*))) = \sum_{l=1}^{L_c} \beta_{lj}\alpha_l(\mathbf{x} - \mathbf{x}^*), \\
\mathrm{ATE}(\mathbf{x})_j &= \mathrm{NDE}(\mathbf{x})_j + \mathrm{NIE}(\mathbf{x})_j = (\sum_{l=1}^{L_c} \beta_{lj}\alpha_l + \theta_j)(\mathbf{x} - \mathbf{x}^*),
\end{aligned} \tag{A.3}$$

where $\tilde{\mathbf{M}}_c$ is the observed matrix for $\tilde{\mathcal{M}}_c, c = 1, \ldots, C$.

Appendix A.5. Algorithm

We summarize the algorithm described in step 1 (Section2.2.1) as in Algorithm 1.

Appendix A.6. Mediation Effect Estimation

Estimation by Partical Correlation: Consider for estimated $\{\mathcal{M}_c, \mathcal{Y}_d\}$ and corresponding factors $\tilde{\mathbf{M}}_c = \mathbf{M}_c\eta_c + \epsilon_c$. Using the same indexes as in (A.3), let $\epsilon_{Yd}^{(j)}$ and $\epsilon_{Mc}^{(l)}$ be the residuals of regressing each mediator factor ($\tilde{\mathbf{M}}_c^{(l)}$) on exposure and each outcome ($\mathbf{Y}_d^{(j)}$) on exposure. The indirect mediation effect calculated by partial correlation then is given by:

$$\hat{IE}_{\rho d} = \frac{1}{J_d}\sum_{j=1}^{J_d}\sum_{l=1}^{L_c} \hat{\rho}_{\epsilon_{Yd}^{(j)}, \epsilon_{Mc}^{(l)}} \hat{\rho}_{X, \tilde{\mathbf{M}}_c^{(l)}}, \tag{A.4}$$

where $\rho_{s,v}$ is the correlation between variables s, v.

Algorithm 1 Mediation dense bi-cluster extraction

repeat
  for λ in a given λ range **do**
    $S \leftarrow U$, $T \leftarrow V$
    for $iter = 1$ to $|S| + |T| - 1$ **do**
      if $|S| > 0$ and $|T| > 0$ **then**
        $o = \min_{s \in S} \sum_{t \in T} |\mathbf{W}(s,t)|, d_S = \sum_{t \in T} |\mathbf{W}(o,t)|$
        $z = \min_{t \in T} \sum_{s \in S} |\mathbf{W}(s,t)|, d_T = \sum_{s \in S} |\mathbf{W}(s,z)|$
        if $\sqrt{k} d_X \leq d_T / \sqrt{k}$ **then**
          $S \leftarrow S / \{o\}$
        else
          $T \leftarrow T / \{z\}$
        end if
      else
        Value of objective function ((2)) is 0
      end if
      Calculate function ((2)) with $U_c = S, V_d = T$
    end for
    Output the subgraph $\{\hat{U}_c \otimes \hat{V}_d(\lambda)\}$ with the maximum value of the objective function
    Calculate the divergence function ((6))
  end for
  Output the subgraph and the corresponding λ with largest divergence $D(P_{\{\hat{U}_c \otimes \hat{V}_d(\lambda)\}})$
  $U \leftarrow U \setminus \{\hat{U}_c\}, V \leftarrow V \setminus \{\hat{V}_d\}$
until $D(P_{\hat{U}_c \otimes \hat{V}_d(\lambda)})$ has an absolute difference of less than ξ

Appendix B. Data Example

Appendix B.1. Data Acquisition and Processing

The Arterial-spin labeling (ASL) data are acquired on a 3-T Siemens Prisma scanner with 64 channels. Three-dimensional (3D) pseudo-continuous ASL (pCASL) with backgrounds suppressed gradient and spin-echo (GRASE) sequence consisting of 13 pairs of labeled and control scans are used. Parameters are spatial resolution equals to 2.5mm $\times$ 2.5mm $\times$ 2.6mm with matrix size 96×96 and 58 axial slices. Time/Echo time repeated is $4,000/37$ms with 120° flip angle, 220mm field of view (FoV) read, 100% FoV phase, $1,700$ms post-label delay and 1650ms labeling duration. The total scan time is about 10min. A 3D T_1 weighted image is acquired for anatomical reference, and the gray and white matter tissue segmentation. The parameters are: TR = $2,400$ms, TE = 2.22ms, TI = $1,000$ms, flip angel = 8°, matrix = 320×320, slices per slab = 208, and 0.8mm $\times$ 0.8mm spatial resolution with 0.8mm slice thickness. The volume of M_0 image is also acquired without background suppression to normalize the control-label difference for CBF quantification. This image is smoothed with a 5mm Gaussian-kernel to suppress the effects of noise (Alsop et al., 2015).Participants with lifetime diagnosis of psychiatric disorders are excluded in order to avoid confounded effects from psychiatric diseases.

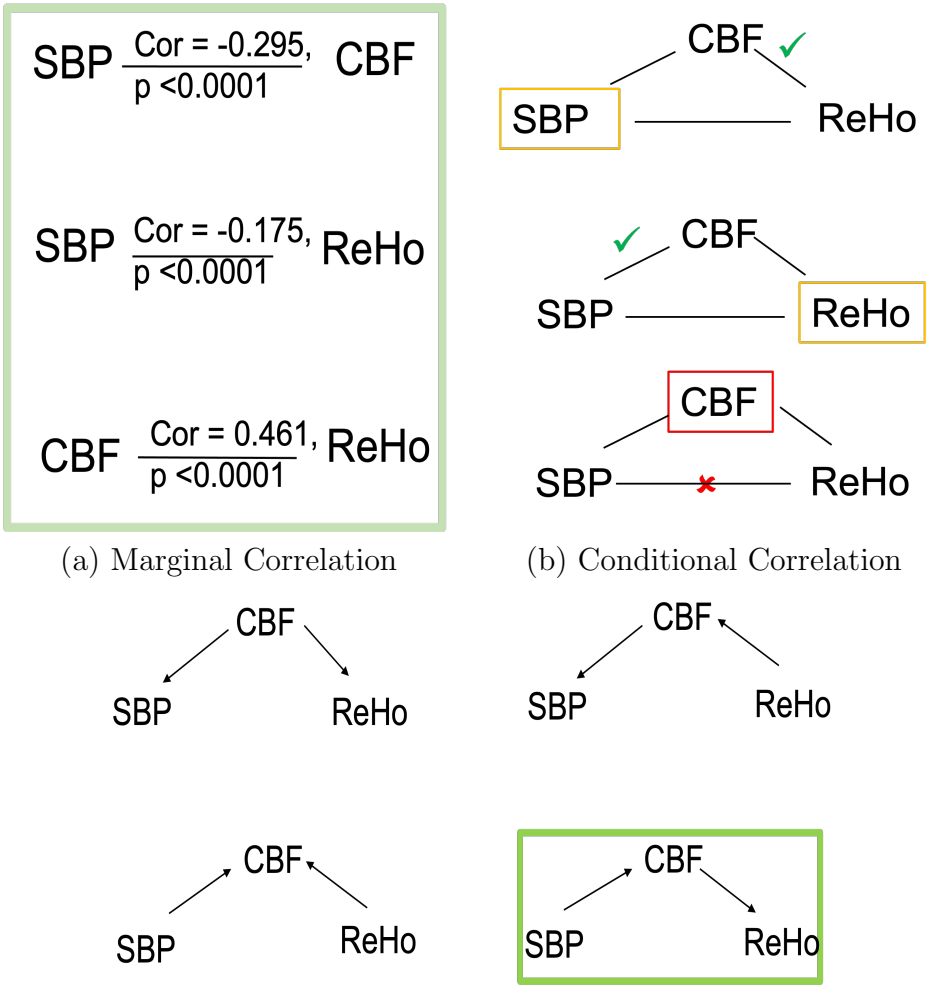
The rsfMRI data acquire from two runs. Oblique axial acquisitions alternated between phase encoding in the anterior-to-posterior(AP) and posterior-to-anterior(PA) directions are from a single run. Separate single-band reference images acquired for phase encoding in AP and PA directions are used for spatial distortion correction. The parameters are: TR=780ms,

TE=34.4ms,spatial resolution of 2-mm isotropic voxels, matrix size $=104 \times 104$ with 72 axial slices, number of volumes=420/run, flip angel=52°, multi-band acceleration factor is 8 and 2, 186, Hz/pixel bandwidth. The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) is used to process the rsfMRI data (Adhikari et al., 2018, 2022). Motions are estimated as the magnitude of the displacement from one-time point to the next, including neighboring time points and outliers voxels fraction (> 0.1). Time points with excessive motion (> 0.2mm) are excluded from the analysis.

Appendix B.2. Mediation Pathways

We used causal discovery method (Glymour et al., 2019) to determine the mediation directed acyclic graph (DAG). We started with the marginal correlation between SBP, average CBF and average ReHo (Glodzik et al., 2019). Correlation tests showed significant relation between pairs of the three variables with p-values all < 0.0001, correlation between SBP and average CBF is -0.295, between SBP and average ReHo is -0.175, and between average CBF and average ReHo is 0.461. [Figure B.2(a)]. Conditional correlation tests revealed no-significant relation between SBP and average ReHo, conditional on average CBF ($p = 0.5306$, correlation is -0.046, **conditional independence**), but still significant relation between SBF and average CBF ($p = 0.001$, correlation is -0.246) and significant relation between average CBF and average ReHo ($p < 0.0001$, correlation is 0.435). [Figure B.2(b)]. The possible DAGs between the three variables then are showed in Figure B.2(c). Biologically, no previous research has suggested that CBF can influence SBP. Therefore, we exclude the first two DAGs in Figure B.2(c). Further, the existence of marginal correlation between SBP

and average ReHo exclude the situation where average CBF is a collider [bottom left DAG in Figure B.2(c)]. Therefore, the DAG in the green box in Figure B.2(c) is the most plausible DAG for our mediation analysis.



(a) Marginal Correlation

(b) Conditional Correlation

(c) Possible DAGs

Figure B.2: The Mediation Discovery procedure for determining the mediation pathway. (a) is the marginal correlation between SBP, CBF, ReHo. (b) is the conditional correlation between the three variables, conditional on the boxed variables. The conditional correlation between SBP and ReHo given CBF is not significant. (c) is all the possible DAGs based on the results in (b). The green box is the most possible DAG for our data.

Appendix B.3. Sensitivity Analysis

In practice, assumptions in (A.2) for an observational study may need to be validated. Imai et al. (2010b) suggested a parametric sensitivity analysis based on the residual correlations. Following this procedure, we conducted a sensitivity analysis for our data example. Specifically, we considered the violation of the sequential ignorability with the correlation of residuals for (1) (defined as ρ) is correlated. When the assumption holds, we should have $\rho = 0$. We let ρ vary and evaluate the performance of our method. The sensitivity analysis results are demonstrated in Figure B.3. In our analysis, the assumption appears valid because $\hat{\rho} \approx 0$. The estimated mediation effect is -0.094. In the sensitivity analysis, ρ deviates from 0, and the corresponding estimation of the mediation effect differs from -0.094. The bias is relatively small when ρ is small to medium (e.g., < 0.3). The sign of mediation effect estimation is changed when $\hat{\rho} > 0.375$. Therefore, we conclude that our results are generally robust to the (mild or moderate) violation of the assumption. (Imai et al., 2010b).

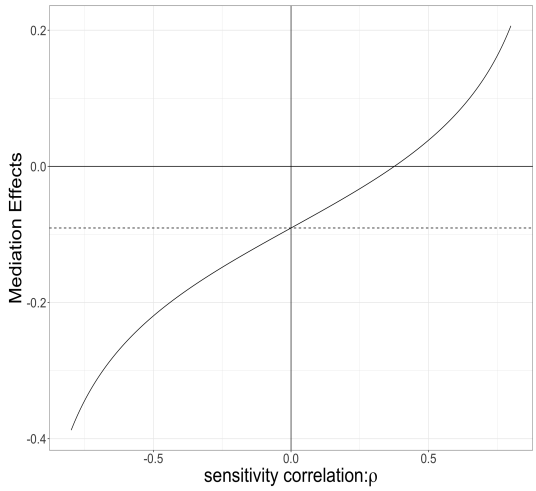


Figure B.3: Sensitivity analysis to assess the impact of the violated assumptions for our data example. The sensitivity analysis is performed based on the procedure described in Imai et al. (2010b). The sequential ignorability holds when $\rho = 0$ (ρ is the correlation of residuals of mediators and outcomes conditional on the exposure). The dashed horizontal line is the estimated mediation effect with the valid assumption of sequential ignorability. When ρ deviates further from 0, the bias is larger. The results suggested that our mediation results were generally robust to (mild to moderate) violation of this assumption (e.g., $\rho = 0$–0.375).

Appendix B.4. Selected Regions

Selected regions for CBF include CGC_L, CingG_L, SMG_R, SMG_L, IOG_R, CingG_R. Selected ReHo regions contain SPG_L, CingG_L, SFG_L, MFG_L IFG_L, PrCG_L, PoC_L, AG_L, PrCu_L, Fu_L, PHG_L, SOG_L, IOG_L, ENT_L, STG_L, ITG_L, LFOG_L, MFOG_L, RG_L, Ins_L, Hippo_L, cerebrellum_L, SCR_L, CGC_L, GCC_L, Caud_L, Put_L, Thal_L, Midbrain_L, SPG_R, CingG_R, SFG_R, MFG_R, IFG_R, FuG_R, PHG_R, ENT_R, STG_R, ITG_R, MTG_R, LFOG_R, MFOG_R, SMG_R, RG_R, Ins_R, Amyg_R, Hippo_R, cerebellum_R, CST_R, ICP_R, SCP_R, SCR_R, CGC_R, CGH_R, EC_R, Put_R, Thal_R, GP_R and Medulla_R. The corresponding region names are available online at https://github.com/zhuivv/MMO/blob/main/Suppl_II_regionnames.

txt or the online supplements.

Appendix C. Simulation

Appendix C.1. Data Generation

To generate $(X, \tilde{\mathbf{M}}, \mathbf{Y}) \sim N(\mu, \Sigma)$, we first specify the precision matrix $\Omega = \Sigma^{-1}$. The off-diagonal elements in Ω is the partial correlations. Let ω_{ij} be the elements in Ω, the subscripts represent the rows and columns of Ω and hence correspond to the univariate X, elements in $\tilde{\mathbf{M}}$ and $\mathbf{Y}$. We set $\omega_{ij} \neq 0$ if $\beta_{ij} \neq 0$ as in Equation 2.1 for i, j representing elements in $\tilde{\mathbf{M}}$ and $\mathbf{Y}$. Similarly, let σ_{ij} be the element in Σ and set $\sigma_{1j} \neq 0$ if $\alpha_j \neq 0$ for j representing elements in $\tilde{\mathbf{M}}$.

Appendix C.2. Additional Simulations

Appendix C.2.1. Non-normally distributed mediators and outcomes

To further assess our method, we considered two additional settings for simulation. Specifically, we generated data from the Cauchy distribution and Laplace distribution $(X, \tilde{\mathbf{M}}, \mathbf{Y}) \sim \text{Cauchy}(\mu, \Sigma)$ and $(X, \tilde{\mathbf{M}}, \mathbf{Y}) \sim \text{Laplace}(\mu, \Sigma)$. We let $n = 200$ for each setting and set the effect size as 0.24 and 0.16. We show the results for bi-cluster mediation pathway extraction in Table C.2, and the results of mediation effect estimation in Table C.3. The performance of MMO under these settings appear similar to the results in normally distributed data settings.

Mediation effect size and Sample size	Method	Cluster Size = 10×10			Cluster Size = 20×20		
		FDR	sens	spec	FDR	sens	spec
Effect=0.24 Cauchy	BH	0.030(0.110)	0.434(0.445)	0.999(0.003)	0.039(0.115)	0.404(0.439)	0.997(0.010)
	3-step	0.049(0.129)	0.510(0.390)	0.999(0.002)	0.103(0.181)	0.787(0.263)	0.993(0.012)
	PathL	0.066(0.085)	0.671(0.330)	0.992(0.011)	0.0922(0.102)	0.743(0.268)	0.980(0.027)
	MMO	0.039(0.101)	0.824(0.254)	0.9995	0.065(0.134)	0.781(0.293)	0.996(0.009)
Effect=0.24 Laplace	BH	0.002(0.014)	0.757(0.425)	1	0.002(0.014)	0.798(0.398)	1^*
	3-step	0.060(0.095)	0.950(0.224)	0.999(0.001)	0.011(0.030)	0.967(0.193)	1^*
	PathL	0.011(0.035)	0.680(0.193)	0.999(0.004)	0.048(0.073)	0.394(0.098)	0.994(0.009)
	MMO	0.020(0.056)	0.998(0.026)	1^*	0.009(0.036)	0.998(0.026)	1^*
Effect=0.16 Cauchy	BH	0.020(0.081)	0.309(0.425)	1^*	0.039(0.109)	0.366(0.432)	0.997(0.009)
	3-step	0.010(0.146)	0.835(0.311)	0.999(0.002)	0.142(0.185)	0.676(0.377)	0.990(0.014)
	PathL	0.045(0.091)	0.544(0.436)	0.949(0.084)	0.026(0.055)	0.680(0.253)	0.996(0.008)
	MMO	0.040(0.103)	0.808(0.267)	1^*	0.084(0.133)	0.714(0.297)	0.996(0.009)
Effect=0.16 Laplace	BH	0.006(0.027)	0.497(0.497)	1^*	0.005(0.018)	0.632(0.474)	1^*
	3-step	0.024(0.035)	0.898(0.316)	1^*	0.020(0.038)	0.785(0.416)	0.999(0.002)
	PathL	0.033(0.703)	0.510(0.129)	0.998(0.005)	0.055(0.091)	0.345(0.086)	0.994(0.011)
	MMO	0.057(0.085)	1^*	0.999(0.001)	0.046(0.071)	0.991(0.053)	0.998(0.004)

Table C.2: Simulation results (Step 1) for simulated datasets with non-normal mediators and outcomes: the accuracy of mediation pattern $U_c \otimes V_d$ extraction. We demonstrate the edge-wise false discovery rate (FDR), sensitivity (sens), and specificity (spec). The standard Cauchy distribution and Laplace distribution are used to generate the mediators and outcomes of the synthetic datasets with the sample size is $n = 200$. $*$ represents a rounded number.

Effect size and Sample size	Method	signal region = 10×10			signal region = 20×20		
		Mean	Bias	Coverage Prob	Mean	Bias	Coverage Prob
effect=0.24 Cauchy	'Oracle Model'	0.285(0.372)			0.255(0.360)		
	medLRM	0.029(0.132)	0.257(0.358)	6.2%	0.040(0.183)	0.215(0.316)	18.0%
	MMO	0.235(0.320)	0.051(0.093)	79.0%	0.211(0.306)	0.045(0.095)	81.2%
effect=0.24 Laplace	'Oracle Model'	0.294(0.060)			0.283(0.058)		
	medLRM	0.081(0.106)	0.213(0.102)	16.5%	0.204(0.110)	0.079(0.093)	68.9%
	MMO	0.287(0.062)	0.008(0.017)	98.8%	0.279(0.060)	0.004(0.010)	99.52%
effect=0.16 Cauchy	'Oracle Model'	0.183(0.313)			0.232(0.333)		
	medLRM	0.011(0.098)	0.172(0.297)	6.0%	0.048(0.162)	0.184(0.293)	11.7%
	MMO	0.151(0.272)	0.032(0.079)	81.2%	0.186(0.255)	0.045(0.113)	73.8%
effect=0.16 Laplace	'Oracle Model'	0.192(0.052)			0.195(0.052)		
	medLRM	0.031(0.051)	0.161(0.060)	6.4%	0.040(0.064)	0.155(0.071)	13.6%
	MMO	0.180(0.053)	0.012(0.018)	96.2%	0.184(0.053)	0.011(0.015)	97.3%

Table C.3: Simulation results (Step 2) for simulated datasets with non-normal mediators and outcomes: mediation effect estimation. We perform Step 2 analysis based on the above results (Step 1 analysis results based on datasets generated by the Cauchy distribution or Laplace distribution) and compare the estimated mediation effects by Low Rank Model (medLRM) and **MMO**, with reference to the estimated mediation effect based on the oracle model (with known mediating imaging factors).

Appendix C.2.2. Non-orthogonal Mediating Factors

To evaluate our method's performance under the setting of non-orthogonal mediating factors, we conducted additional simulation analysis. Specifically, we set the correlations between correlated mediating factors $\tilde{\mathbf{M}}$ between 0.5–0.8. We simulated datasets with $n = 200$ and the two effect sizes 0.16 and 0.24. The results are shown in Table C.4 and Table C.5. The results suggest that the performance of Step 1 in our model is invariant to the correlated mediating factors. However, the estimation of the mediation effects in Step 2 can be affected because the regression models can be unstable due to collinearity.

Mediation effect size and Sample size	Method	Cluster Size = 10×10			Cluster Size = 20×20		
		FDR	sens	spec	FDR	sens	spec
Effect=0.24	BH	0.054(0.093)	0.547(0.483)	0.999(0.001)	0.052(0.061)	0.736(0.426)	0.998(0.003)
	3-step	0.285(0.150)	0.992(0.034)	0.995(0.003)	0.187(0.068)	0.997(0.001)	0.990(0.004)
	PathL	0.099(0.151)	0.520(0.132)	0.991(0.017)	0.036(0.083)	0.325(0.089)	0.996(0.008)
	MMO	0.094(0.160)	0.923(0.100)	0.999(0.003)	0.075(0.104)	0.937(0.077)	0.996(0.006)
Effect=0.16	BH	0	0.305(0.454)	1	0^*	0.368(0.478)	1
	3-step	0.025(0.096)	0.675(0.452)	1^*	0	0.775(0.329)	1
	PathL	0.05(0.158)	0.490(0.160)	0.997(0.011)	0	0.270(0.098)	1
	MMO	0.029(0.090)	0.990(0.074)	1^*	0.048(0.128)	0.994(0.058)	0.997(0.011)

Table C.4: Simulation results (Step 1) for simulated datasets with non-orthogonal mediators: the accuracy of mediation pattern $U_c \otimes V_d$ extraction. We demonstrate the edge-wise false discovery rate (FDR), sensitivity (sens) and specificity (spec). The correlations of non-orthogonal factors vary from 0.5 to 0.8 with sample size is $n = 200$. $*$ represents a rounded number.

Effect size and Sample size	Method	signal region = 10×10		signal region = 20×20	
		Mean	Bias	Mean	Bias
effect=0.24	'Oracle Model'	0.226(0.032)		0.228(0.030)	
	medLRM	0.373(0.058)	0.143(0.038)	0.394(0.047)	0.162(0.028)
	MMO	0.388(0.057)	0.159(0.035)	0.396(0.048)	0.164(0.029)
effect=0.16	'Oracle Model'	0.162(0.031)		0.159(0.032)	
	medLRM	0.193(0.041)	0.032(0.015)	0.172(0.058)	0.021(0.039)
	MMO	0.199(0.043)	0.038(0.015)	0.174(0.059)	0.023(0.039)

Table C.5: Simulation results (Step 2) for simulated datasets with non-orthogonal mediators: the estimated mediation effects. We compare the estimated mediation effects by Low Rank Model (medLRM) and **MMO**, with reference to the estimated mediation effect based on the oracle model (with known mediating imaging factors). The correlations of simulated non-orthogonal factors vary from 0.5 to 0.8 with$n = 200$. $*$ represents rounded numbers.

Appendix C.3. Low-rank mediators

Low-rank models are commonly used to reduce the dimensionality of the multivariate mediators in the neuroimaging study. For example, PCA and factor analysis methods have been widely applied to multivariate mediation analysis with neuroimaging data as mediators (Chén et al., 2018; Zhao et al., 2020). Alternatively, a composite mediator can be constructed based on the weighted sum from the **W** matrix. We further evaluate the performance of mediation effect estimation using these methods. Figure C.4 illustrates the results: both methods perform well while the mediating factors by factorization method slightly outperforms the cumulative weight composite mediator. One possible reason could be that the cumulative weight method may have included insignificant weights as noise.

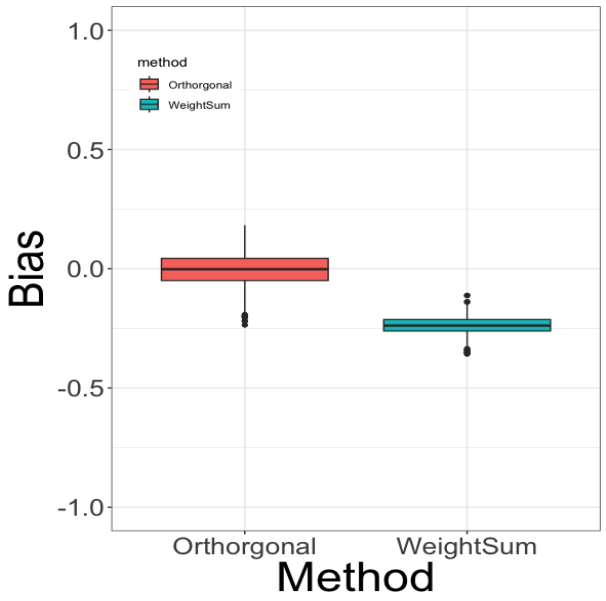


Figure C.4: Results of mediation effect estimation based on mediators by orthogonal factorization method and a composite mediator based on the weighted sum method. Based on 1 000 simulations, both methods perform well, while the factorization method slightly outperforms the weighted sum component method.

References

Adhikari, B.M., Hong, L.E., Zhao, Z., Wang, D.J., Thompson, P.M., Jahanshad, N., Zhu, A.H., Holiga, S., Turner, J.A., van Erp, T.G., et al., 2022. Cerebral blood flow and cardiovascular risk effects on resting brain regional homogeneity. NeuroImage 262, 119555.

Adhikari, B.M., Jahanshad, N., Shukla, D., Glahn, D.C., Blangero, J., Reynolds, R.C., Cox, R.W., Fieremans, E., Veraart, J., Novikov, D.S., et al., 2018. Heritability estimates on resting state fmri data using enigma analysis pipeline, in: PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium, World Scientific. pp. 307–318.

Alsop, D.C., Detre, J.A., Golay, X., Günther, M., Hendrikse, J., Hernandez-Garcia, L., Lu, H., MacIntosh, B.J., Parkes, L.M., Smits, M., et al., 2015. Recommended implementation of arterial spin-labeled perfusion mri for clinical applications: A consensus of the ismrm perfusion study group and the european consortium for asl in dementia. Magnetic resonance in medicine 73, 102–116.

Bickel, P., Choi, D., Chang, X., Zhang, H., 2013. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. The Annals of Statistics 41, 1922–1943.

Chén, O.Y., Crainiceanu, C., Ogburn, E.L., Caffo, B.S., Wager, T.D., Lindquist, M.A., 2018. High-dimensional multivariate mediation with application to neuroimaging data. Biostatistics 19, 121–136.

Glodzik, L., Rusinek, H., Tsui, W., Pirraglia, E., Kim, H.J., Deshpande, A., Li, Y., Storey, P., Randall, C., Chen, J., et al., 2019. Different relationship between systolic blood pressure and cerebral perfusion in subjects with and without hypertension. Hypertension 73, 197–205.

Glymour, C., Zhang, K., Spirtes, P., 2019. Review of causal discovery methods based on graphical models. Frontiers in genetics 10, 524.

Huang, Y.T., Pan, W.C., 2016. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. Biometrics 72, 402–413.

Imai, K., Keele, L., Tingley, D., 2010a. A general approach to causal mediation analysis. Psychological methods 15, 309.

Imai, K., Keele, L., Yamamoto, T., 2010b. Identification, inference and sensitivity analysis for causal mediation effects. Statistical science 25, 51–71.

Imai, K., Yamamoto, T., 2013. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. Political Analysis 21, 141–171.

Kenley, E.C., Cho, Y.R., 2011. Entropy-based graph clustering: Application to biological and social networks, in: 2011 IEEE 11th International Conference on Data Mining, IEEE. pp. 1116–1121.

Rubin, D.B., 2005. Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association 100, 322–331.

Wang, Y.R., Bickel, P.J., 2017. Likelihood-based model selection for stochastic block models. The Annals of Statistics 45, 500–528.

Wu, Q., Huang, X., Culbreth, A., Waltz, J., Hong, L.E., Chen, S., 2020. Extracting brain disease-related connectome subgraphs by adaptive dense subgraph discovery. bioRxiv .

Wu, Q., Ma, T., Liu, Q., Milton, D.K., Zhang, Y., Chen, S., 2021a. Icn: extracting interconnected communities in gene co-expression networks. Bioinformatics 37, 1997–2003.

Wu, Q., Zhang, Y., Huang, X., Ma, T., Hong, L., Kochunov, P., Chen, S., 2021b. A multivariate to multivariate approach for voxel-wise genome-wide association analysis .

Zhao, Y., Lindquist, M.A., Caffo, B.S., 2020. Sparse principal component based high-dimensional mediation analysis. Computational statistics & data analysis 142, 106835.