Matrix Graph Hypothesis Testing and Application in Brain Connectivity Alternation Detection

Yin Xia and Lexin Li

Fudan University
University of California at Berkeley

Supplementary Material

S1 Lemmas

Define $U_{i,j,g} = \frac{1}{n_g q}\sum_{k=1}^{n_g}\sum_{l=1}^{q}\{\epsilon_{k,i,l}^{(g)}\epsilon_{k,j,l}^{(g)} - E(\epsilon_{k,i,l}^{(g)}\epsilon_{k,j,l}^{(g)})\}$, for $g = 1,2$. The following lemma states the results in the oracle case. Its proof can be obtained from Xia et al. (2015), with $n_g q$ inverse regression models instead.

Lemma 1. *Suppose that (C1) and (C4) hold. Then we have*

$$\max_{1\le i\le p}|\hat{r}_{i,i,g} - r_{i,i,g}| = o_{\rm p}[\{\log p/(nq)\}^{1/2}],$$

and

$$\tilde{r}_{i,j,g} = \tilde{R}_{i,j,g} - \tilde{r}_{i,i,g}(\hat{\beta}_{i,j,g} - \beta_{i,j,g}) - \tilde{r}_{j,j,g}(\hat{\beta}_{j-1,i,g} - \beta_{j-1,i,g}) + o_{\rm p}\{(nq\log p)^{-1/2}\},$$

for $1 \le i < j \le p$, $g = 1,2$, where $\tilde{R}_{i,j,g}$ is the empirical covariance between $\{\epsilon_{k,i,l}^{(g)}, k = 1,\ldots,n_g, l = 1,\ldots,q\}$ and $\{\epsilon_{k,j,l}^{(g)}, k = 1,\ldots,n_g, l = 1,\ldots,q\}$. Consequently, uniformly in $1 \le i < j \le p$,

$$\hat{r}_{i,j,g} - (\omega_{S_g,i,i}\hat{\sigma}_{i,i,\epsilon}^{(g)} + \omega_{S_g,j,j}\hat{\sigma}_{j,j,\epsilon}^{(g)} - 1)r_{i,j,g} = -U_{i,j,g} + o_{\rm p}\{(nq\log p)^{-1/2}\},$$

where $(\hat{\sigma}_{i,j,\epsilon}^{(g)}) = \frac{1}{n_g q}\sum_{k=1}^{n_g}\sum_{l=1}^{q}(\boldsymbol{\epsilon}_{k,,l}^{(g)} - \bar{\boldsymbol{\epsilon}}^{(g)})(\boldsymbol{\epsilon}_{k,,l}^{(g)} - \bar{\boldsymbol{\epsilon}}^{(g)})'$, $\boldsymbol{\epsilon}_{k,,l}^{(g)} = (\epsilon_{k,1,l}^{(g)},\ldots,\epsilon_{k,p,l}^{(g)})$ and $\bar{\boldsymbol{\epsilon}}^{(g)} = \frac{1}{n_g q}\sum_{k=1}^{n_g}\sum_{l=1}^{q}\boldsymbol{\epsilon}_{k,,l}^{(g)}$.

Lemma 2. *Let* $\boldsymbol{X}_k \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ *for* $k = 1, \ldots, n_1$ *and* $\boldsymbol{Y}_k \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ *for* $k = 1, \ldots, n_2$. *Define* $\tilde{\boldsymbol{\Sigma}}_1 = (\tilde{\sigma}_{i,j,1})_{p\times p} = \frac{1}{n_1}\sum_{k=1}^{n_1}(\boldsymbol{X}-\boldsymbol{\mu}_1)(\boldsymbol{X}-\boldsymbol{\mu}_1)^{\mathrm{T}}$ *and* $\tilde{\boldsymbol{\Sigma}}_2 = (\tilde{\sigma}_{i,j,2})_{p\times p} = \frac{1}{n_2}\sum_{k=1}^{n_2}(\boldsymbol{Y}-\boldsymbol{\mu}_2)(\boldsymbol{Y}-\boldsymbol{\mu}_2)^{\mathrm{T}}$. *Then, for some constant* $C > 0$, $\tilde{\sigma}_{i,j,1}-\tilde{\sigma}_{i,j,2}$ *satisfies the large deviation bound,*

$$\begin{aligned}&\mathrm{pr}\left[\max_{(i,j)\in\mathcal{S}} \frac{(\tilde{\sigma}_{i,j,1}-\tilde{\sigma}_{i,j,2}-\sigma_{i,j,1}+\sigma_{i,j,2})^2}{\mathrm{var}\{(X_{k,i}-\mu_{1,i})(X_{k,j}-\mu_{1,j})\}/n_1+\mathrm{var}\{(Y_{k,i}-\mu_{2,i})(Y_{k,j}-\mu_{2,j})\}/n_2} \geq x^2\right]\\ &\leq C|\mathcal{S}|\{1-\Phi(x)\}+O(p^{-1}),\end{aligned}$$

uniformly for $0 \leq x \leq (8\log p)^{1/2}$ *and any subset* $\mathcal{S} \subseteq \{(i,j) : 1 \leq i \leq j \leq p\}$.

Lemma 3. *Suppose that (C1), (C5) and (C6) hold. Then we have, for* $g = 1, 2$, *uniformly in* $1 \leq i \leq j \leq p$,

$$\hat{r}^d_{i,j,g} - (\omega_{S_g,i,i}\hat{\sigma}^{(g)}_{i,i,\epsilon} + \omega_{S_g,j,j}\hat{\sigma}^{(g)}_{j,j,\epsilon} - 1)r_{i,j,g} = -U_{i,j,g} + o_{\mathrm{p}}\{(nq\log p)^{-1/2}\},$$

where $\hat{r}^d_{i,j,g} = -(\tilde{r}^d_{i,j,g} + \tilde{r}^d_{i,i,g}\hat{\beta}^d_{i,j,g} + \tilde{r}^d_{j,j,g}\hat{\beta}^d_{j-1,i,g})$, $\tilde{r}^d_{i,j,g} = 1/(nq)\sum_{k=1}^{n}\sum_{l=1}^{q}\hat{\epsilon}^d_{k,i,l,g}\hat{\epsilon}^d_{k,j,l,g}$, *and* $\hat{\epsilon}^d_{k,i,l,g} = Y^{(g)}_{k,i,l} - (\boldsymbol{Y}^{(g)}_{k,-i,l})^{\mathrm{T}}\hat{\boldsymbol{\beta}}^d_{i,l,g}$ *with* $\boldsymbol{Y}^{(g)}_k = \boldsymbol{X}^{(g)}_k\hat{\boldsymbol{\Sigma}}^{-1/2}_{T_g}$.

This lemma is essentially proved in Theorems 5 and 6 in Xia and Li (2017).

S2 Proofs

S2.1 Proof of Theorem 1

By separation of the spatial and temporal dependence structures, we have the following $2nq$ inverse regression models

$$(\boldsymbol{X}^{(g)}_k\boldsymbol{\Sigma}_T^{-1/2})_{i,l} = (\boldsymbol{X}^{(g)}_k\boldsymbol{\Sigma}_T^{-1/2})_{-i,l}\boldsymbol{\beta}_{i,g} + \epsilon^{(g)}_{k,i,l}, \quad 1 \leq k \leq n, 1 \leq l \leq q, g = 1, 2.$$

We first show that $\hat{t}$, as defined in (9) in Algorithm 1, can be obtained in the range $(0, 2(\log p)^{1/2})$. Then we illustrate that the number of false rejections is close to $2\{1-\Phi(t)\}|\mathcal{H}_0|$, by first showing the terms in A_τ are negligible. We then focus on the set $\mathcal{H}_0 \setminus A_\tau$ and prove the result.

Without loss of generality, throughout this proof, we assume that $\omega_{S_g,i,i} = 1$ for $g = 1, 2$ and $i = 1, \ldots, p$. For $g = 1, 2$, let

$$V_{i,j} = (U_{i,j,2} - U_{i,j,1})/\{\mathrm{var}(\epsilon^{(1)}_{k,i,l}\epsilon^{(1)}_{k,j,l})/(n_1q) + \mathrm{var}(\epsilon^{(2)}_{k,i,l}\epsilon^{(2)}_{k,j,l})/(n_2q)\}^{1/2},$$

where $\mathrm{var}(\epsilon_{k,i,l}^{(g)}\epsilon_{k,j,l}^{(g)}) = r_{i,i,g}r_{j,j,g}(1+\eta_{i,j,g}^2)$ with $\eta_{i,j,g}^2 = \beta_{i,j,g}^2 r_{i,i,g}/r_{j,j,g}$. By Lemma 1, we have

$$\max_{1\le i\le p}|\hat{r}_{i,i,g} - r_{i,i,g}| = O_{\mathrm{p}}[\{\log p/(nq)\}^{\frac{1}{2}}],$$

and

$$\max_{1\le i\le p}|\hat{r}_{i,i,g} - \tilde{R}_{i,i,g}| = o_{\mathrm{p}}\{(nq\log p)^{-1/2}\}.$$

Note that

$$\max_{1\le i<j\le p}(\hat{\beta}_{i,j,g}^2\hat{r}_{i,i,g}/\hat{r}_{j,j,g} - \eta_{i,j,g}^2) = o_{\mathrm{p}}(1/\log p),$$

and

$$\max_{1\le i\le j\le p}|\omega_{S_g,i,i}\hat{\sigma}_{i,i,\epsilon}^{(g)} + \omega_{S_g,j,j}\hat{\sigma}_{j,j,\epsilon}^{(g)} - 2| = O_{\mathrm{p}}\{(\log p/(nq))^{1/2}\}.$$

As noted in Section 4.2, in the two-sample case, $\rho_{S_1,i,j}$ and $\rho_{S_2,i,j}$ are not necessarily equal to 0 under the null, and additional corrections are crucial. Toward that end, we divide the set of indices into two subsets, $\mathcal{H}_0\setminus A_\tau$ and A_τ, the former with a negligible correction and the latter requiring a major correction by $b_{i,j}$. Also note that for $(i,j)\in\mathcal{H}_0\setminus A_\tau$, we have $|\omega_{S_g,i,j}| = o\{(\log p)^{-1}\}$. Then by Lemma 1, it is straightforward to see that, under Conditions (C1) and (C2), we have, for $(i,j)\in\mathcal{H}_0\setminus A_\tau$,

$$\max_{(i,j)\in\mathcal{H}_0\setminus A_\tau}||W_{i,j}| - |V_{i,j}|| = o_{\mathrm{p}}\{(\log p)^{-1/2}\}.$$

For $(i,j)\in A_\tau$, as a result of Lemma 1, we have

$$W_{i,j} = V_{i,j} + b_{i,j} + o_{\mathrm{p}}(\log p^{-1/2}),$$

where $b_{i,j} = 2\{\omega_{i,j}(\hat{\sigma}_{i,i,\epsilon}^{(1)} - \hat{\sigma}_{i,i,\epsilon}^{(2)}) + \omega_{i,j}(\hat{\sigma}_{j,j,\epsilon}^{(1)} - \hat{\sigma}_{j,j,\epsilon}^{(2)})\}/(\hat{\theta}_{i,j,1}+\hat{\theta}_{i,j,2})^{1/2}$. Note that

$$\begin{aligned}|b_{i,j}| \;\le\; & 2\left(\frac{2\eta_{i,j}^2}{1+\eta_{i,j}^2}\right)^{\frac{1}{2}}\left[\frac{|\tilde{\sigma}_{i,i,\epsilon}^{(1)} - \tilde{\sigma}_{i,i,\epsilon}^{(2)}|}{\{\mathrm{var}((\epsilon_{k,i,l}^{(1)})^2)/(n_1q) + \mathrm{var}((\epsilon_{k,i,l}^{(2)})^2)/(n_2q)\}^{\frac{1}{2}}}\right.\\ &\left.+\frac{|\tilde{\sigma}_{j,j,\epsilon}^{(1)} - \tilde{\sigma}_{j,j,\epsilon}^{(2)}|}{\{\mathrm{var}((\epsilon_{k,j,l}^{(1)})^2)/(n_1q) + \mathrm{var}((\epsilon_{k,j,l}^{(2)})^2)/(n_2q)\}^{\frac{1}{2}}}\right] + o\{(\log p)^{-1/2}\},\end{aligned}$$

where $\tilde{\sigma}_{i,i,\epsilon}^{(s)} = n_s^{-1}\sum_{k=1}^{n_s}\sum_{l=1}^{q}(\epsilon_{k,i,l}^{(s)})^2$. Thus, we have

$$\begin{aligned}&\mathrm{pr}(\max_{(i,j)\in A_\tau} W_{i,j}^2 \ge 4\log p - \log\log p + t)\\ &\quad\le \mathrm{Card}(A_\tau)\{\mathrm{pr}(V_{i,j}^2 \ge \log p/8) + \mathrm{pr}(b_{i,j}^2 \ge 2\log p)\} = o(1), \qquad \text{(S2.1)}\end{aligned}$$

where the last equality is a direct result of Lemma 2.

Under the conditions of Theorem 1, we have

$$\sum_{1\le i<j\le p} I\{|W_{i,j}| \ge 2(\log p)^{1/2}\} \ge [1/\{(8\pi)^{1/2}\alpha\} + \delta](\log_2 p)^{1/2},$$

with probability going to one. Henceforth, with probability going to one, we have

$$\frac{(p^2-p)/2}{\max\{\sum_{1\le i<j\le p} I\{|W_{i,j}| \ge 2(\log p)^{1/2}\},1\}} \le \frac{p^2-p}{2}\left\{\frac{1}{(8\pi)^{1/2}\alpha}+\delta\right\}^{-1}(\log_2 p)^{-1/2}.$$

Let $t_p = (4\log p - \log_2 p - \log_3 p)^{1/2}$. Because $1-\Phi(t_p) \sim 1/\{(2\pi)^{1/2}t_p\}\exp(-t_p^2/2)$, we have $\mathrm{pr}(1 \le \hat{t} \le t_p) \to 1$ according to the definition of $\hat{t}$ in (9). Note that, for $0 \le \hat{t} \le t_p$,

$$\frac{2\{1-\Phi(\hat{t})\}(p^2-p)/2}{\max\{\sum_{1\le i<j\le p} I\{|W_{i,j}| \ge 2(\log p)^{1/2}\},1\}} = \alpha.$$

Thus to prove Theorem 1, it suffices to show that

$$\frac{|\sum_{(i,j)\in\mathcal{H}_0}\{I(|W_{i,j}| \ge t) - G(t)\}|}{\{l_0 G(t)\}} \to 0$$

in probability, for $0 \le t \le \{4\log p + o(\log p)\}^{1/2}$, where $G(t) = 2\{1-\Phi(t)\}$.

If $t = \{4\log p + o(\log p)\}^{1/2}$, by (S2.1) and Condition (C2), it suffices to show

$$\frac{|\sum_{(i,j)\in\mathcal{H}_0\backslash A_\tau}\{I(|V_{i,j}| \ge t) - G(t)\}|}{\{l_0 G(t)\}} \to 0 \tag{S2.2}$$

If $t \le (C\log p)^{1/2}$ with $C < 4$, we have

$$\left|\frac{\sum_{(i,j)\in A_\tau\cap\mathcal{H}_0}\{I(|W_{i,j}| \ge t) - I(|V_{i,j}| \ge t)\}}{l_0 G(t)}\right| \le \frac{2|A_\tau\cap\mathcal{H}_0|}{O(p^{2-C/2})} \to 0$$

in probability. Thus, it is again enough to show that (S2.2) holds. This shows the negligibility of the highly dependent pairs. We arrange the indices $\{(i,j) : (i,j) \in \mathcal{H}_0\backslash A_\tau\}$ in any ordering and set them as $\{(i_m, j_m) : m = 1,\ldots,s\}$ with $s =$Card$(\mathcal{H}_0 \setminus A_\tau)$. Let $n_1/n_2 \le K$ with $K \ge 1$, $\theta_{m,g} = \mathrm{var}(\epsilon^{(g)}_{k,i_m,l}\epsilon^{(g)}_{k,j_m,l})$, and define $Z_{k,m,l} = (n_1/n_2)\{\epsilon^{(2)}_{k,i_m,l}\epsilon^{(2)}_{k,j_m,l} - E(\epsilon^{(2)}_{k,i_m,l}\epsilon^{(2)}_{k,j_m,l})\}$ for $1 \le k \le n_2$ and $1 \le l \le q$, and $Z_{k,m,l} = -\{\epsilon^{(1)}_{k,i_m,l}\epsilon^{(1)}_{k,j_m,l} - E(\epsilon^{(1)}_{k,i_m,l}\epsilon^{(1)}_{k,j_m,l})\}$ for $n_2+1 \le k \le n_1+n_2$ and $1 \le l \le q$. Define

$$V_m = (n_1^2 q\theta_{m,2}/n_2 + n_1 q\theta_{m,1})^{-1/2}\sum_{k=1}^{n_1+n_2}\sum_{l=1}^{q} Z_{k,m,l},$$

and

$$\hat{V}_m = (n_1^2 q\theta_{m,2}/n_2 + n_1 q\theta_{m,1})^{-1/2} \sum_{k=1}^{n_1+n_2} \sum_{l=1}^{q} \hat{Z}_{k,m,l},$$

where $\hat{Z}_{k,m,l} = Z_{k,m,l} I(|Z_{k,m,l}| \le \tau_n) - E\{Z_{k,m,l} I(|Z_{k,m,l}| \le \tau_n)\}$, and $\tau_n = 32\log(p+nq)$. Note that $\max_{(i,j)\in\mathcal{H}_0\setminus A_\tau} V_{i,j}^2 = \max_{1\le m\le s} V_m^2$, and that

$$\begin{aligned}
\max_{1\le m\le s} (nq)^{-1/2} \sum_{k=1}^{n_1+n_2} \sum_{l=1}^{q} \quad & E[|Z_{k,m,l}| I\{|Z_{k,m,l}| \ge 32\log(p+nq)\}] \\
& \le C(nq)^{1/2} \max_{1\le k\le n} \max_{1\le l\le q} \max_{1\le m\le s} E[|Z_{k,m,l}| I\{|Z_{k,m,l}| \ge 32\log(p+nq)\}] \\
& \le C(nq)^{1/2} (p+nq)^{-4} \max_{1\le k\le n} \max_{1\le l\le q} \max_{1\le m\le s} E[|Z_{k,m,l}| \exp\{|Z_{k,m,l}|/8\}] \\
& \le C(nq)^{1/2}(p+nq)^{-4}.
\end{aligned}$$

This yields to

$$\mathrm{pr}\Big\{ \max_{1\le m\le s} |V_m - \hat{V}_m| \ge (\log p)^{-1} \Big\} \le \mathrm{pr}\Big(\max_{1\le m\le s} \max_{1\le k\le n} \max_{1\le l\le q} |Z_{k,m,l}| \ge \tau_n \Big) = O(p^{-1}).$$

By Lemma 6.1 in Liu (2013), we have

$$\max_m \sum_{0\le t\le t_p} \Big| \frac{\mathrm{pr}(|V_m| \ge t)}{G(t)} - 1 \Big| \le C(\log p)^{-1-\tau_1},$$

with $\tau_1 = \min\{\tau, 1/2\}$ and $t_p = (4\log p - \log_2 p - \log_3 p)^{1/2}$. Thus to prove Theorem 1, by the fact that $G(t + o((\log p)^{-1/2}))/G(t) = 1 + o(1)$ uniformly in $0 \le t \le 2\sqrt{\log p}$, it suffices to prove that

$$\frac{|\sum_{1\le m\le s}\{I(|\hat{V}_m| \ge t) - G(t)\}|}{l_0 G(t)} \to 0$$

in probability, for $0 \le t \le t_p$, where $G(t) = 2\{1 - \Phi(t)\}$. Let $0 \le t_0 < t_1 < \cdots < t_b = t_p$ such that $t_\iota - t_{\iota-1} = v_p$ for $1 \le \iota \le b-1$ and $t_b - t_{b-1} \le v_p$, where $v_p = 1/\{\log p(\log_4 p)^2\}^{1/2}$. Thus we have $b \sim t_p/v_p$. For any t such that $t_{\iota-1} \le t \le t_\iota$, we have

$$\begin{aligned}
\frac{\sum_{1\le m\le s} I(|\hat{V}_m| \ge t_\iota)}{l_0 G(t_\iota)} \frac{G(t_\iota)}{G(t_{\iota-1})} &\le \frac{\sum_{1\le m\le s} I(|\hat{V}_m| \ge t)}{l_0 G(t)} \\
&\le \frac{\sum_{1\le m\le s} I(|\hat{V}_m| \ge t_{\iota-1})}{l_0 G(t_{\iota-1})} \frac{G(t_{\iota-1})}{G(t_\iota)}.
\end{aligned}$$

Thus it suffices to prove

$$\max_{0\le\iota\le b} \Big| \frac{\sum_{1\le m\le s}[I(|\hat{V}_m| \ge t_\iota) - G(t_\iota)]}{l_0 G(t_\iota)} \Big| \to 0 \tag{S2.3}$$

in probability. Note that

$$\begin{aligned}
&\text{pr}\left\{\max_{1\leq\iota\leq b}\left|\frac{\sum_{1\leq m\leq s}[I(|\hat{V}_m|\geq t_\iota)-G(t_\iota)]}{l_0G(t_\iota)}\right|\geq\epsilon\right\}\\
&\leq\sum_{l=1}^{m}\text{pr}\left\{\left|\frac{\sum_{1\leq m\leq s}[I(|\hat{V}_m|\geq t_\iota)-G(t_\iota)]}{l_0G(t_\iota)}\right|\geq\epsilon\right\}\\
&\leq\frac{1}{v_p}\int_0^{t_p}\text{pr}\left\{\left|\frac{\sum_{1\leq m\leq s}[I(|\hat{V}_m|\geq t)-G(t)]}{l_0G(t)}\right|\geq\epsilon\right\}dt\\
&\quad+\sum_{\iota=b-1}^{b}\text{pr}\left\{\left|\frac{\sum_{1\leq m\leq s}[I(|\hat{V}_m|\geq t_\iota)-G(t_\iota)]}{l_0G(t_\iota)}\right|\geq\epsilon\right\}.
\end{aligned}$$

As noted in Section 4.2, in the two-sample setting, the test statistics can be highly dependent since $\boldsymbol{R}_{S_g}$ is not necessarily an identity matrix. To show the error rate control, we reorganize the set of test statistics into a number of subgroups according to the level of dependency, as shown in the proof of Theorem 4 in Xia et al. (2015), with $2nq$ regression models. As a result, for any $\epsilon>0$ that,

$$\begin{aligned}
\sum_{0\leq t\leq t_p}\text{pr}\left[\left|\frac{\sum_{1\leq m\leq s}\{I(|\hat{V}_m|\geq t)-\text{pr}(|\hat{V}_m|\geq t)\}}{2l_0\{1-\Phi(t)\}}\right|\geq\epsilon\right] &= o(1),\\
\int_0^{t_p}\text{pr}\left[\left|\frac{\sum_{1\leq m\leq s}\{I(|\hat{V}_m|\geq t)-\text{pr}(|\hat{V}_m|\geq t)\}}{2l_0\{1-\Phi(t)\}}\right|\geq\epsilon\right]dt &= o(v_p).
\end{aligned}$$

Thus (S2.3) is proved. Then Theorem 1 follows. □

S2.2 Proof of Theorem 2

By the proof of Theorem 1, together with Lemma 3, Theorem 2 is proved. □

References

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41:2948–2978.

Xia, Y., Cai, T., and Cai, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika*, 102:247–266.

Xia, Y. and Li, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics*, in press.