A Unified Debiasing Framework for Inference with Jointly Regularized Multi-Graph Estimators

Qiao Zheng[1,†], Gefei Li[2,†], Yibo Yan[1] and Riquan Zhang[1]

[1] *Shanghai University of International Business and Economics*

[2] *East China Normal University*

Supplementary Material

The online Supplementary Material includes algorithms, technical proofs, and additional simulation studies. Specifically, it contains Algorithm 1 (debiasing procedure), Algorithm 2 (FWER control), and Algorithm 3 (FDR control); detailed proofs of Proposition 2.1, Propositions 3.1 and 3.2, Theorems 3.1 – 3.4 and 4.1, and Corollaries 3.1 and 4.1; additional simulation results for $p = 100$ under Models II–III and for $p = 50, 200, 400$ under Models I – III; and a real-data example with $p = 100$.

S1 Related Algorithm

Algorithm 1 Debiasing Procedure for Multi-Graph Estimators

1: **Input:** data $\{\mathbf{x}^{(m)}\}_{m=1}^{M}$; an initial estimation routine that produces $\widehat{\mathbf{\Theta}}$ from $\{\mathbf{x}^{(m)}\}_{m=1}^{M}$ and a covariance-type estimator $\widehat{S}^{(m)}$ (default $\widehat{S}^{(m)} = \widehat{\Sigma}^{(m)}$).
2: **for** $m = 1, \ldots, M$ **do**
3: Construct the debiased estimator: $\widetilde{\Theta}^{(m)} \leftarrow 2\widehat{\Theta}^{(m)} - \widehat{\Theta}^{(m)}\widehat{S}^{(m)}\widehat{\Theta}^{(m)}$.
4: Symmetrize $\widetilde{\Theta}^{(m)} \leftarrow \{\widetilde{\Theta}^{(m)} + (\widetilde{\Theta}^{(m)})^{\top}\}/2$.
5: **end for**
6: **Output:** $\widetilde{\mathbf{\Theta}} = (\widetilde{\Theta}^{(1)}, \ldots, \widetilde{\Theta}^{(M)})$.

Corresponding author: Riquan Zhang. E-mail: zhangriquan@163.com
† These authors contributed equally to this work.

Algorithm 2 FWER Control via Gaussian Multiplier Bootstrap

Input: Debiased estimators $\{\widetilde{\Theta}^{(m)}\}_{m=1}^{M}$; standard errors $\{\widehat{\tau}_{m,jl}\}_{(m,j,l)\in\mathcal{S}}$; estimated linear-representation terms $\{\widehat{\psi}_{i,jl}^{(m)}\}_{i=1}^{n}$ for all $(m,j,l)\in\mathcal{S}$; index set $\mathcal{S}$ with $d_s=|\mathcal{S}|$; sample size n; nominal level $\alpha\in(0,1)$; bootstrap replicates B.
Step 1 (Studentized statistics).
for each $(m,j,l)\in\mathcal{S}$ **do**
  $T_{m,jl}\leftarrow\sqrt{n}\,\widetilde{\Theta}_{jl}^{(m)}/\widehat{\tau}_{m,jl}$ $(H_0:\Theta_{jl}^{(m)}=0)$.
end for
$T_{\max}\leftarrow\max_{(m,j,l)\in\mathcal{S}}\left|T_{m,jl}\right|$.
Step 2 (Gaussian multiplier bootstrap for $c_{1-\alpha}$).
for $b=1$ **to** B **do**
  for $m=1$ **to** M **do**
    Generate multipliers $\xi_1^{(b,m)},\ldots,\xi_n^{(b,m)}\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$.
  end for
  for each $(m,j,l)\in\mathcal{S}$ **do**
    $T_{m,jl}^{*(b)}\leftarrow\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i^{(b,m)}\,\widehat{\psi}_{i,jl}^{(m)}\right)\Big/\widehat{\tau}_{m,jl}$.
  end for
  $T_{\max}^{*(b)}\leftarrow\max_{(m,j,l)\in\mathcal{S}}\left|T_{m,jl}^{*(b)}\right|$.
end for
$c_{1-\alpha}\leftarrow$ empirical $(1-\alpha)$-quantile of $\{T_{\max}^{*(b)}\}_{b=1}^{B}$.
Step 3 (Simultaneous confidence intervals / rejection rule).
for each $(m,j,l)\in\mathcal{S}$ **do**
  $\mathrm{CI}_{m,jl}^{\mathrm{sim}}(1-\alpha)\leftarrow\left[\widetilde{\Theta}_{jl}^{(m)}\pm c_{1-\alpha}\,\widehat{\tau}_{m,jl}/\sqrt{n}\right]$.
end for
Reject $H_0:\Theta_{jl}^{(m)}=0$ if $|T_{m,jl}|>c_{1-\alpha}$. (Two-sided test)
Output: critical value $c_{1-\alpha}$ and simultaneous CIs $\{\mathrm{CI}_{m,jl}^{\mathrm{sim}}(1-\alpha)\}_{(m,j,l)\in\mathcal{S}}$.

Algorithm 3 FDR Control via the Benjamini–Yekutieli (BY) Procedure

Input: Debiased estimators $\{\widetilde{\Theta}^{(m)}\}_{m=1}^{M}$; standard errors $\{\widehat{\tau}_{m,jl}\}_{(m,j,l)\in\mathcal{S}}$; index set $\mathcal{S}\subset\{(m,j,l):1\le m\le M,\ 1\le j<l\le p\}$ with $d_s=|\mathcal{S}|$; sample size n; target FDR level $q\in(0,1)$.
Step 1 (Compute studentized statistics and p-values).
for each $(m,j,l)\in\mathcal{S}$ **do**
  $T_{m,jl}\leftarrow\sqrt{n}\,\widetilde{\Theta}_{jl}^{(m)}/\widehat{\tau}_{m,jl}$ $(H_0:\Theta_{jl}^{(m)}=0)$.
  $\widehat{p}_{m,jl}\leftarrow 2\{1-\Phi(|T_{m,jl}|)\}$.
end for
Step 2 (Order p-values).
Sort $\{\widehat{p}_{m,jl}:(m,j,l)\in\mathcal{S}\}$ increasingly to obtain $\widehat{p}_{(1)}\le\cdots\le\widehat{p}_{(d_s)}$, and record the corresponding indices $\{(m_r,j_r,l_r)\}_{r=1}^{d_s}$.
Step 3 (Compute BY threshold).
$c_{d_s}\leftarrow\sum_{r=1}^{d_s}\frac{1}{r}$.
$\widehat{r}_{\mathrm{BY}}\leftarrow\max\left\{1\le r\le d_s:\ \widehat{p}_{(r)}\le\frac{r}{d_s}\cdot\frac{q}{c_{d_s}}\right\}$;
if the set is empty, set $\widehat{r}_{\mathrm{BY}}=0$.
Step 4 (Reject/Select edges).
Reject H_0 for all indices $\{(m_r,j_r,l_r):1\le r\le\widehat{r}_{\mathrm{BY}}\}$.
Output: rejection set $\mathcal{R}_{\mathrm{BY}}=\{(m_r,j_r,l_r):1\le r\le\widehat{r}_{\mathrm{BY}}\}$ (with asymptotic FDR control at level q).

S2 Proof

Proof of Proposition 2.1. Fix m and suppress the superscript (m). We know that

$$\begin{aligned}\widetilde{\Theta} - \Theta_0 &= 2\widehat{\Theta} - \widehat{\Theta}\widehat{S}\widehat{\Theta} - \Theta_0 = \Delta + (\Delta + \Theta_0) - (\Delta + \Theta_0)(V + \Sigma_0)(\Delta + \Theta_0) \\ &= \Delta - (\Delta + \Theta_0)(V\Delta + V\Theta_0 + \Sigma_0\Delta) = -\Theta_0 V\Theta_0 - (\Delta\Sigma_0\Delta + \Delta V\Theta_0 + \Theta_0 V\Delta + \Delta V\Delta),\end{aligned}$$

where $\Delta = \widehat{\Theta} - \Theta_0$, and $V = \widehat{S} - \Sigma_0$. Thus, we complete the proof. □

Proof of Proposition 3.1. Recall $\widetilde{\boldsymbol{\Theta}} = (\widetilde{\Theta}^{(1)}, \ldots, \widetilde{\Theta}^{(M)})$ with $\widetilde{\Theta}^{(m)} = 2\widehat{\Theta}^{(m)} - \widehat{\Theta}^{(m)}\widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}$ for $m = 1, \ldots, M$. We focus on the m^{th} component and consider the error $\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}$ relative to the true precision matrix $\Theta_*^{(m)}$. Similar to the proof of Proposition 2.1, a direct algebraic manipulation yields

$$\begin{aligned}\widetilde{\Theta}^{(m)} - \Theta_*^{(m)} &= 2\widehat{\Theta}^{(m)} - \widehat{\Theta}^{(m)}\widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)} - \Theta_*^{(m)} \\ &= -\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)} + \big(\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\big)\big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big) - \big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big) W^{(m)} \Theta_*^{(m)},\end{aligned}$$

where $W^{(m)} = \widehat{\Sigma}^{(m)} - \Sigma^{(m)}$ and $\mathbf{I}$ denotes the $p \times p$ identity matrix.

By the triangle inequality and the basic matrix norm bounds $\|AB\|_{\max} \le \|A\|_{\max}\|B\|_1$ and $\|ABC\|_{\max} \le \|A\|_1\|B\|_{\max}\|C\|_1$, we obtain

$$\begin{aligned}\big\|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\big\|_{\max} &\le \big\|\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)}\big\|_{\max} + \big\|\big(\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\big)\big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big)\big\|_{\max} \\ &\qquad + \big\|\big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big) W^{(m)} \Theta_*^{(m)}\big\|_{\max} \\ &\lesssim \|W^{(m)}\|_{\max} + \big\|\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\big\|_{\max}\big\|\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big\|_1 + \|W^{(m)}\|_{\max}\big\|\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big\|_1,\end{aligned}$$

where we used Assumption 3.2 to absorb $\|\Theta_*^{(m)}\|_1$ into the implicit constant.

Taking the maximum over $m = 1, \ldots, M$ yields

$$\max_{1\le m\le M} \left\|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\right\|_{\max} \lesssim \max_{1\le m\le M} \|W^{(m)}\|_{\max} + \delta_{n,1} \max_{1\le m\le M} \left\|\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\right\|_{\max} + \delta_{n,1} \max_{1\le m\le M} \|W^{(m)}\|_{\max}, \tag{S1}$$

where $\max_{1\le m\le M} \|\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\|_1 = O_p(\delta_{n,1})$ by Assumption 3.2.

By standard concentration for Gaussian (or sub-Gaussian) samples and a union bound over m, we have

$$\max_{1\le m\le M} \|W^{(m)}\|_{\max} = \max_{1\le m\le M} \|\widehat{\Sigma}^{(m)} - \Sigma^{(m)}\|_{\max} = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Moreover, via $\|AB\|_{\max} \le \|A\|_{\max}\|B\|_1$, we obtain that

$$\max_{1\le m\le M} \left\|\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\right\|_{\max} = O_p\left(\delta_{n,1}\sqrt{\frac{\log p}{n}} + \sqrt{\frac{\log p}{n}} + \delta_{n,1}\right).$$

Substituting these bounds into Eq. (S1) gives

$$\begin{aligned}\max_{1\le m\le M} \left\|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\right\|_{\max} &= O_p\left(\sqrt{\frac{\log p}{n}} + \delta_{n,1}\left(\delta_{n,1}\sqrt{\frac{\log p}{n}} + \sqrt{\frac{\log p}{n}} + \delta_{n,1}\right) + \delta_{n,1}\sqrt{\frac{\log p}{n}}\right) \\ &= O_p\left(\sqrt{\frac{\log p}{n}} + \delta_{n,1}\sqrt{\frac{\log p}{n}} + \delta_{n,1}^2\right),\end{aligned}$$

where in the last step we absorbed the term $\delta_{n,1}^2\sqrt{\frac{\log p}{n}}$ into $\delta_{n,1}^2$. This proves the stated bound and completes the proof. □

Proof of Theorem 3.1. We begin by establishing the asymptotic linear representation

stated in part (i). Under the proof of Proposition 3.1, we know that

$$\widetilde{\Theta}^{(m)} - \Theta_*^{(m)} = -\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)} + R^{(m)},$$

where the remainder term $R^{(m)}$ is defined by

$$R^{(m)} = \big(\mathbf{I} - \widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}\big)\big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big) - \big(\widehat{\Theta}^{(m)} - \Theta_*^{(m)}\big) W^{(m)} \Theta_*^{(m)}.$$

According to the discussion in the proof of Proposition 3.1, we obtain uniformly in m that

$$\|R^{(m)}\|_{\max} = O_p\Big(\delta_{n,1}\sqrt{\tfrac{\log p}{n}} + \delta_{n,1}^2\Big).$$

Multiplying both sides by $\sqrt{n}$ yields

$$\sqrt{n}\|R^{(m)}\|_{\max} = O_p\Big(\delta_{n,1}\sqrt{\log p} + \delta_{n,1}^2\sqrt{n}\Big).$$

By Assumption 3.4(b), we obtain that

$$\delta_{n,1} = o\bigg(\frac{1}{\sqrt{\log p}}\bigg) \quad \text{and} \quad \delta_{n,1} = o\big(n^{-1/4}\big)\,,$$

we have

$$\delta_{n,1}\sqrt{\log p} \to 0, \quad \delta_{n,1}^2\sqrt{n} \to 0.$$

Hence,

$$\sqrt{n}\,\|R^{(m)}\|_{\max} = o_p(1), \quad m = 1, \ldots, M,$$

which proves the asymptotic linear representation in part (i).

Then we establish the entry-wise asymptotic normality in part (ii). From the asymptotic linear representation in part (i), we have

$$\widetilde{\Theta}^{(m)} - \Theta_*^{(m)} = -\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)} + R^{(m)}, \quad \sqrt{n}\|R^{(m)}\|_{\max} = o_p(1), \tag{S2}$$

where $W^{(m)} = \widehat{\Sigma}^{(m)} - \Sigma^{(m)}$ and $\widehat{\Sigma}^{(m)} = n^{-1}\sum_{k=1}^n \boldsymbol{x}_k^{(m)} (\boldsymbol{x}_k^{(m)})^\top$.

Define

$$\boldsymbol{v}_k^{(m)} := \Theta_*^{(m)} \boldsymbol{x}_k^{(m)}, \quad k = 1, \ldots, n.$$

Then $\mathbb{E}(\boldsymbol{v}_k^{(m)} (\boldsymbol{v}_k^{(m)})^\top) = \Theta_*^{(m)} \Sigma^{(m)} \Theta_*^{(m)} = \Theta_*^{(m)}$, and hence,

$$\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)} = \frac{1}{n}\sum_{k=1}^n \Big(\boldsymbol{v}_k^{(m)} (\boldsymbol{v}_k^{(m)})^\top - \mathbb{E}\big\{ \boldsymbol{v}_k^{(m)} (\boldsymbol{v}_k^{(m)})^\top \big\} \Big). \tag{S3}$$

Taking the $(j,l)^{\text{th}}$ entry in Eq. (S2) and using Eq. (S3), we obtain that

$$\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)} = -\frac{1}{n}\sum_{k=1}^n \Big\{ v_{k,j}^{(m)} v_{k,l}^{(m)} - \mathbb{E}\big(v_{k,j}^{(m)} v_{k,l}^{(m)} \big) \Big\} + R_{jl}^{(m)},$$

where $v_{k,j}^{(m)}$ denotes the j^{th} component of $\boldsymbol{v}_k^{(m)}$. Multiplying both sides by $\sqrt{n}$ yields

$$\sqrt{n}\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)} \big) = -\frac{1}{\sqrt{n}}\sum_{k=1}^n \xi_{k,jl}^{(m)} + \sqrt{n} R_{jl}^{(m)},$$

where we define $\xi_{k,jl}^{(m)} := v_{k,j}^{(m)} v_{k,l}^{(m)} - \mathbb{E}\big(v_{k,j}^{(m)} v_{k,l}^{(m)} \big)$.

By construction, $\{\xi_{k,jl}^{(m)}\}_{k=1}^{n}$ are i.i.d. mean-zero random variables with variance

$$\tau_{m,jl}^2 = \mathrm{Var}\big(v_{k,j}^{(m)} v_{k,l}^{(m)}\big) = \mathrm{Var}\left\{\big(\Theta_*^{(m)}\boldsymbol{x}^{(m)}\big)_j \big(\Theta_*^{(m)}\boldsymbol{x}^{(m)}\big)_l\right\},$$

Hence, by the classical Lindeberg–Feller (or Lyapunov) central limit theorem, we obtain

$$-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\xi_{k,jl}^{(m)} \xrightarrow{d} \mathcal{N}(0,\tau_{m,jl}^2).$$

On the other hand, by part (i), we have $\sqrt{n}\,\|R^{(m)}\|_{\max} = o_p(1)$, and in particular $\sqrt{n}\,R_{jl}^{(m)} = o_p(1)$. Therefore, from the Slutsky's lemma, it follows that

$$\sqrt{n}\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\big) \xrightarrow{d} \mathcal{N}(0,\tau_{m,jl}^2).$$

It remains to identify the explicit form of $\tau_{m,jl}^2$ under the Gaussian model. Since $\boldsymbol{v}_k^{(m)} \sim \mathcal{N}(\mathbf{0},\Theta_*^{(m)})$, via Isserlis' formula for Gaussian fourth moments, we have

$$\mathbb{E}\big\{(v_{k,j}^{(m)})^2(v_{k,l}^{(m)})^2\big\} = [\Theta_*^{(m)}]_{jj}[\Theta_*^{(m)}]_{ll} + 2\big([\Theta_*^{(m)}]_{jl}\big)^2, \quad \mathbb{E}\big(v_{k,j}^{(m)}v_{k,l}^{(m)}\big) = [\Theta_*^{(m)}]_{jl}.$$

Thus, we have

$$\begin{aligned}
\tau_{m,jl}^2 &= \mathbb{E}\big\{(v_{k,j}^{(m)})^2(v_{k,l}^{(m)})^2\big\} - \Big(\mathbb{E}\big(v_{k,j}^{(m)}v_{k,l}^{(m)}\big)\Big)^2 \\
&= [\Theta_*^{(m)}]_{jj}[\Theta_*^{(m)}]_{ll} + 2\big([\Theta_*^{(m)}]_{jl}\big)^2 - \big([\Theta_*^{(m)}]_{jl}\big)^2 \\
&= [\Theta_*^{(m)}]_{jj}[\Theta_*^{(m)}]_{ll} + \big([\Theta_*^{(m)}]_{jl}\big)^2,
\end{aligned}$$

which gives the stated variance expression and completes the proof of part (ii). □

Proof of Proposition 3.2. Under the conditions of Theorem 3.1, according to Proposition 3.1, we have

$$\max_{1\le m\le M} \left\|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\right\|_{\max} = o_p(1).$$

Therefore, for any fixed $m \in \{1, \ldots, M\}$ and indices (j, l),

$$\widetilde{\Theta}_{jj}^{(m)} \xrightarrow{p} \Theta_{*,jj}^{(m)}, \quad \widetilde{\Theta}_{ll}^{(m)} \xrightarrow{p} \Theta_{*,ll}^{(m)}, \quad \widetilde{\Theta}_{jl}^{(m)} \xrightarrow{p} \Theta_{*,jl}^{(m)}.$$

Thus, by the continuous mapping theorem, we have

$$\widehat{\tau}_{m,jl}^2 = f\left(\widetilde{\Theta}_{jj}^{(m)}, \widetilde{\Theta}_{ll}^{(m)}, \widetilde{\Theta}_{jl}^{(m)}\right) \xrightarrow{p} f\left(\Theta_{*,jj}^{(m)}, \Theta_{*,ll}^{(m)}, \Theta_{*,jl}^{(m)}\right) = \tau_{m,jl}^2,$$

where $f(a, b, c) = ab + c^2, (a, b, c) \in \mathbb{R}^3$. This proves the consistency of $\widehat{\tau}_{m,jl}^2$ and completes the proof. □

Proof of Theorem 3.2. Fix any $m \in \{1, \ldots, M\}$ and index pair (j, l). By Theorem 3.1, the debiased estimator achieves the entry-wise asymptotic normality

$$\sqrt{n}\left(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\right) \xrightarrow{d} \mathcal{N}\left(0, \tau_{m,jl}^2\right),$$

where $\tau_{m,jl}^2$ denotes the asymptotic variance defined therein.

Moreover, by Proposition 3.2, the plug-in estimator $\widehat{\tau}_{m,jl}^2$ satisfies

$$\widehat{\tau}_{m,jl}^2 \xrightarrow{p} \tau_{m,jl}^2.$$

Applying Slutsky's lemma yields

$$\sqrt{n}\frac{\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}}{\widehat{\tau}_{m,jl}} \xrightarrow{d} \mathcal{N}(0,1).$$

Consequently, for any $\alpha \in (0,1)$,

$$\Pr\left(\Theta_{*,jl}^{(m)} \in \mathrm{CI}_{m,jl}(1-\alpha)\right) = \Pr\left(\left|\sqrt{n}\frac{\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}}{\widehat{\tau}_{m,jl}}\right| \leq z_{1-\alpha/2}\right) \to \Pr\left(|Z| \leq z_{1-\alpha/2}\right) = 1-\alpha,$$

where $Z \sim \mathcal{N}(0,1)$. This establishes the asymptotic validity of the confidence interval $\mathrm{CI}_{m,jl}(1-\alpha)$ and completes the proof. □

Proof of Corollary 3.1. The result follows directly from Theorem 3.1, Proposition 3.2, and Slutsky's theorem. □

Proof of Theorem 3.3. For any $(m,j,l) \in \mathcal{S}$, recall that

$$T_{m,jl} = \sqrt{n}\frac{\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}}{\widehat{\tau}_{m,jl}}, \quad T_{\max} = \max_{(m,j,l)\in\mathcal{S}} |T_{m,jl}|.$$

By the definition of the simultaneous confidence intervals,

$$\Theta_{*,jl}^{(m)} \in \mathrm{CI}_{m,jl}^{\mathrm{sim}}(1-\alpha) \quad \Longleftrightarrow \quad \left|\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\right| \leq \frac{c_{1-\alpha}\widehat{\tau}_{m,jl}}{\sqrt{n}}.$$

Thus,

$$\begin{aligned}
&\Pr\left(\Theta_{*,jl}^{(m)} \in \mathrm{CI}_{m,jl}^{\mathrm{sim}}(1-\alpha) \text{ for all } (m,j,l) \in \mathcal{S}\right) \\
&\qquad = \Pr\left(\max_{(m,j,l)\in\mathcal{S}} \sqrt{n}\frac{|\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}|}{\widehat{\tau}_{m,jl}} \leq c_{1-\alpha}\right) = \Pr\left(T_{\max} \leq c_{1-\alpha}\right).
\end{aligned}$$

Therefore, it is equivalent to prove that

$$\Pr(T_{\max} \leq c_{1-\alpha}) \longrightarrow 1-\alpha.$$

By the asymptotic linear representation established in Theorem 3.1, for each $(m,j,l) \in \mathcal{S}$, we have

$$\sqrt{n}\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\big) = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\xi_{k,jl}^{(m)} + \sqrt{n}R_{jl}^{(m)},$$

where $\max\limits_{(m,j,l)\in\mathcal{S}}\sqrt{n}\left|R_{jl}^{(m)}\right| = o_p(1)$. Therefore, $T_{m,jl}$ can be rewritten as

$$T_{m,jl} = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_{k,jl}^{(m)}}{\widehat{\tau}_{m,jl}} + \frac{\sqrt{n}R_{jl}^{(m)}}{\widehat{\tau}_{m,jl}}.$$

Further, we rewrite $T_{m,jl}$ as

$$T_{m,jl} = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_{k,jl}^{(m)}}{\tau_{m,jl}} + \Delta_{n,mjl},$$

where

$$\Delta_{n,mjl} = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\xi_{k,jl}^{(m)}\left(\frac{1}{\widehat{\tau}_{m,jl}} - \frac{1}{\tau_{m,jl}}\right) + \frac{\sqrt{n}R_{jl}^{(m)}}{\widehat{\tau}_{m,jl}}.$$

Define

$$T_{\max}^{0} = \max_{(m,j,l)\in\mathcal{S}}\left|-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_{k,jl}^{(m)}}{\tau_{m,jl}}\right|,$$

then we have

$$|T_{\max} - T_{\max}^{0}| \leq T_{\max}^{0}\cdot\max_{(m,j,l)\in\mathcal{S}}\left|\frac{\tau_{m,jl}}{\widehat{\tau}_{m,jl}} - 1\right| + \max_{(m,j,l)\in\mathcal{S}}\frac{\sqrt{n}|R_{jl}^{(m)}|}{\widehat{\tau}_{m,jl}}.$$

Let $a_n := \max_{(m,j,l)\in\mathcal{S}} \left| \frac{\tau_{m,jl}}{\widehat{\tau}_{m,jl}} - 1 \right|$ and $b_n := \max_{(m,j,l)\in\mathcal{S}} \sqrt{n}\,|R_{jl}^{(m)}|$. Then, we have

$$|T_{\max} - T_{\max}^0| \le a_n T_{\max}^0 + b_n \max_{(m,j,l)\in\mathcal{S}} \frac{1}{\widehat{\tau}_{m,jl}}.$$

By Assumption 3.2 and Proposition 3.2, we have

$$\max_{(m,j,l)\in\mathcal{S}} \frac{1}{\widehat{\tau}_{m,jl}} = O_p(1).$$

Therefore, we obtain that

$$|T_{\max} - T_{\max}^0| = O_p\big(a_n T_{\max}^0 + b_n\big).$$

Under the Gaussian model, the standardized variables $\xi_{k,jl}^{(m)}/\tau_{m,jl}$ are mean zero with unit variance and uniformly bounded sub-exponential tails. Hence, by Bernstein's inequality, for some constant $c > 0$,

$$\Pr\left(\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\xi_{k,jl}^{(m)}}{\tau_{m,jl}} \right| > t \right) \le 2\exp\left\{ -c\min(t^2, \sqrt{n}\,t) \right\}.$$

Because $d_s = |\mathcal{S}| \le Mp^2$ and M is fixed, Assumption 3.3 implies $\log d_s = o(n)$. Applying the union bound over the d_s indices in $\mathcal{S}$ and taking $t = C\sqrt{\log d_s}$ therefore yields

$$T_{\max}^0 = O_p\big(\sqrt{\log d_s}\big).$$

Consequently,

$$|T_{\max} - T_{\max}^0| = O_p\left(a_n\sqrt{\log d_s} + b_n\right).$$

Next, we derive the convergence rates of a_n and b_n. By Theorem 3.1,

$$\max_{1\le m\le M}\left\|\widetilde{\Theta}^{(m)}-\Theta_*^{(m)}\right\|_{\max}=O_p\left(\sqrt{\frac{\log p}{n}}+\delta_{n,1}\sqrt{\frac{\log p}{n}}+\delta_{n,1}^2\right).$$

Since $\widehat{\tau}_{m,jl}^2=\widetilde{\Theta}_{jj}^{(m)}\widetilde{\Theta}_{ll}^{(m)}+\left(\widetilde{\Theta}_{jl}^{(m)}\right)^2$ and $\tau_{m,jl}$ is uniformly bounded away from zero under Assumption 3.2, it follows that

$$\max_{m,j,l}|\widehat{\tau}_{m,jl}-\tau_{m,jl}|=O_p\left(\sqrt{\frac{\log p}{n}}+\delta_{n,1}\sqrt{\frac{\log p}{n}}+\delta_{n,1}^2\right).$$

Consequently,

$$a_n=O_p\left(\sqrt{\frac{\log p}{n}}+\delta_{n,1}\sqrt{\frac{\log p}{n}}+\delta_{n,1}^2\right).$$

Moreover, by the proof of Theorem 3.1, we obtain that

$$b_n=O_p\left(\delta_{n,1}\sqrt{\log p}+\delta_{n,1}^2\sqrt{n}\right).$$

Combining the above rates with $|T_{\max}-T_{\max}^0|=O_p\left(a_n\sqrt{\log d_s}+b_n\right)$ and using $\log d_s=O(\log p)$ together with Assumption 3.3, we obtain that

$$|T_{\max}-T_{\max}^0|=O_p\left(\frac{\log p}{\sqrt{n}}+\delta_{n,1}\sqrt{\log p}+\delta_{n,1}^2\sqrt{n}\right).$$

Let $r_{n,0}:=\dfrac{\log p}{\sqrt{n}}+\delta_{n,1}\sqrt{\log p}+\delta_{n,1}^2\sqrt{n}$. Then the preceding result can be written as

$$|T_{\max}-T_{\max}^0|=O_p(r_{n,0}).$$

Under the rate conditions of Theorem 3.3, $r_{n,0}=o(1)$. Hence, there exists a deterministic

sequence $L_n \to \infty$, sufficiently slowly, such that $L_n r_{n,0} \to 0$ and

$$\Pr\left(|T_{\max} - T^0_{\max}| > L_n r_{n,0}\right) \to 0.$$

Let $\zeta_{1,n} := L_n r_{n,0}$. Then $\zeta_{1,n} \to 0$, and there exists a deterministic sequence $\zeta_{2,n} \to 0$ such that, for all sufficiently large n,

$$\Pr\left(|T_{\max} - T^0_{\max}| > \zeta_{1,n}\right) < \zeta_{2,n}.$$

Therefore, condition (14) of Chernozhukov et al. (2013) is satisfied.

Recall the Gaussian multiplier bootstrap statistic

$$T^*_{m,jl} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} e_k \frac{\xi^{(m)}_{k,jl}}{\widehat{\tau}_{m,jl}}, \quad (m,j,l) \in \mathcal{S},$$

and

$$T^*_{\max} = \max_{(m,j,l)\in\mathcal{S}} |T^*_{m,jl}|,$$

where $\{e_k\}_{k=1}^n$ are *i.i.d.* standard normal random variables independent of the data. Define the corresponding oracle multiplier bootstrap statistic by

$$T^{*,0}_{m,jl} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} e_k \frac{\xi^{(m)}_{k,jl}}{\tau_{m,jl}}, \quad (m,j,l) \in \mathcal{S},$$

and

$$T^{*,0}_{\max} = \max_{(m,j,l)\in\mathcal{S}} |T^{*,0}_{m,jl}|.$$

By the definition of a_n, we have

$$|T^*_{\max} - T^{*,0}_{\max}| \le \max_{(m,j,l)\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} e_k \xi^{(m)}_{k,jl} \left(\frac{1}{\widehat{\tau}_{m,jl}} - \frac{1}{\tau_{m,jl}} \right) \right| \le a_n T^{*,0}_{\max}.$$

Conditionally on the data, $\{T^{*,0}_{m,jl} : (m,j,l) \in \mathcal{S}\}$ is a centered Gaussian vector with conditional variances

$$\mathrm{Var}_e(T^{*,0}_{m,jl}) = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\xi^{(m)}_{k,jl}}{\tau_{m,jl}} \right)^2,$$

where $\mathrm{Var}_e(\cdot)$ denotes the variance with respect to the Gaussian multipliers $\{e_k\}_{k=1}^n$, conditional on the observed data. For $\varrho = (m,j,l) \in \mathcal{S}$, let $X_{k,\varrho} := \frac{\xi^{(m)}_{k,jl}}{\tau_{m,jl}}$, and define

$$\Delta_n := \max_{\varrho_1,\varrho_2\in\mathcal{S}} \left| \frac{1}{n} \sum_{k=1}^{n} X_{k,\varrho_1} X_{k,\varrho_2} - \mathbb{E}\big(X_{k,\varrho_1} X_{k,\varrho_2}\big) \right|.$$

Since $\mathbb{E}(X^2_{k,\varrho}) = 1$, we have

$$\mathrm{Var}_e(T^{*,0}_{m,jl}) = \frac{1}{n} \sum_{k=1}^{n} X^2_{k,\varrho} \le 1 + \Delta_n.$$

Therefore, we have

$$\max_{(m,j,l)\in\mathcal{S}} \mathrm{Var}_e(T^{*,0}_{m,jl}) \le 1 + \Delta_n.$$

Under the Gaussian model, the variables $X_{k,mjl}$ have uniformly bounded sub-exponential tails. Thus, condition (E.1) of Chernozhukov et al. (2013) is satisfied. Together with the dimensionality condition in Theorem 3.3, Lemma C.1 of Chernozhukov et al. (2013) yields $\Delta_n = o_p(1)$. Consequently,

$$\max_{(m,j,l)\in\mathcal{S}} \mathrm{Var}_e(T^{*,0}_{m,jl}) = O_p(1).$$

Hence, with probability tending to one, there exists a constant $C > 0$ such that

$$\Pr_e\left(T^{*,0}_{\max} > t\right) \leq 2d_s \exp\left(-\frac{t^2}{2C}\right),$$

where $\Pr_e(\cdot)$ denotes probability with respect to the Gaussian multipliers $\{e_k\}_{k=1}^n$, conditional on the observed data. Therefore, for any $\varepsilon > 0$, there exists a sufficiently large constant $C > 0$ such that

$$\Pr_e\left(T^{*,0}_{\max} > C\sqrt{\log d_s}\right) \leq \varepsilon$$

with probability tending to one. Because $|T^*_{\max} - T^{*,0}_{\max}| \leq a_n T^{*,0}_{\max}$, it follows that

$$\Pr_e\left(|T^*_{\max} - T^{*,0}_{\max}| > C a_n \sqrt{\log d_s}\right) \leq \varepsilon$$

with probability tending to one.

By the convergence rate of a_n established above and $\log d_s = O(\log p)$, we have

$$a_n\sqrt{\log d_s} = O_p(r_{n,0}).$$

Since $\zeta_{1,n} = L_n r_{n,0}$ and $L_n \to \infty$, it follows that

$$\frac{\zeta_{1,n}}{a_n\sqrt{\log d_s}} \longrightarrow \infty$$

in probability. Therefore,

$$\Pr_e\left(|T^*_{\max} - T^{*,0}_{\max}| > \zeta_{1,n}\right) \longrightarrow 0$$

in probability. By enlarging $\zeta_{2,n}$ if necessary while preserving $\zeta_{2,n} \to 0$, we may further ensure that, for all sufficiently large n,

$$\Pr\left\{\Pr_e\left(|T^*_{\max} - T^{*,0}_{\max}| > \zeta_{1,n}\right) > \zeta_{2,n}\right\} < \zeta_{2,n}.$$

Hence, condition (15) of Chernozhukov et al. (2013) is satisfied.

We next verify the remaining terms in Theorem 3.2 of Chernozhukov et al. (2013). Note that $T^0_{\max}$ can be represented as the maximum of a $2d_s$-dimensional normalized sum with coordinates

$$\pm\frac{\xi^{(m)}_{k,jl}}{\tau_{m,jl}}, \quad (m,j,l) \in \mathcal{S}.$$

These variables have mean zero and unit variance, so the variance condition in Theorem 3.2 of Chernozhukov et al. (2013) is satisfied. Moreover, they have uniformly bounded sub-exponential tails, which implies condition (E.1) of Chernozhukov et al. (2013) with $B_n = O(1)$.

Since $\log(2d_s n) = O(\log(pn))$, Theorem 2.2 and Lemma 2.2 of Chernozhukov et al. (2013), together with condition (E.1) and $B_n = O(1)$, imply that the Gaussian approximation error appearing in Theorem 3.2 of Chernozhukov et al. (2013) converges to zero under $\dfrac{\{\log(pn)\}^7}{n} \to 0$. Indeed, taking $\gamma = n^{-1}$ in Theorem 2.2 of Chernozhukov et al. (2013), the resulting bound is controlled by terms of orders

$$\left\{\frac{\{\log(pn)\}^7}{n}\right\}^{1/8} \quad \text{and} \quad \left\{\frac{\{\log(pn)\}^7}{n}\right\}^{1/2},$$

together with n^{-1}, all of which converge to zero under the above dimensionality condition. Moreover, Lemma C.1 of Chernozhukov et al. (2013) gives

$$\mathbb{E}(\Delta_n) = O\left(\sqrt{\frac{\log p}{n}} \vee \frac{\{\log(pn)\}^2 \log p}{n}\right).$$

Together with $\{\log(pn)\}^7/n \to 0$, this bound and Markov's inequality show that, by taking the auxiliary tolerance in Lemma 3.2 of Chernozhukov et al. (2013) to be $\{\log(pn)\}^{-3}$, the covariance deviation probability converges to zero. For the same choice, the quantile-comparison bound in Lemma 3.2 of Chernozhukov et al. (2013) is of order $\{\log(pn)\}^{-1/3}$ and hence also converges to zero. Therefore, both corresponding terms in Theorem 3.2 of Chernozhukov et al. (2013) vanish. Finally, we verify that the remaining terms involving $\zeta_{1,n}$ and $\zeta_{2,n}$ in Theorem 3.2 of Chernozhukov et al. (2013) converge to zero.

Note that $r_{n,0} \geq (\log p)/\sqrt{n}$, we have

$$1 \vee \log\left(\frac{2d_s}{r_{n,0}}\right) = O(\log(pn)).$$

It follows from the definition of $r_{n,0}$ that

$$r_{n,0}\sqrt{1 \vee \log\left(\frac{2d_s}{r_{n,0}}\right)} = O\left(\frac{\{\log(pn)\}^{3/2}}{\sqrt{n}} + \delta_{n,1}\log(pn) + \delta_{n,1}^2\sqrt{n\log(pn)}\right) = o(1).$$

Indeed, the first term converges to zero since $\{\log(pn)\}^7/n \to 0$, while

$$\delta_{n,1}\log(pn) = \left\{\delta_{n,1}^2\sqrt{n\log(pn)}\right\}^{1/2}\left\{\frac{\{\log(pn)\}^3}{n}\right\}^{1/4} = o(1),$$

and the last term converges to zero by $\delta_{n,1}^2\sqrt{n\log(pn)} \to 0$. Therefore, the sequence L_n

introduced above can be chosen sufficiently slowly such that

$$L_n r_{n,0} \sqrt{1 \vee \log\left(\frac{2d_s}{r_{n,0}}\right)} \to 0.$$

Such a choice also guarantees $L_n r_{n,0} \to 0$, as required above, because the square-root factor is at least one. Moreover, since $L_n \geq 1$ for all sufficiently large n, $\zeta_{1,n} = L_n r_{n,0} \geq r_{n,0}$, and hence

$$\log\left(\frac{2d_s}{\zeta_{1,n}}\right) \leq \log\left(\frac{2d_s}{r_{n,0}}\right).$$

Consequently,

$$\zeta_{1,n} \sqrt{1 \vee \log\left(\frac{2d_s}{\zeta_{1,n}}\right)} \leq L_n r_{n,0} \sqrt{1 \vee \log\left(\frac{2d_s}{r_{n,0}}\right)} \to 0.$$

Together with $\zeta_{2,n} \to 0$, this yields

$$\zeta_{1,n} \sqrt{1 \vee \log\left(\frac{2d_s}{\zeta_{1,n}}\right)} + \zeta_{2,n} \to 0.$$

Hence, all the error terms in Theorem 3.2 of Chernozhukov et al. (2013) converge to zero.

Let $c_{1-\alpha} := \inf\left\{t \in \mathbb{R} : \Pr_e(T^*_{\max} \leq t) \geq 1-\alpha\right\}$. Applying Theorem 3.2 of Chernozhukov et al. (2013) with dimension $2d_s$, we obtain

$$\sup_{\alpha \in (0,1)} \left|\Pr\left(T_{\max} \leq c_{1-\alpha}\right) - (1-\alpha)\right| \longrightarrow 0.$$

Therefore,

$$\Pr\left(\Theta^{(m)}_{*,jl} \in \mathrm{CI}^{\mathrm{sim}}_{m,jl}(1-\alpha) \text{ for all } (m,j,l) \in \mathcal{S}\right) = \Pr(T_{\max} \leq c_{1-\alpha}) \longrightarrow 1-\alpha.$$

Thus, we complete the proof.

□

Proof of Theorem 3.4. We first show that the Gaussian-based p-values $\{\widehat{p}_{m,jl}\}_{(m,j,l)\in\mathcal{H}_0}$ are asymptotically uniform under the null hypotheses, and then invoke the Benjamini–Yekutieli (BY) procedure to obtain asymptotic FDR control.

Fix any $(m,j,l) \in \mathcal{H}_0$. By Theorem 3.1 and the consistency of $\widehat{\tau}_{m,jl}$ established in Proposition 3.2, we have

$$T_{m,jl} = \sqrt{n}\frac{\widetilde{\Theta}_{jl}^{(m)}\Theta_{*,jl}^{(m)}}{\widehat{\tau}_{m,jl}} \xrightarrow{d} \mathcal{N}(0,1) \text{ under } H_0 : \Theta_{*,jl}^{(m)} = 0.$$

Consequently, for any fixed $t \in (0,1)$,

$$\begin{aligned}
\Pr\left(\widehat{p}_{m,jl} \le t\right) &= \Pr\Big(2\{1-\Phi(|T_{m,jl}|)\} \le t\Big) \\
&= \Pr\Big(|T_{m,jl}| \ge \Phi^{-1}(1-t/2)\Big) \\
&= t + o(1),
\end{aligned}$$

where the remainder term $o(1)$ is uniform over all $(m,j,l) \in \mathcal{H}_0$ under the conditions of Theorem 3.1. In particular,

$$\sup_{t\in(0,1)} \Big\{ \Pr\left(\widehat{p}_{m,jl} \le t\right) - t \Big\}_+ \longrightarrow 0, \quad (m,j,l) \in \mathcal{H}_0,$$

which implies that the null p-values are asymptotically uniform, or equivalently, stochastically dominated by the Uniform$(0,1)$ distribution.

Let $\widehat{p}_{(1)} \le \cdots \le \widehat{p}_{(d_s)}$ be the ordered p-values with associated index triples (m_r, j_r, l_r).

Recall the BY thresholds $t_r = \frac{r}{d_s}\frac{q}{c_{d_s}}$, $c_{d_s} = \sum_{r=1}^{d_s}\frac{1}{r}$, and the data-dependent cutoff $\widehat{r}_{\mathrm{BY}} = \max\left\{1 \le r \le d_s : \ \widehat{p}_{(r)} \le t_r\right\}$, with the convention that $\widehat{r}_{\mathrm{BY}} = 0$ if the set is empty. The BY procedure rejects all hypotheses with indices (m_r, j_r, l_r) for $1 \le r \le \widehat{r}_{\mathrm{BY}}$.

Combining this asymptotic uniformity with the Theorem 1.2 and 1.3 of Benjamini and Yekutielli (2001), which establishes FDR control for the step-up procedure with thresholds $t_r = \frac{r}{d_s}\frac{q}{c_{d_s}}$ under arbitrary dependence, we obtain

$$\limsup_{n\to\infty} \mathrm{FDR} \le q.$$

This completes the proof. □

Proof of Theorem 4.1. We start from the asymptotic linear representation in Theorem 3.1. For each $m = 1, \ldots, M$ and fixed (j, l), we have

$$\sqrt{n}\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\big) = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\xi_{k,jl}^{(m)} + o_p(1), \tag{S4}$$

where

$$\xi_{k,jl}^{(m)} = v_{k,j}^{(m)}v_{k,l}^{(m)} - \mathbb{E}\left\{v_{k,j}^{(m)}v_{k,l}^{(m)}\right\}, \quad \boldsymbol{v}_k^{(m)} = \Theta_*^{(m)}\boldsymbol{x}_k^{(m)},$$

and the remainder term $o_p(1)$ is uniform in m.

Let

$$\widetilde{\boldsymbol{\Theta}}_{jl} = \big(\widetilde{\Theta}_{jl}^{(1)}, \ldots, \widetilde{\Theta}_{jl}^{(M)}\big)^\top, \quad \boldsymbol{\Theta}_{*,jl} = \big(\Theta_{*,jl}^{(1)}, \ldots, \Theta_{*,jl}^{(M)}\big)^\top.$$

For any fixed contrast vector $\boldsymbol{c} \in \mathbb{R}^M$ with $\|\boldsymbol{c}\|_2 = 1$, the contrast-based error can be

written as

$$\boldsymbol{c}^\top\big(\widetilde{\boldsymbol{\Theta}}_{jl} - \boldsymbol{\Theta}_{*,jl}\big) = \sum_{m=1}^{M} c_m\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\big).$$

Multiplying Eq. (S4) by c_m and summing over $m = 1, \ldots, M$ gives

$$\begin{aligned}\sqrt{n}\Big\{\boldsymbol{c}^\top\widetilde{\boldsymbol{\Theta}}_{jl} - \boldsymbol{c}^\top\boldsymbol{\Theta}_{*,jl}\Big\} &= \sum_{m=1}^{M} c_m\sqrt{n}\big(\widetilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}\big) \\ &= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{m=1}^{M} c_m\,\xi_{k,jl}^{(m)} + o_p(1).\end{aligned}$$

Define $\zeta_{k,jl}(\boldsymbol{c}) = \sum_{m=1}^{M} c_m\,\xi_{k,jl}^{(m)}, k = 1, \ldots, n$. Then $\{\zeta_{k,jl}(\boldsymbol{c})\}_{k=1}^{n}$ are i.i.d. mean-zero random variables with variance

$$\mathrm{Var}\{\zeta_{k,jl}(\boldsymbol{c})\} = \boldsymbol{c}^\top\boldsymbol{\Sigma}_{jl}\boldsymbol{c},$$

where $\boldsymbol{\Sigma}_{jl} = \mathrm{Cov}\left(\boldsymbol{v}_{k,j}^{(1)}\boldsymbol{v}_{k,l}^{(1)}, \ldots, \boldsymbol{v}_{k,j}^{(M)}\boldsymbol{v}_{k,l}^{(M)}\right)$.

Via the Lindeberg-Feller (or Lyapunov) central limit theorem, we obtain that

$$-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\zeta_{k,jl}(\boldsymbol{c}) \xrightarrow{d} \mathcal{N}\big(0, \boldsymbol{c}^\top\boldsymbol{\Sigma}_{jl}\boldsymbol{c}\big).$$

Therefore,

$$\sqrt{n}\Big\{\boldsymbol{c}^\top\widetilde{\boldsymbol{\Theta}}_{jl} - \boldsymbol{c}^\top\boldsymbol{\Theta}_{*,jl}\Big\} = -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\zeta_{k,jl}(\boldsymbol{c}) + o_p(1) \xrightarrow{d} \mathcal{N}\big(0, \boldsymbol{c}^\top\boldsymbol{\Sigma}_{jl}\boldsymbol{c}\big).$$

This proves the theorem. □

Proof of Corollary 4.1. Consider two graphs $m_1 \neq m_2$ and the null hypothesis

$$H_0 : \Theta_{*,jl}^{(m_1)} = \Theta_{*,jl}^{(m_2)}.$$

Define the contrast vector

$$\boldsymbol{c} = \frac{1}{\sqrt{2}}\big(0, \ldots, 1, \ldots, -1, \ldots, 0\big)^\top,$$

where the nonzero entries $+1/\sqrt{2}$ and $-1/\sqrt{2}$ correspond to graphs m_1 and m_2, respectively. Then

$$\Delta_{jl}(\boldsymbol{c}) = \boldsymbol{c}^\top\big(\Theta^{(1)}_{*,jl}, \ldots, \Theta^{(M)}_{*,jl}\big) = \frac{1}{\sqrt{2}}\big(\Theta^{(m_1)}_{*,jl} - \Theta^{(m_2)}_{*,jl}\big).$$

Under H_0 we have $\Theta^{(m_1)}_{*,jl} = \Theta^{(m_2)}_{*,jl}$ and hence $\Delta_{jl}(\boldsymbol{c}) = 0$.

By Theorem 4.1, it follows that

$$\sqrt{n}\boldsymbol{c}^\top\widetilde{\boldsymbol{\Theta}}_{jl} = \sqrt{n}\frac{\widetilde{\Theta}^{(m_1)}_{jl} - \widetilde{\Theta}^{(m_2)}_{jl}}{\sqrt{2}} \xrightarrow{d} \mathcal{N}\big(0, \sigma^2_{jl}(\boldsymbol{c})\big),$$

where $\sigma^2_{jl}(\boldsymbol{c}) = \boldsymbol{c}^\top\boldsymbol{\Sigma}_{jl}\boldsymbol{c}$.

Under $\boldsymbol{c} = \frac{1}{\sqrt{2}}\big(0, \ldots, 1, \ldots, -1, \ldots, 0\big)^\top$,the corresponding asymptotic variance simplifies to $\sigma^2_{jl}(\boldsymbol{c}) = \frac{1}{2}\big(\tau^2_{m_1,jl} + \tau^2_{m_2,jl}\big)$, and by Proposition 3.2,

$$\widehat{\sigma}^2_{jl}(m_1, m_2) = \frac{1}{2}\big(\widehat{\tau}^2_{m_1,jl} + \widehat{\tau}^2_{m_2,jl}\big) \xrightarrow{p} \sigma^2_{jl}(\boldsymbol{c}).$$

Define the test statistic

$$T^{(m_1,m_2)}_{jl} = \sqrt{n}\frac{\widetilde{\Theta}^{(m_1)}_{jl} - \widetilde{\Theta}^{(m_2)}_{jl}}{\widehat{\sigma}_{jl}(m_1, m_2)}.$$

Combining the above asymptotic normality with the consistency of $\widehat{\sigma}_{jl}(m_1, m_2)$ and applying Slutsky's theorem, we obtain that

$$T^{(m_1,m_2)}_{jl} \xrightarrow{d} \mathcal{N}(0, 1) \text{ under } H_0.$$

Therefore, a level-α two-sided test rejects H_0 whenever

$$\left|T_{jl}^{(m_1,m_2)}\right| > z_{1-\alpha/2},$$

which completes the proof. □

S3 More Simulation Results

S3.1 p=50, Model I

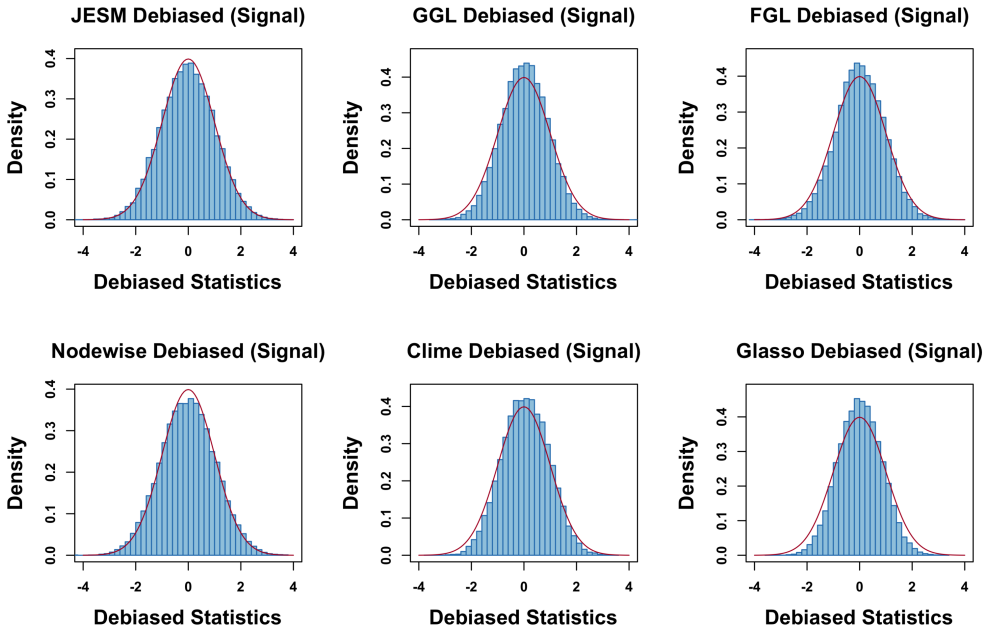


Figure S1: Empirical distributions of debiased statistics for signal edges.

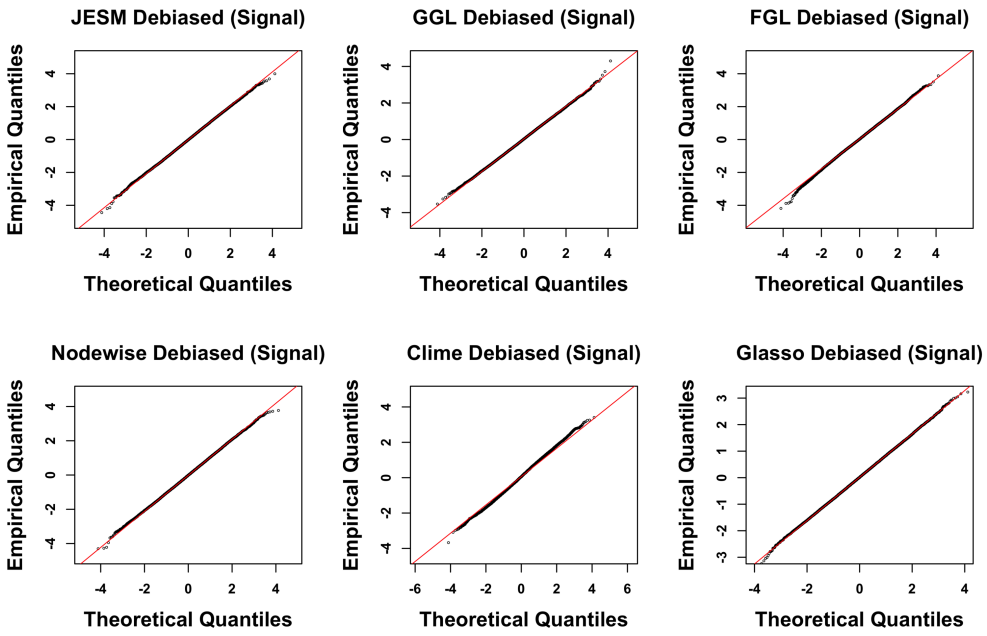


Figure S2: Quantile-quantile plots of debiased statistics for signal edges.

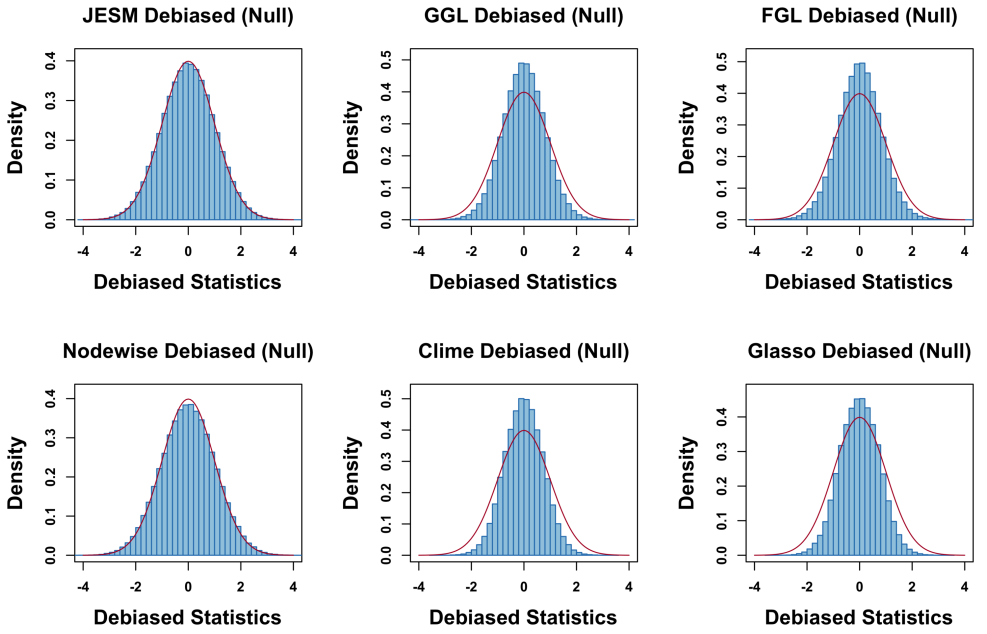


Figure S3: Empirical distributions of debiased statistics for null edges.

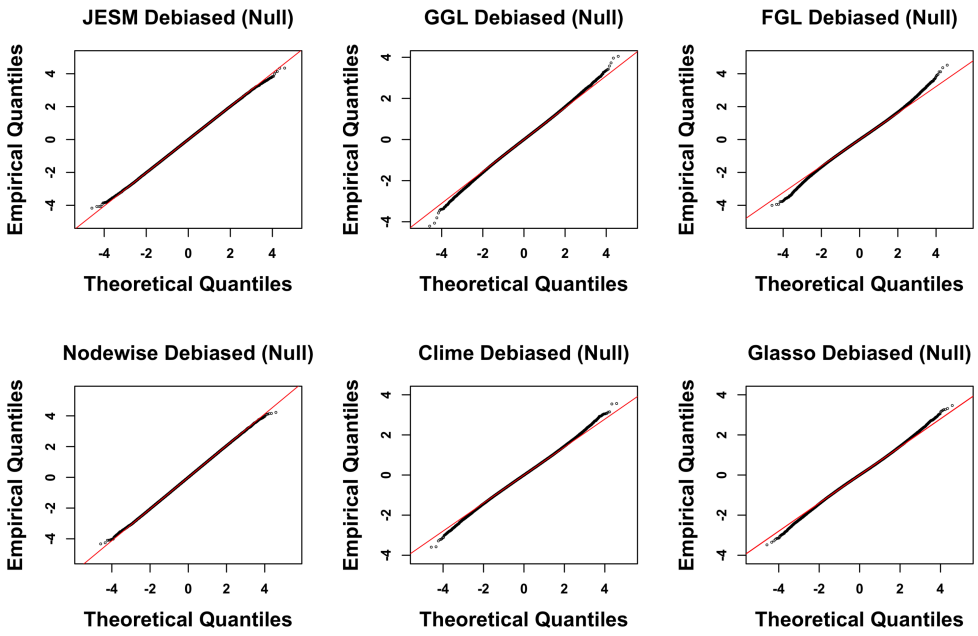


Figure S4: Quantile-quantile plots of debiased statistics for null edges.

S3.2 p=50, Model II

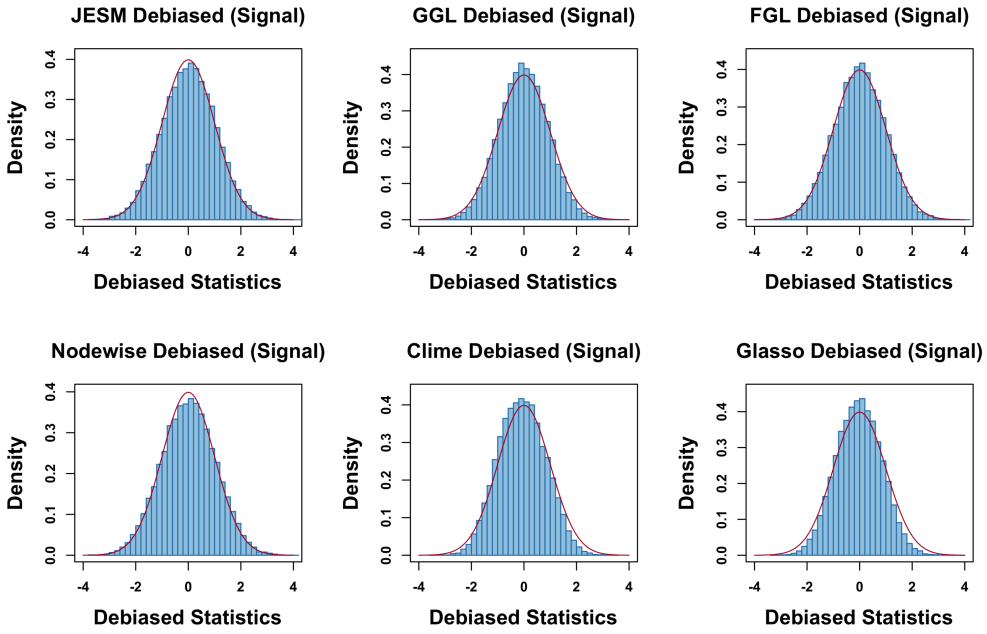


Figure S5: Empirical distributions of debiased statistics for signal edges.

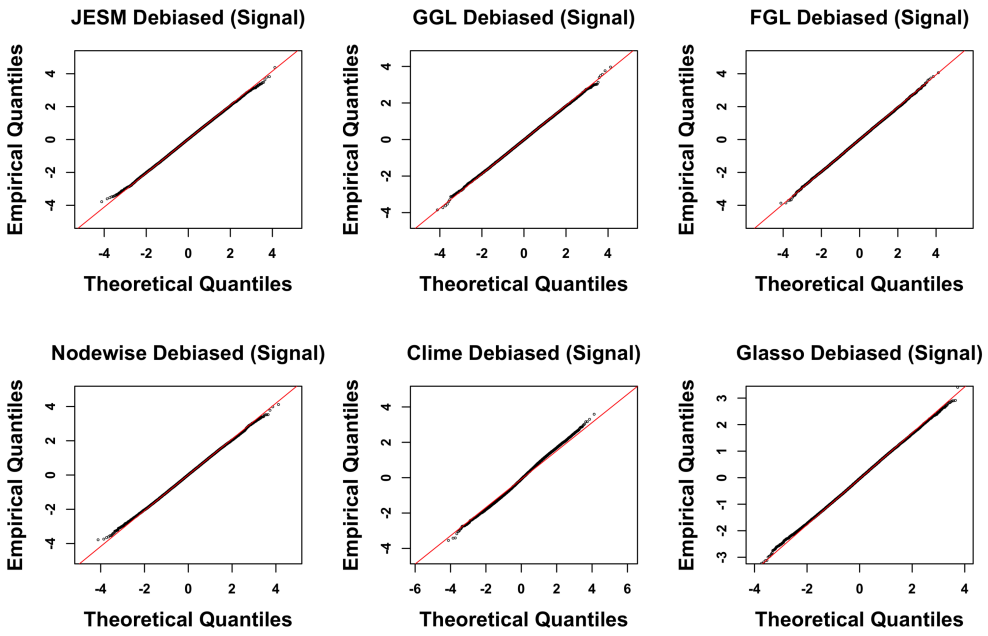


Figure S6: Quantile-quantile plots of debiased statistics for signal edges.

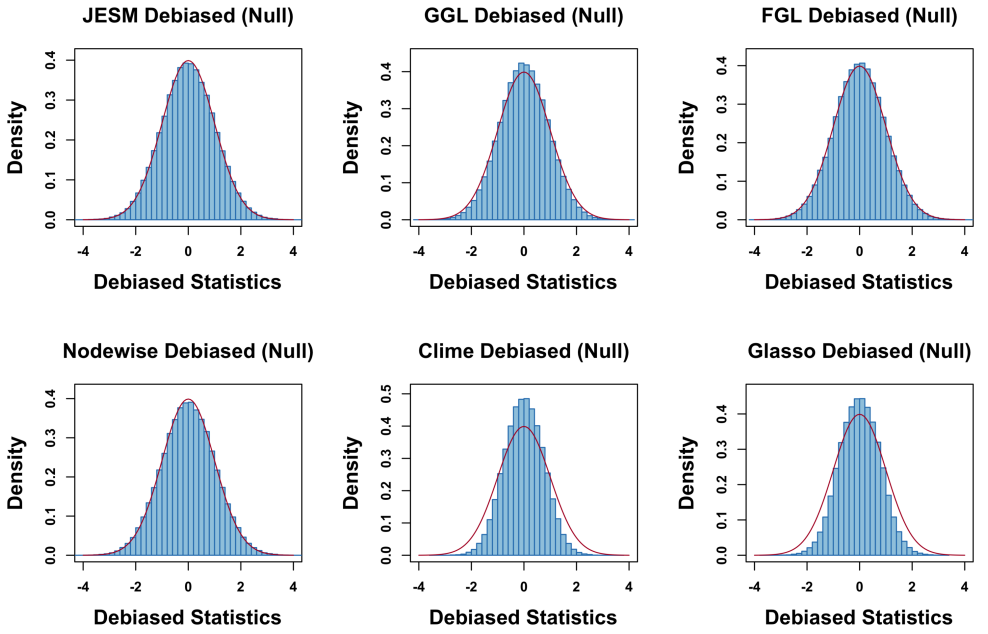


Figure S7: Empirical distributions of debiased statistics for null edges.

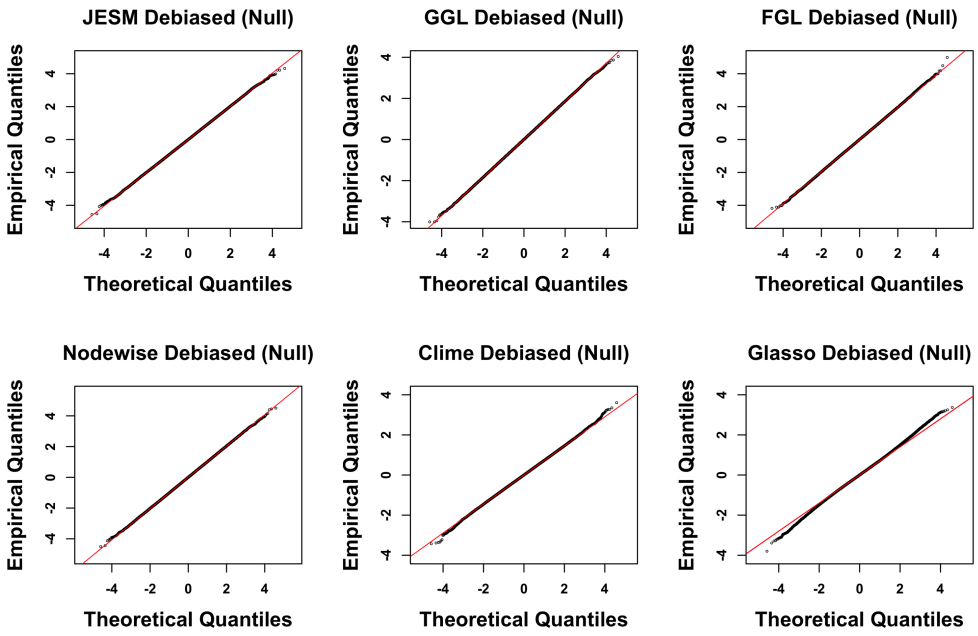


Figure S8: Quantile-quantile plots of debiased statistics for null edges.

S3.3 p=50, Model III

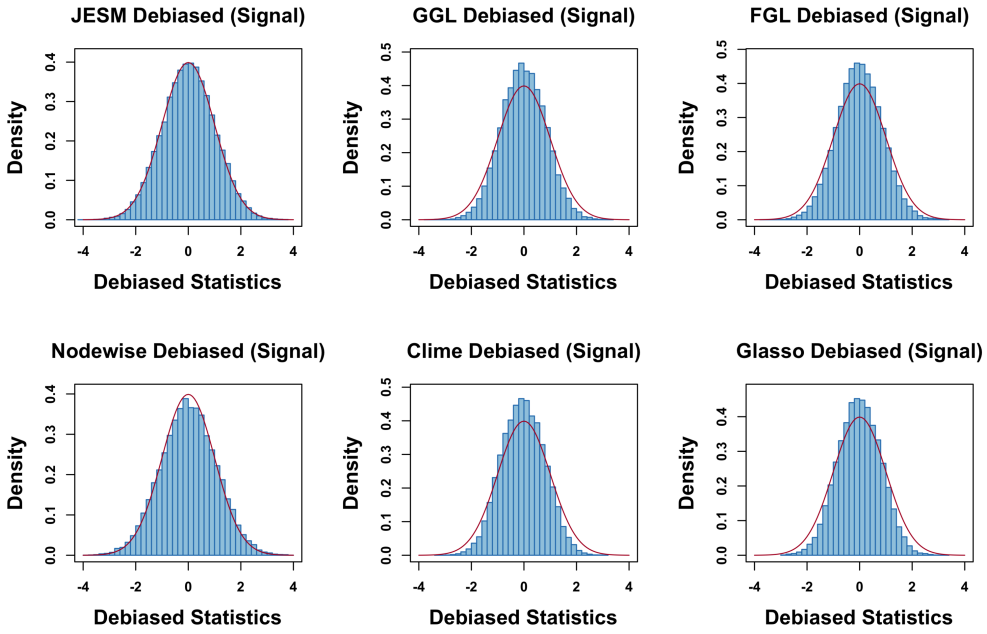


Figure S9: Empirical distributions of debiased statistics for signal edges.

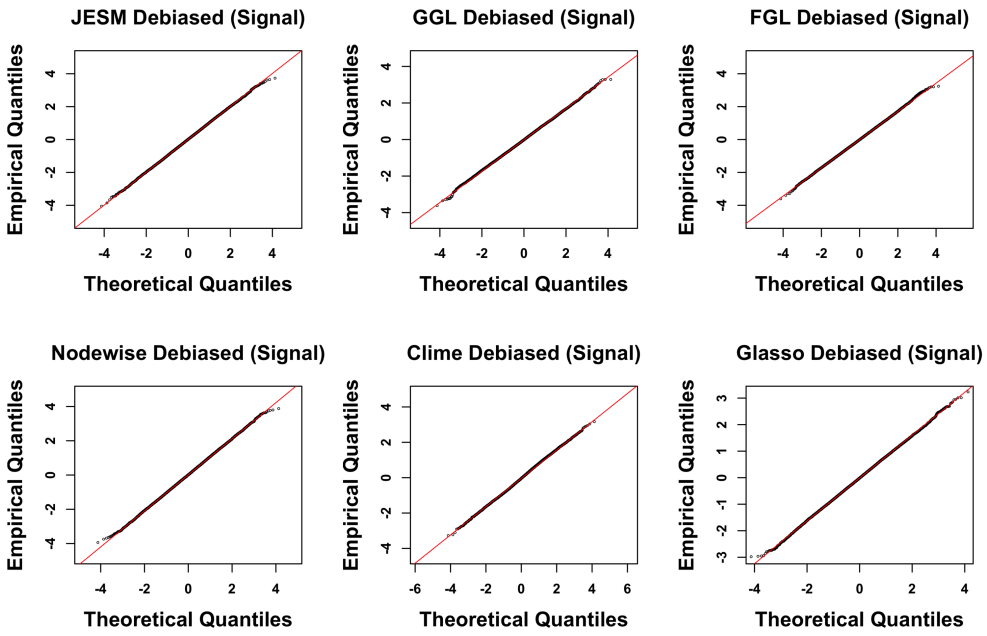


Figure S10: Quantile-quantile plots of debiased statistics for signal edges.

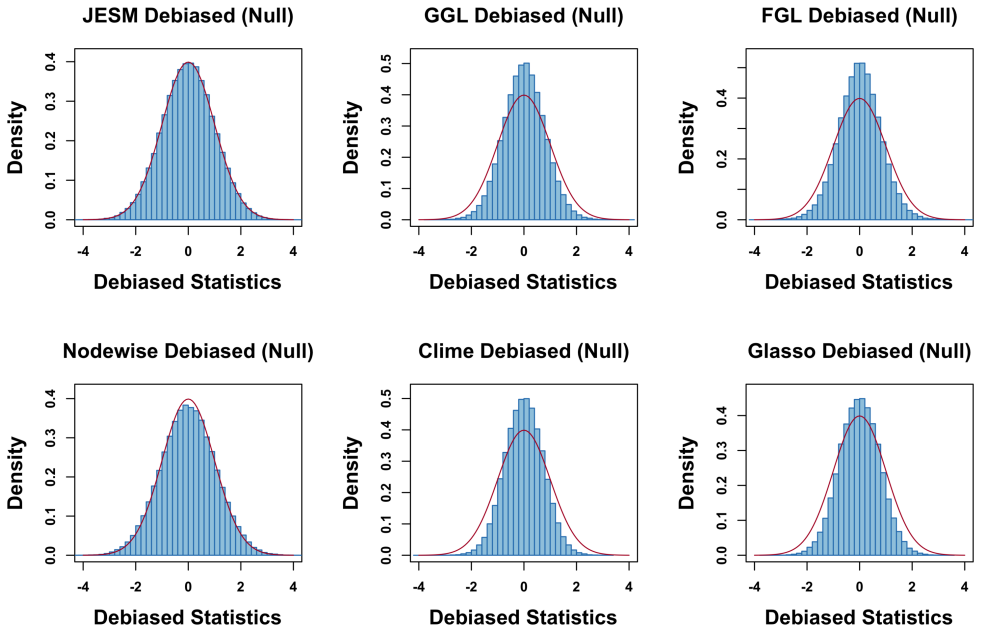


Figure S11: Empirical distributions of debiased statistics for null edges.

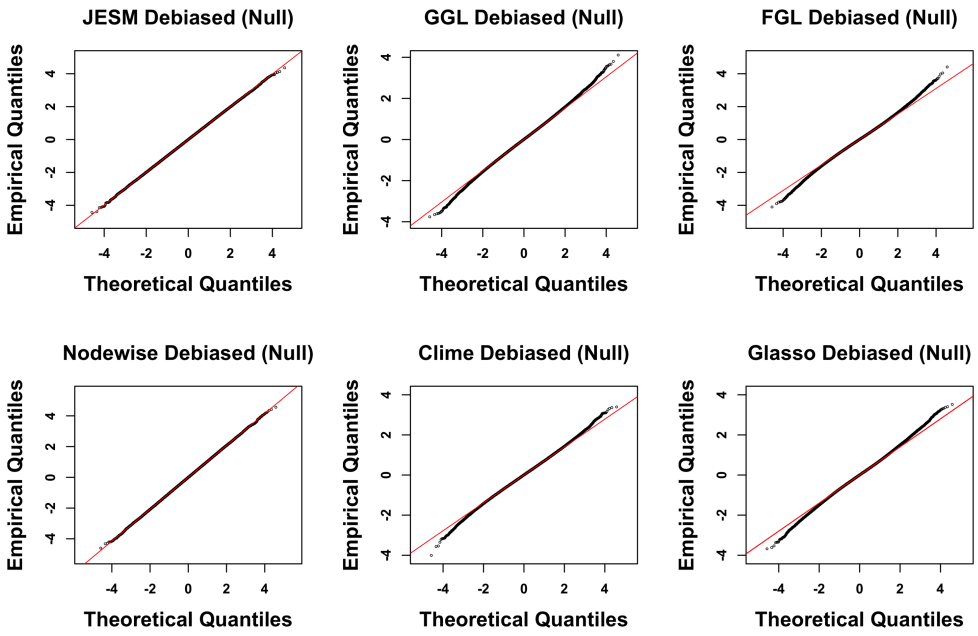


Figure S12: Quantile-quantile plots of debiased statistics for null edges.

S3.4 p=100, Model II

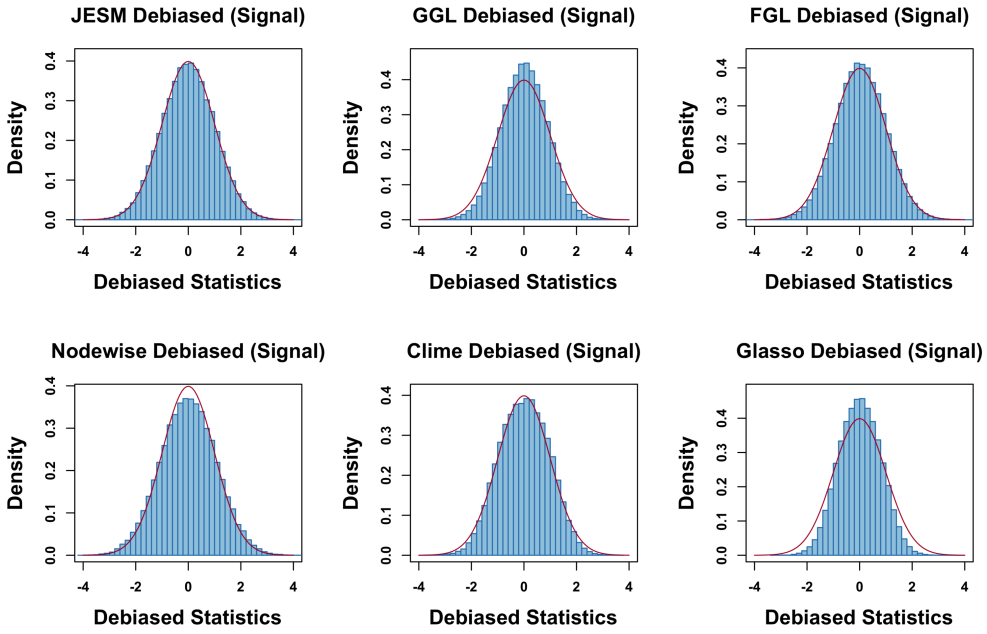


Figure S13: Empirical distributions of debiased statistics for signal edges.

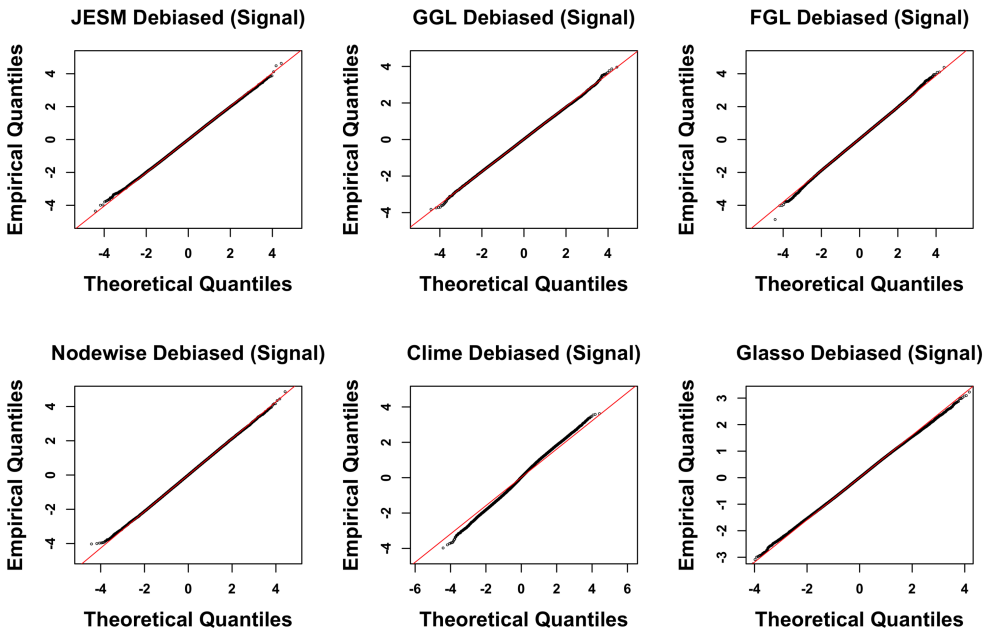


Figure S14: Quantile-quantile plots of debiased statistics for signal edges.

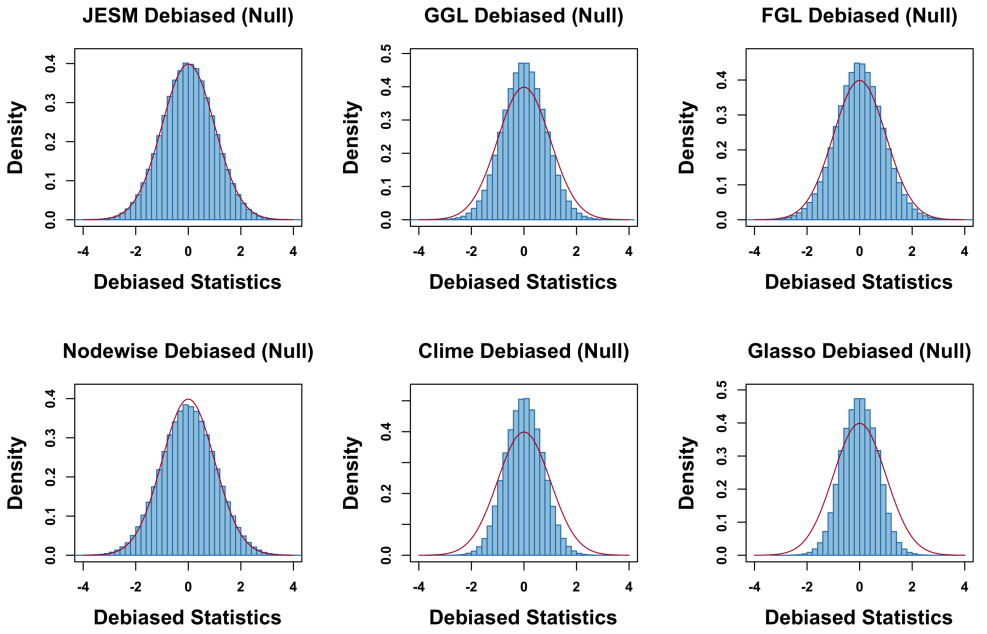


Figure S15: Empirical distributions of debiased statistics for null edges.

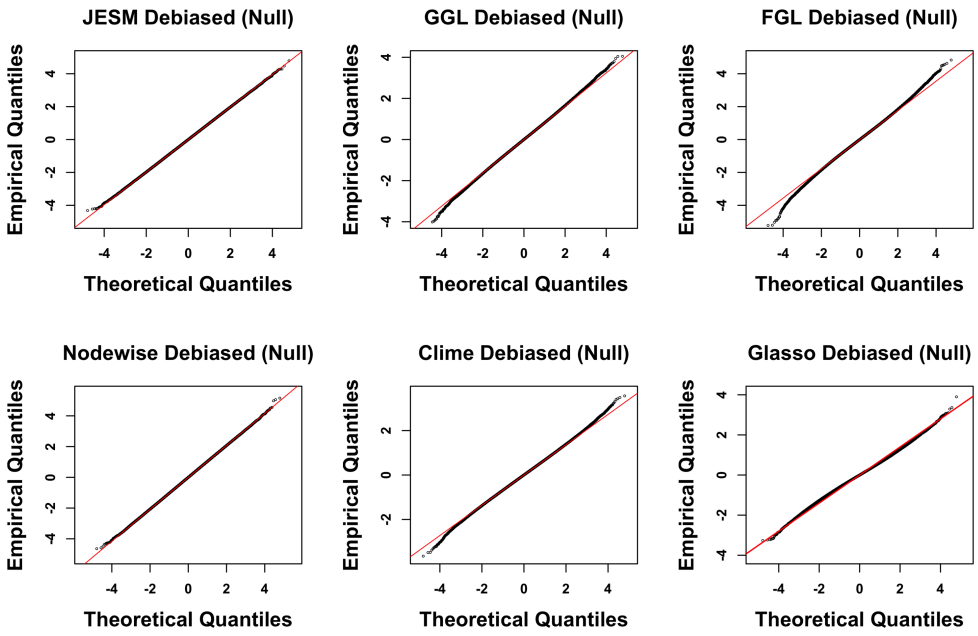


Figure S16: Quantile-quantile plots of debiased statistics for null edges.

S3.5 p=100, Model III

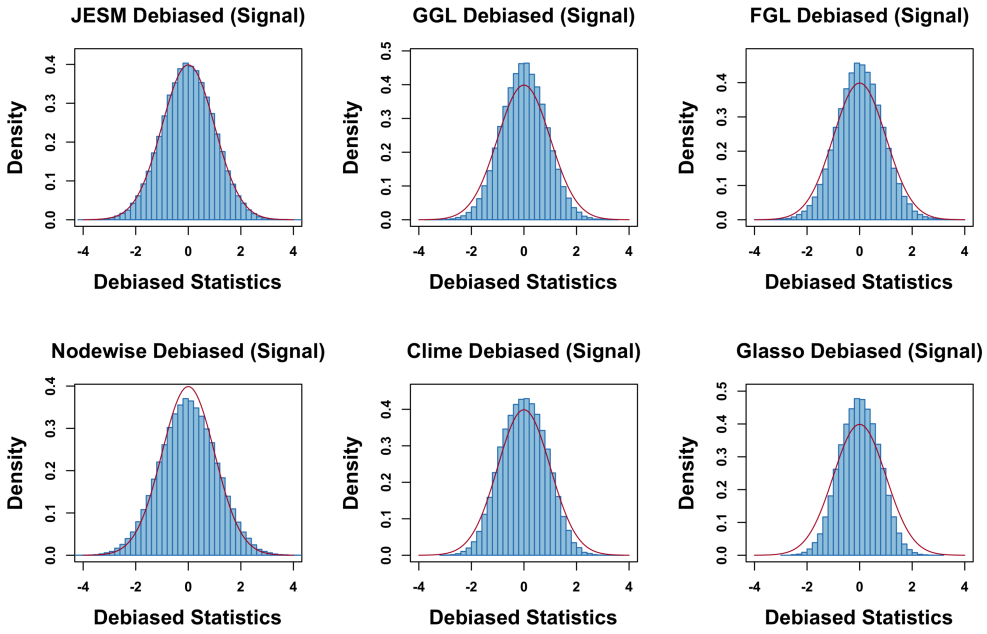


Figure S17: Empirical distributions of debiased statistics for signal edges.

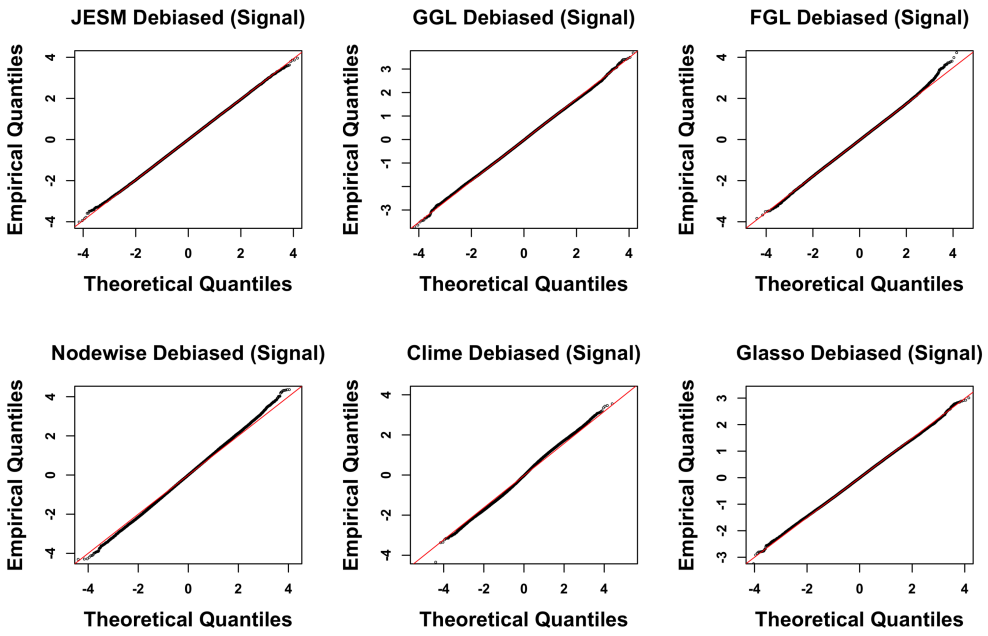


Figure S18: Quantile-quantile plots of debiased statistics for signal edges.

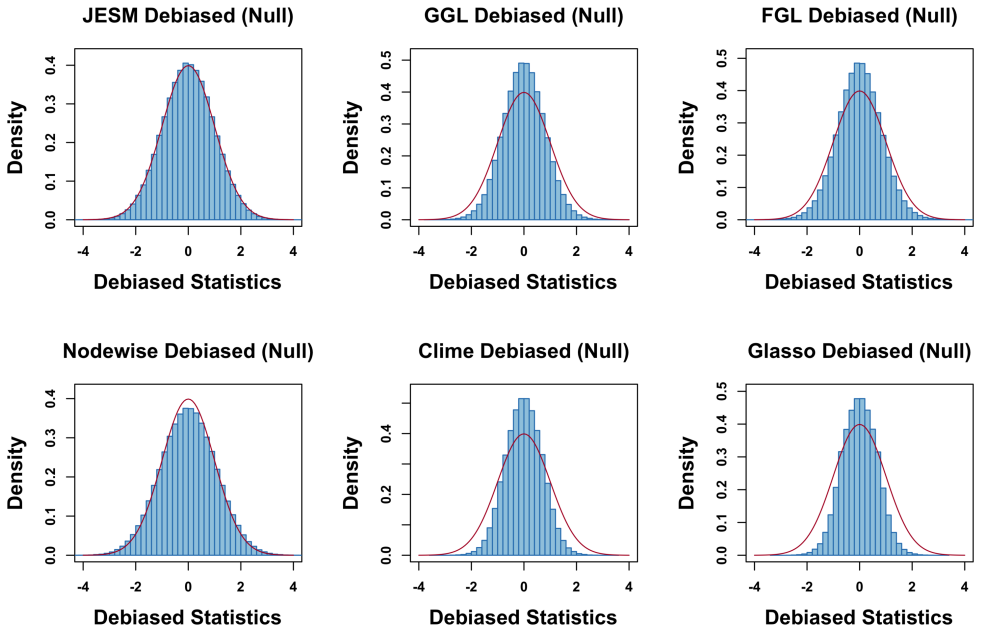


Figure S19: Empirical distributions of debiased statistics for null edges.

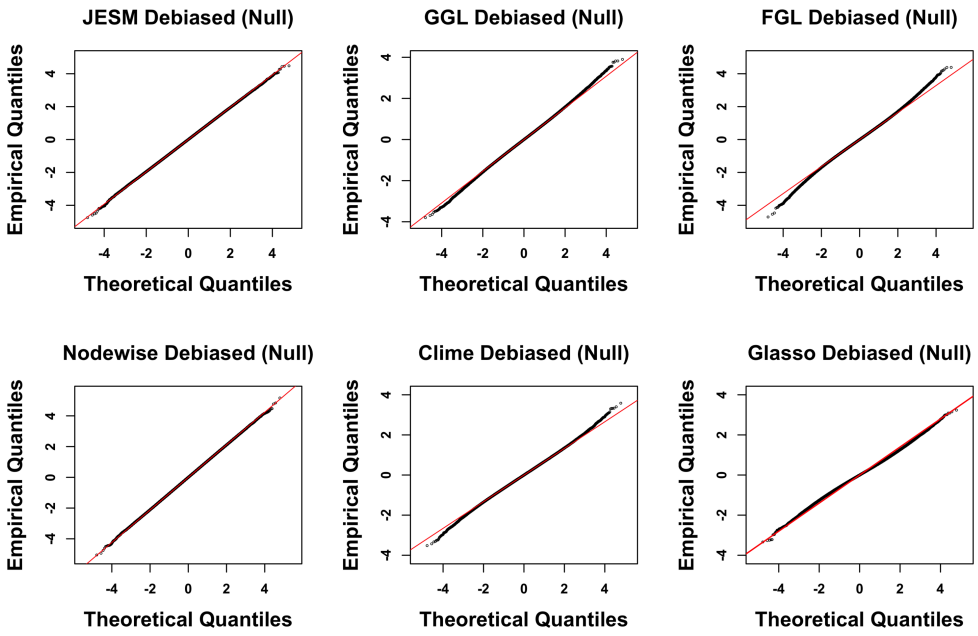


Figure S20: Quantile-quantile plots of debiased statistics for null edges.

S3.6 p=200, Model I

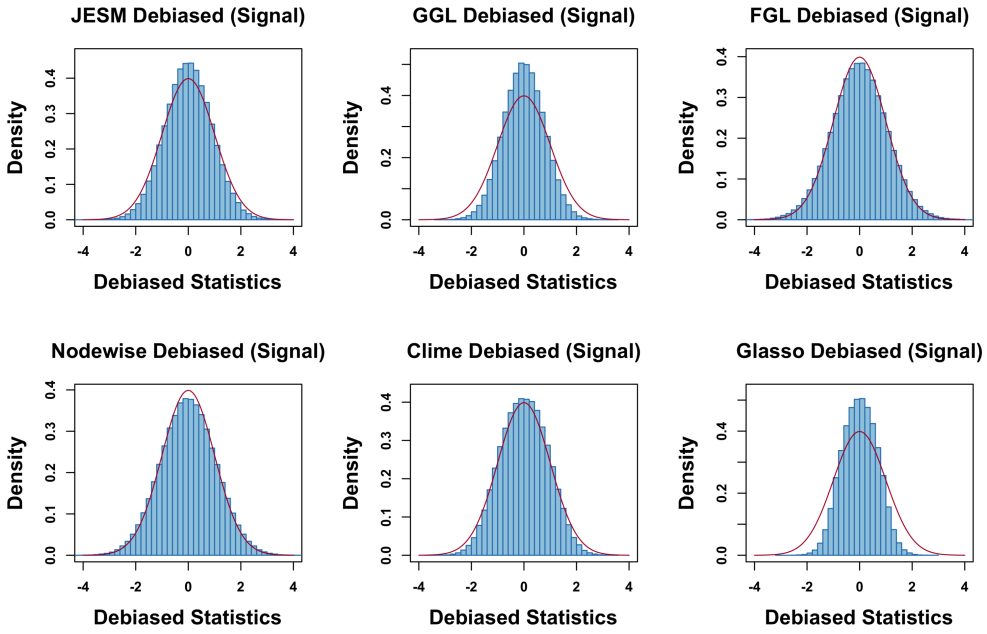


Figure S21: Empirical distributions of debiased statistics for signal edges.

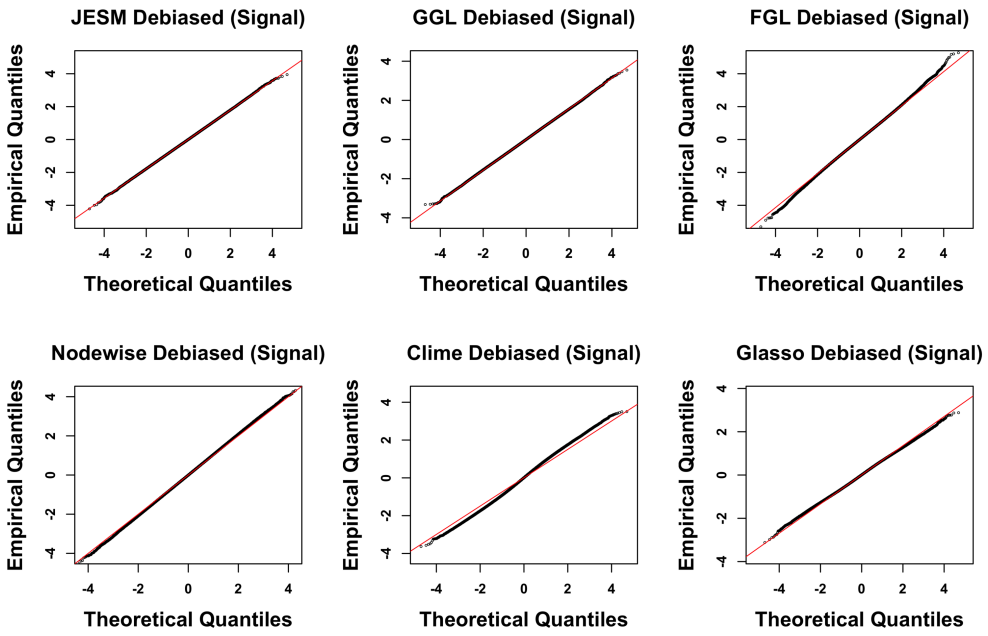


Figure S22: Quantile-quantile plots of debiased statistics for signal edges.

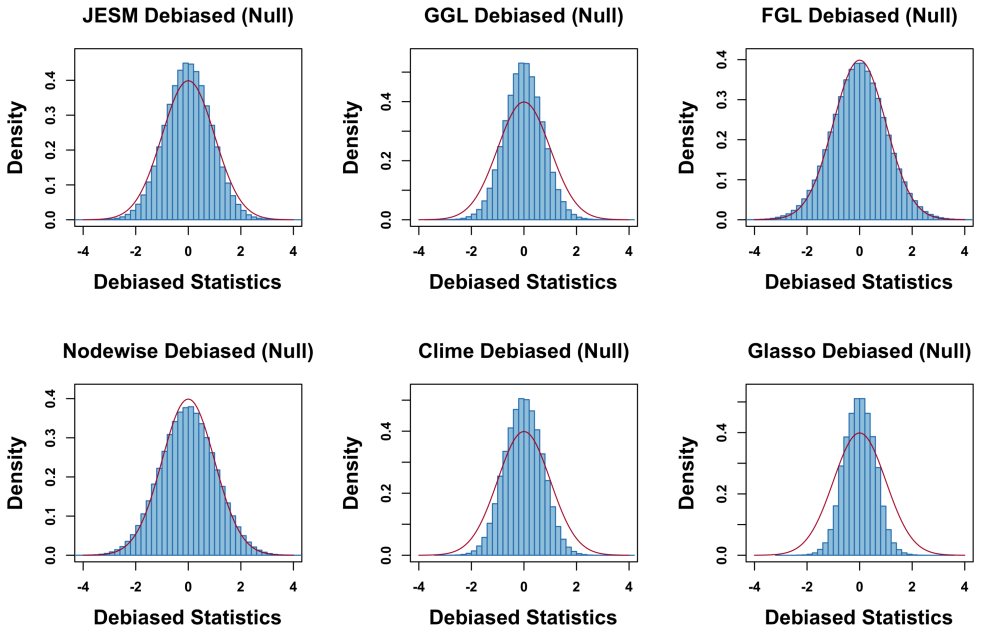


Figure S23: Empirical distributions of debiased statistics for null edges.

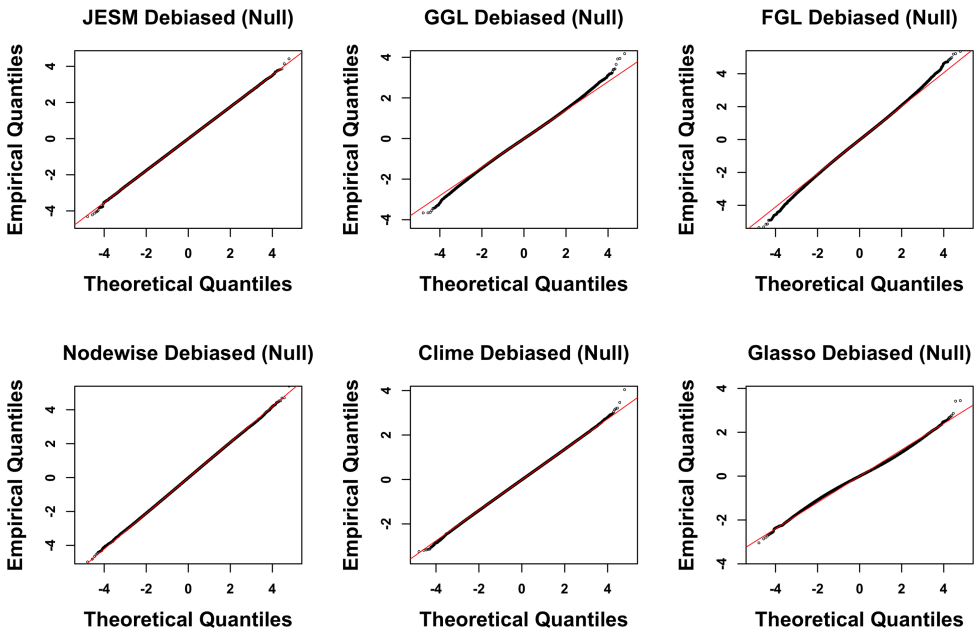


Figure S24: Quantile-quantile plots of debiased statistics for null edges.

S3.7 p=200, Model II

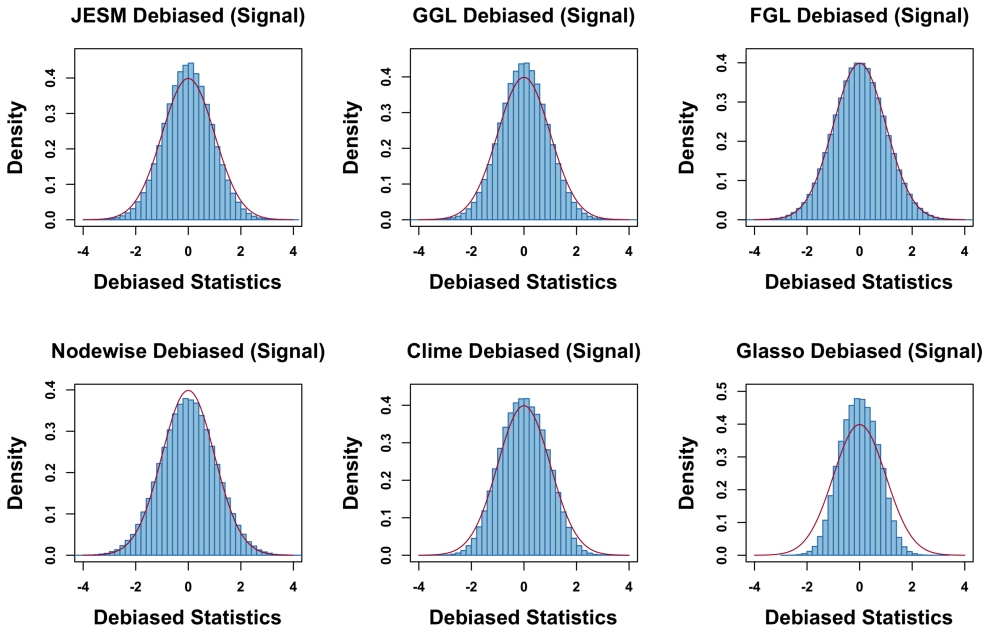


Figure S25: Empirical distributions of debiased statistics for signal edges.

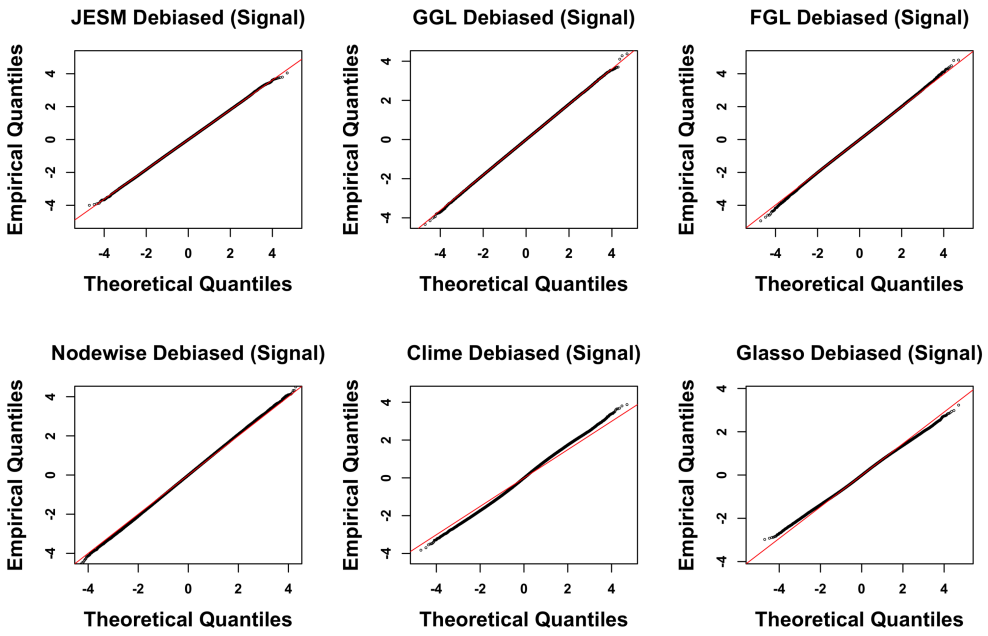


Figure S26: Quantile-quantile plots of debiased statistics for signal edges.

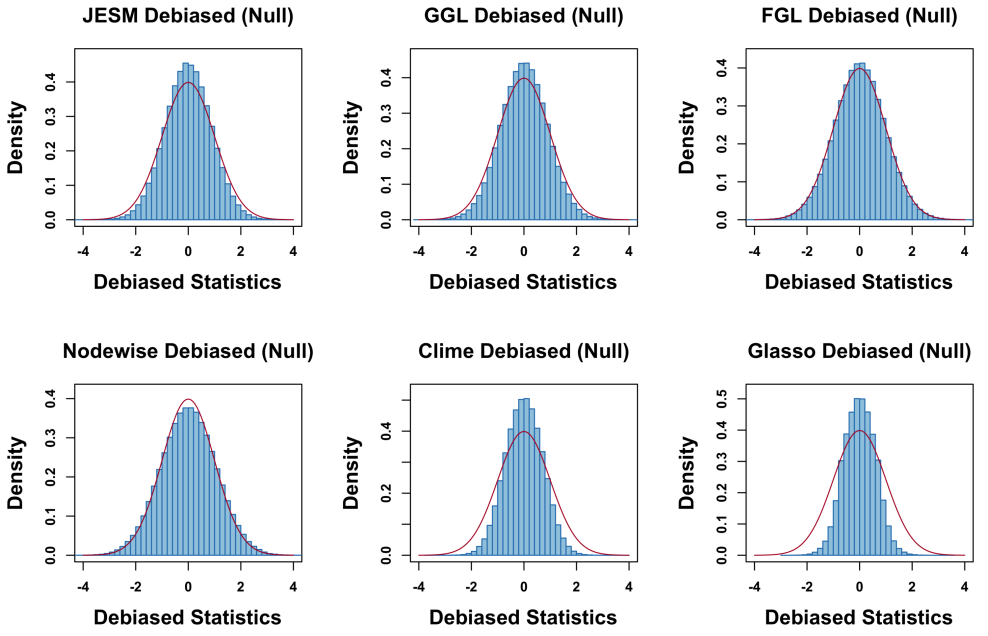


Figure S27: Empirical distributions of debiased statistics for null edges.

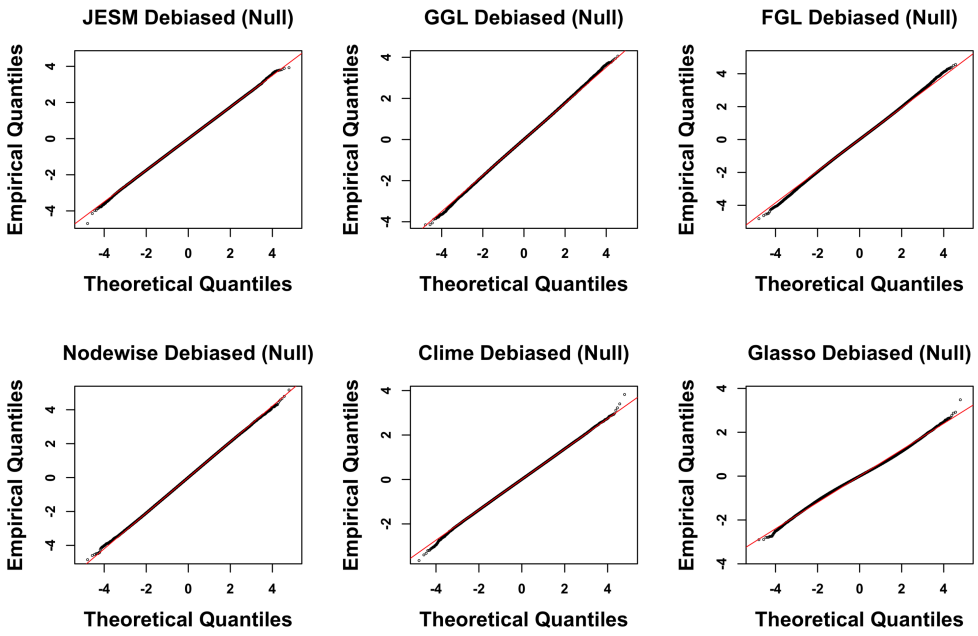


Figure S28: Quantile-quantile plots of debiased statistics for null edges.

S3.8 p=200, Model III

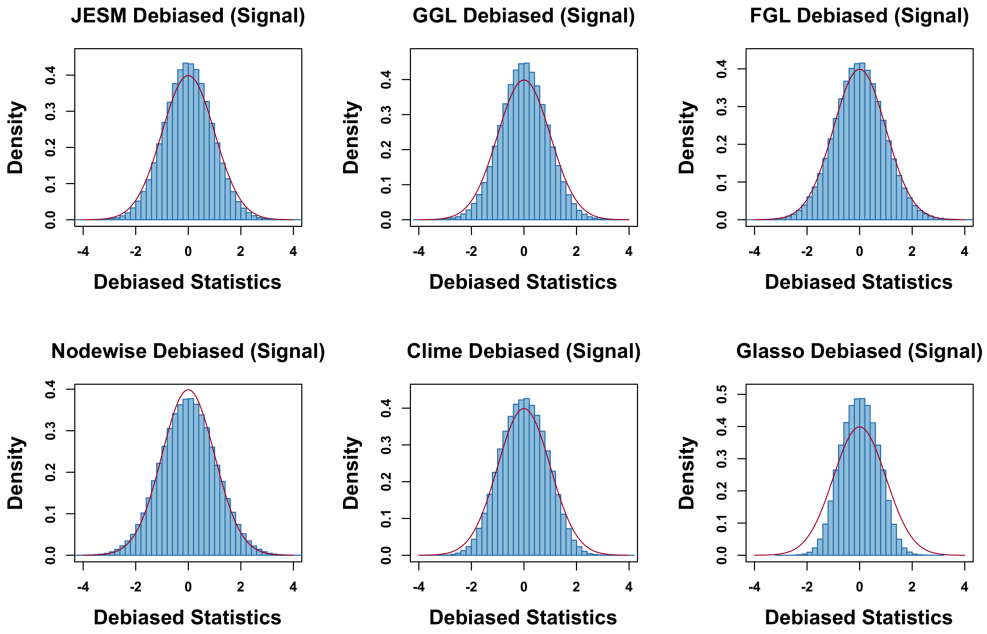


Figure S29: Empirical distributions of debiased statistics for signal edges.

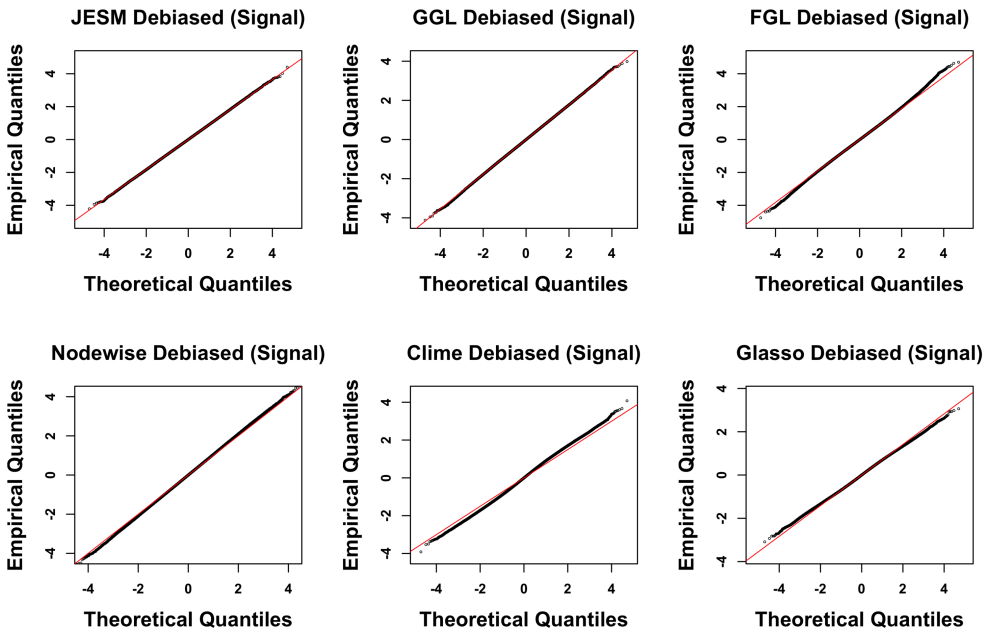


Figure S30: Quantile-quantile plots of debiased statistics for signal edges.

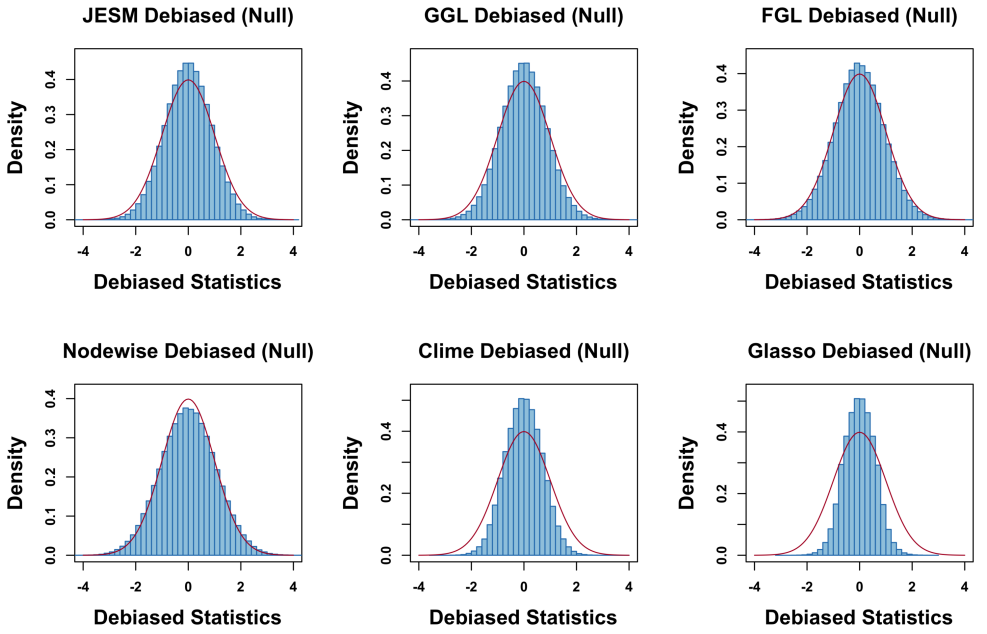


Figure S31: Empirical distributions of debiased statistics for null edges.

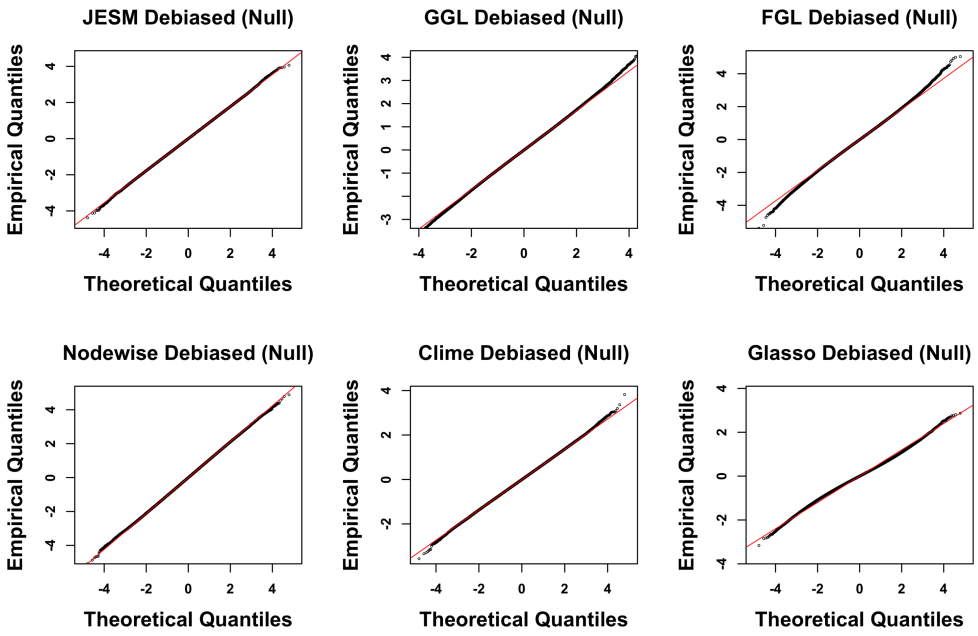


Figure S32: Quantile-quantile plots of debiased statistics for null edges.

S3.9 p=400, Model I

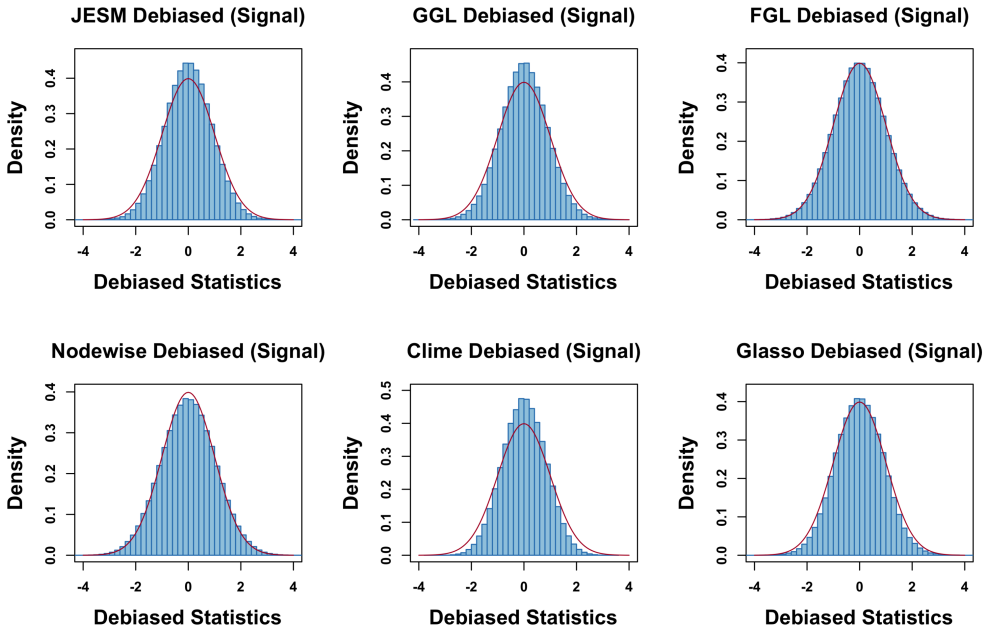


Figure S33: Empirical distributions of debiased statistics for signal edges.

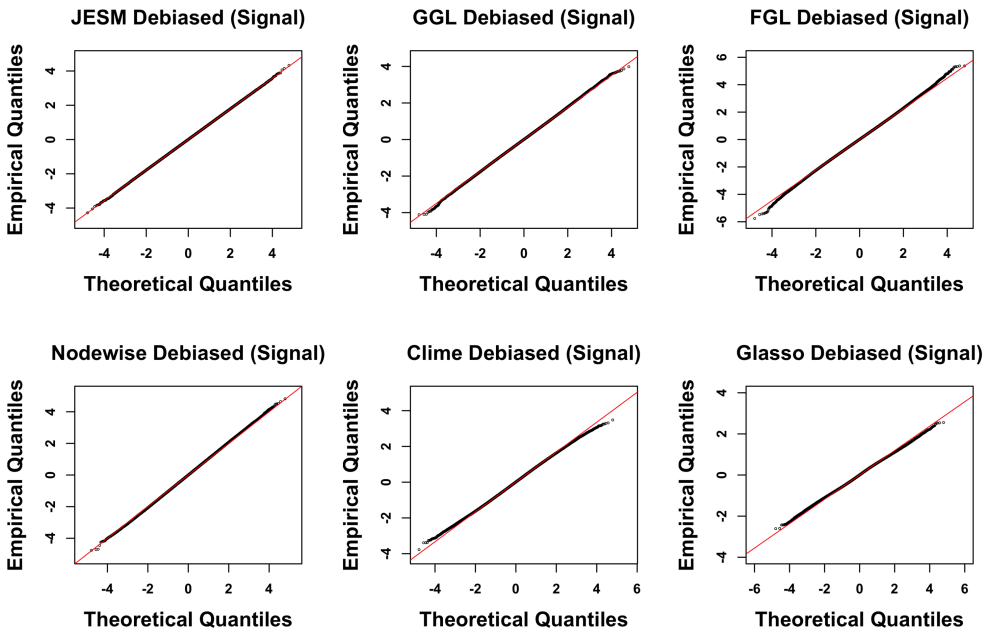


Figure S34: Quantile-quantile plots of debiased statistics for signal edges.

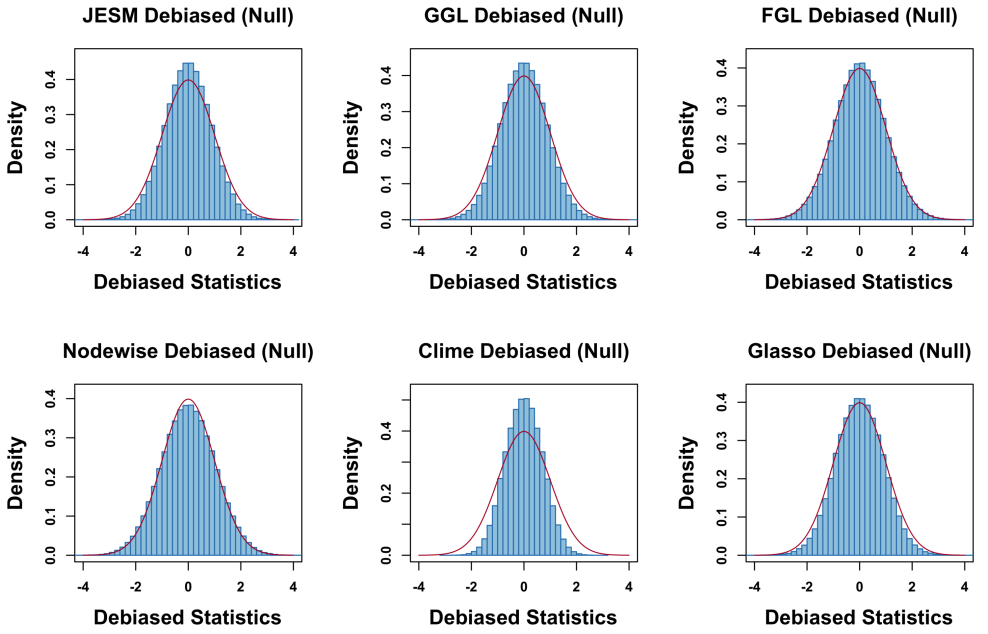


Figure S35: Empirical distributions of debiased statistics for null edges.

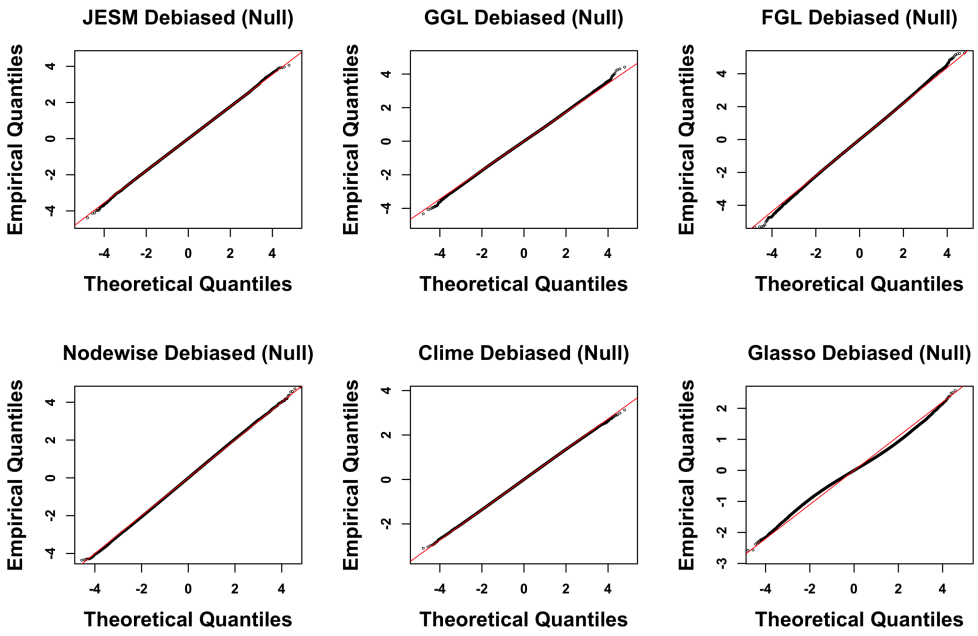


Figure S36: Quantile-quantile plots of debiased statistics for null edges.

S3.10 p=400, Model II

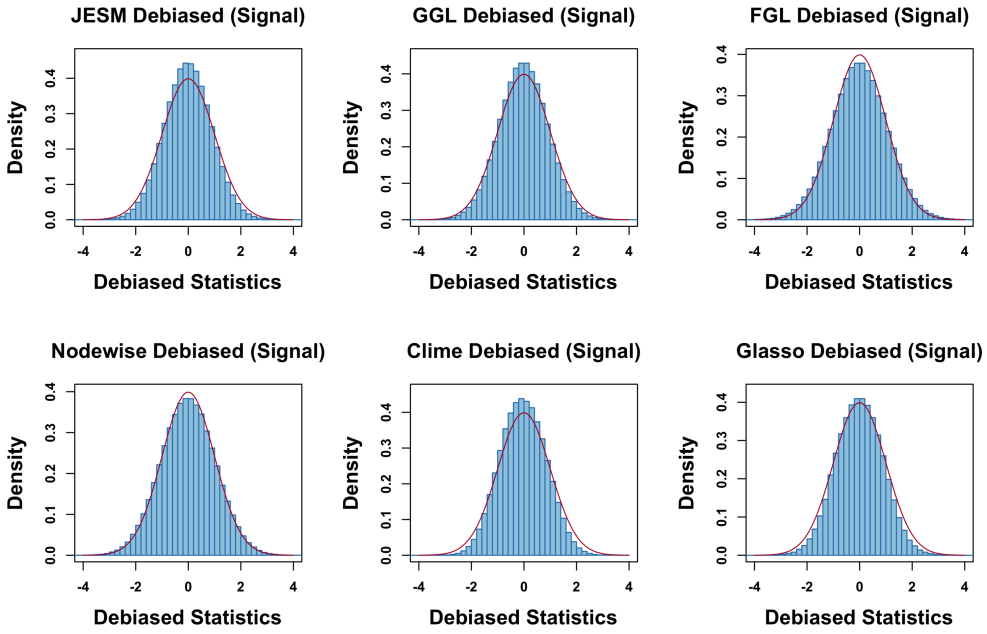


Figure S37: Empirical distributions of debiased statistics for signal edges.

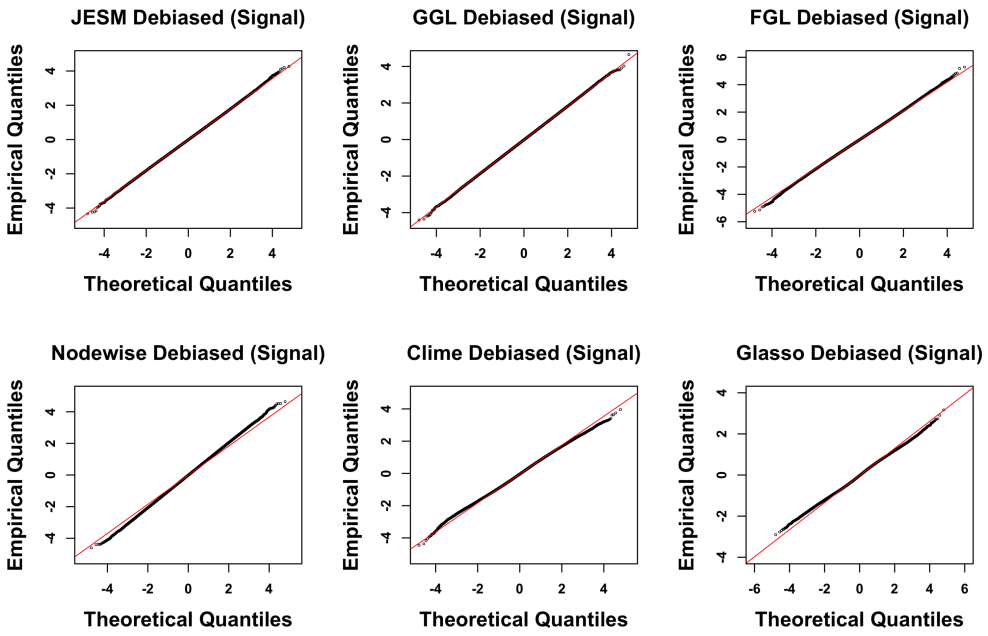


Figure S38: Quantile-quantile plots of debiased statistics for signal edges.

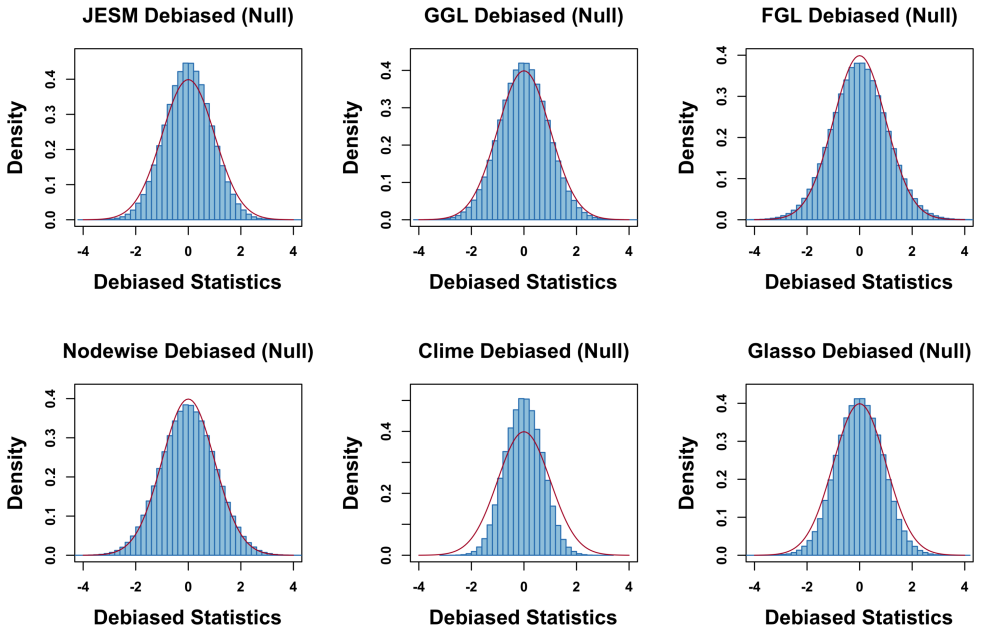


Figure S39: Empirical distributions of debiased statistics for null edges.

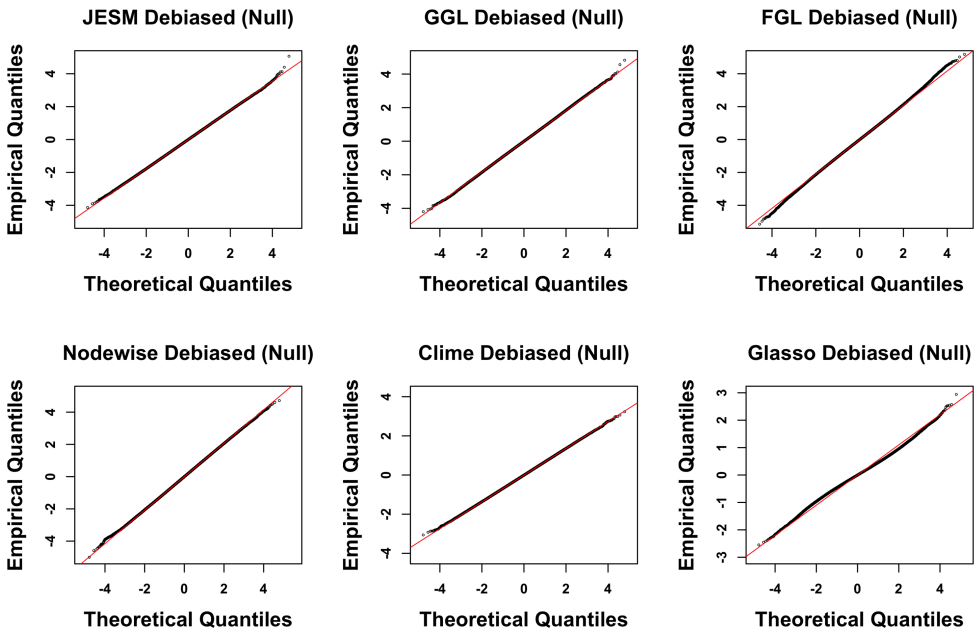


Figure S40: Quantile-quantile plots of debiased statistics for null edges.

S3.11 p=400, Model III

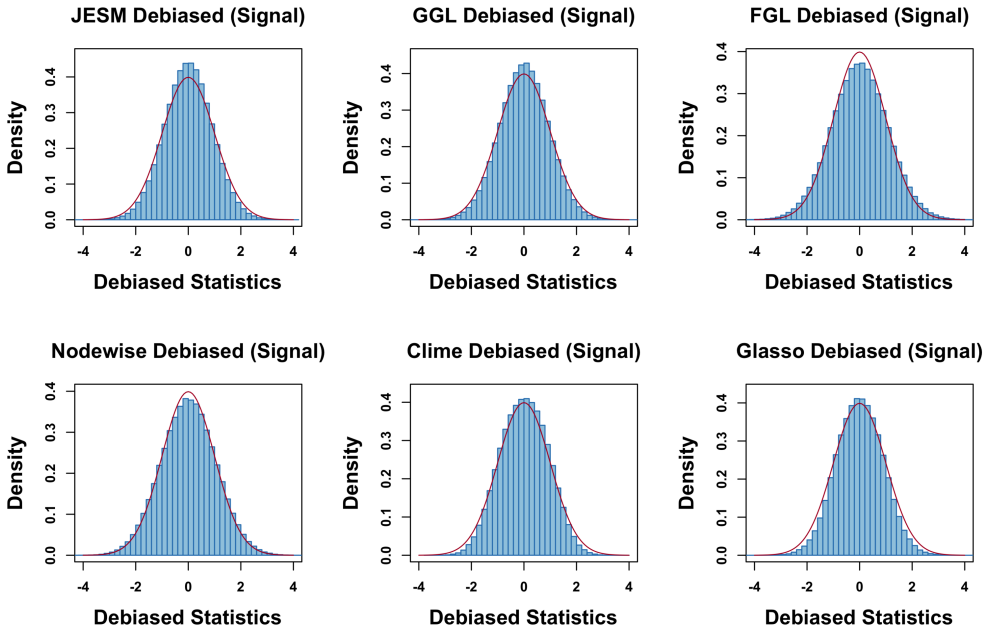


Figure S41: Empirical distributions of debiased statistics for signal edges.

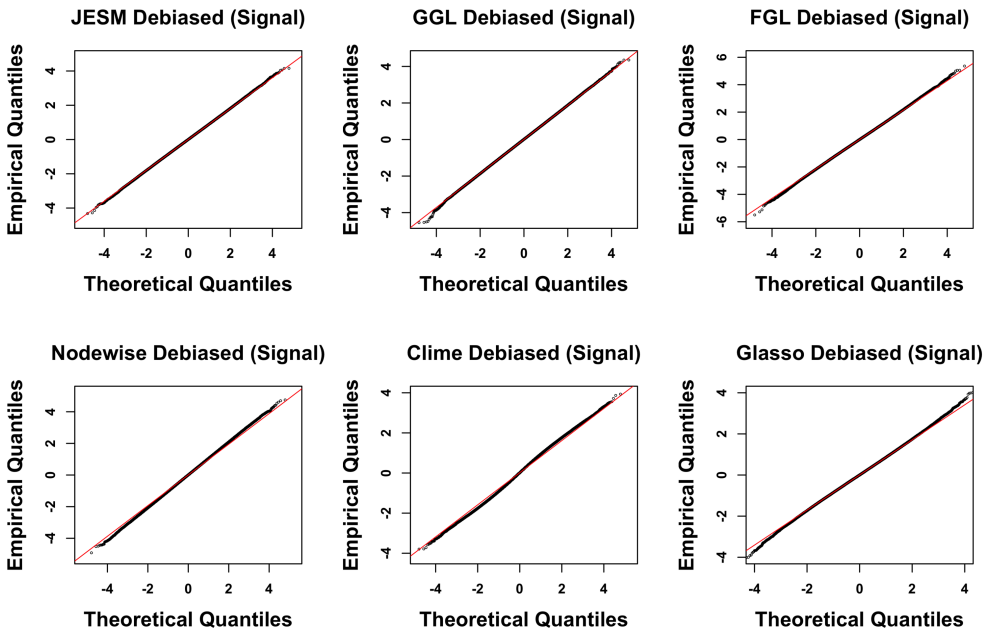


Figure S42: Quantile-quantile plots of debiased statistics for signal edges.

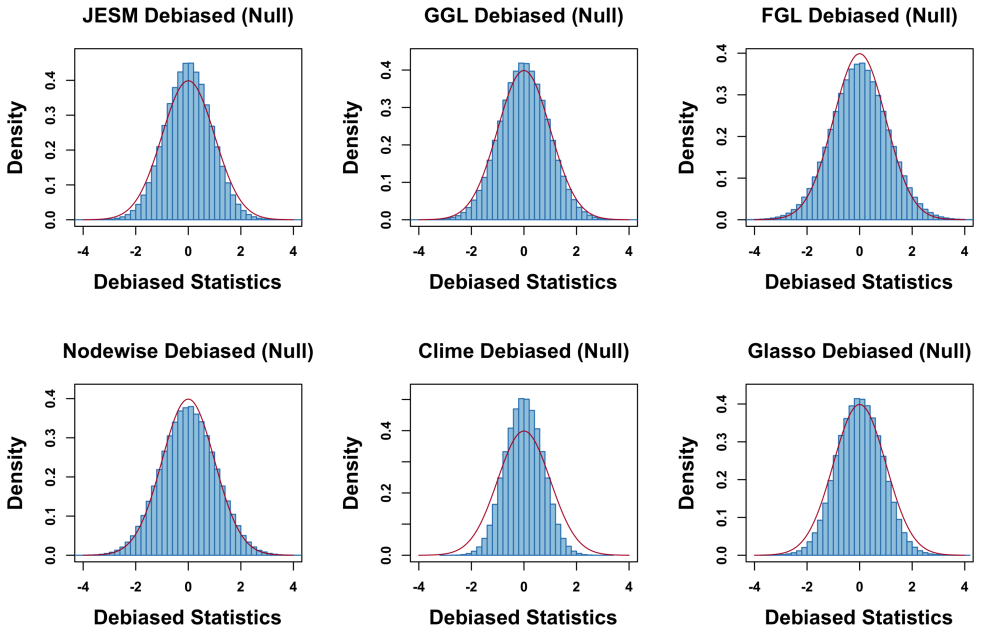


Figure S43: Empirical distributions of debiased statistics for null edges.

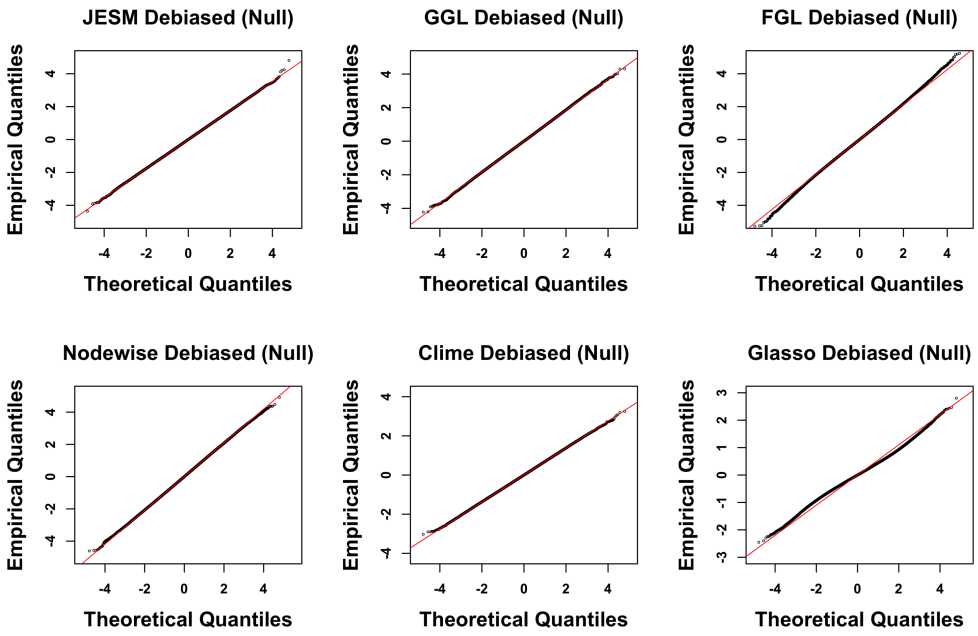


Figure S44: Quantile-quantile plots of debiased statistics for null edges.

S4 Additional Real-Data Results

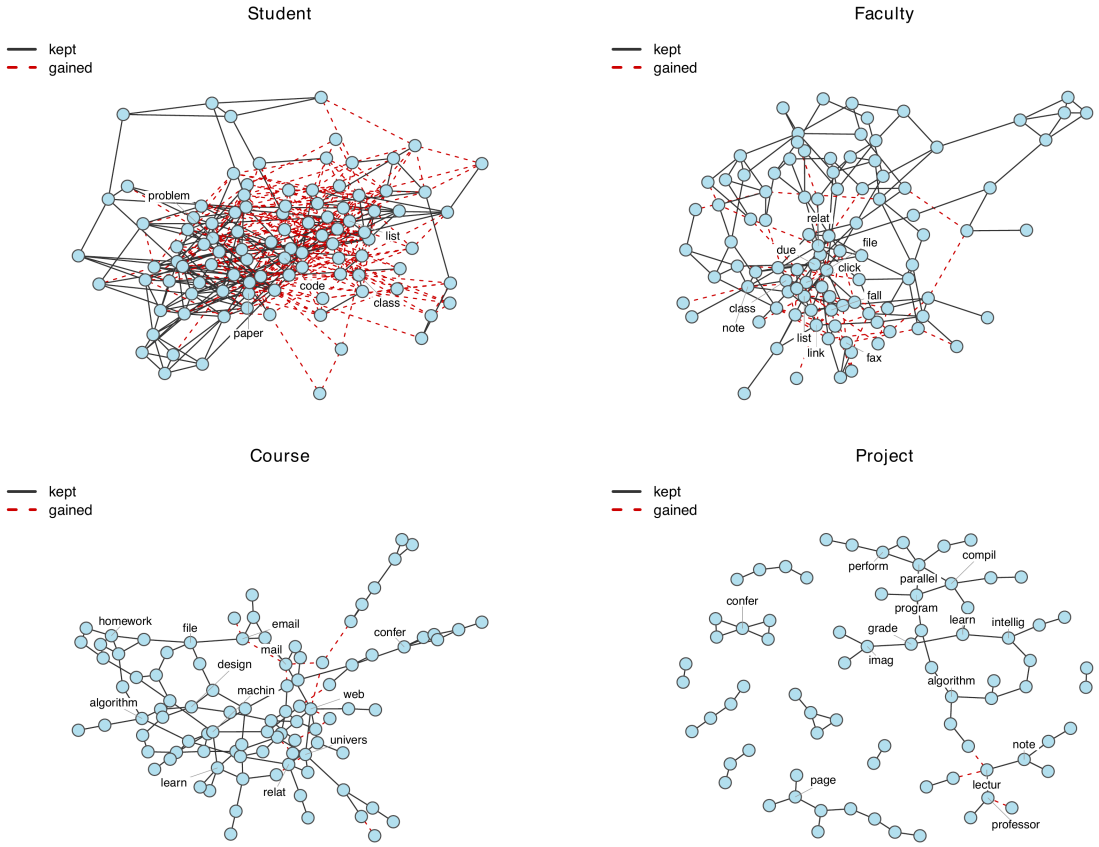


Figure S45: Visualization of WebKB data: Debiased edge discoveries with FDR control at level $q = 0.1$ under $p = 100$.

Category k	n_k	q	$\lvert E^{(k)}_{\text{init}}\rvert$	$\lvert E^{(k)}_{\text{sig}}\rvert$	$\lvert E^{(k)}_{\text{kept}}\rvert$	$\lvert E^{(k)}_{\text{lost}}\rvert$	$\lvert E^{(k)}_{\text{gained}}\rvert$
Student	1625	0.1	400	474	192	208	282
Faculty	1116	0.1	605	217	152	453	65
Course	926	0.1	687	128	115	572	13
Project	501	0.1	516	72	69	447	3

Table S1: Summary of WebKB data: Debiased edge inference with FDR control at level $q = 0.1$ under $p = 100$.

References

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29 (4), 1165–1188.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Annals of Statistics 41 (6), 2786–2819.