5. Supplementary Materials

5.1 *A new algorithm to optimize the objective function (1)*

The objective function

$$\arg\max_{C,\{G_c\}} \sum_{C=1}^{C} \exp\left[\log\left\{\sum(w_{i,j}|e_{i,j} \in G_c)\right\} - \lambda_0 \log(|E_c|)\right]$$

is non-convex and NP hard. We solve it in two steps.

Firstly, we optimize $\{G_c\}$ with given C:

$$\begin{aligned}
&\arg\max_{\{G_c\}} \sum_{C=1}^{C} \exp\{\log(\sum(w_{i,j}|e_{i,j} \in G_c)) - \lambda_0 \log(|E_c|)\} \\
=&\arg\max_{\{G_c\}} \sum_{C=1}^{C} \left(\frac{\sum(w_{i,j}|e_{i,j} \in G_c)}{|E_c|}\right)^{\lambda_0} \left(\sum(w_{i,j}|e_{i,j} \in G_c)\right)^{1-\lambda_0} \\
\doteq&\arg\max_{\{G_c\}} \sum_{C=1}^{C} \rho_{CC}|V_c|, (\lambda_0 = 1/2, \rho_{CC} = \sum(w_{i,j}|e_{i,j} \in G_c)/|E_c|) \\
=&\arg\max_{\{G_c\}} \sum(w_{i,j}|e_{i,j} \in G)/|V| - \sum_{C=1}^{C} \sum_{C' \neq C} \rho_{CC'}(|V_c| + |V_{C'}|) \\
\Leftrightarrow&\arg\min_{\{G_c\}} \sum_{C=1}^{C} \sum_{C' \neq C} \rho_{CC'}(|V_c| + |V_{C'}|) \\
=&\arg\min_{\{G_c\}} \sum_{C=1}^{C} \sum_{C' \neq C} \frac{\sum(w_{i,j}|i \in G_c, j \in G_{C'})}{|V_c||V_{C'}|}(|V_c| + |V_{C'}|) \\
=&\arg\min_{\{G_c\}} \sum_{C=1}^{C} \frac{\sum(w_{i,j}|i \in G_c, j \notin G_C)}{|V_c|}
\end{aligned} \tag{1}$$

We solve objective function 1 by using spectral clustering algorithm RatioCut (Chen *et al*, 2015a).

Next, we select C by grid searching that maximizes the criteria:

$$\sum_{C=1}^{C}\left(\frac{\sum(w_{i,j}|e_{i,j}\in G_c)}{|E_c|}\right)^{\lambda_0}\left(\sum(w_{i,j}|e_{i,j}\in G_c)\right)^{1-\lambda_0} \quad (2)$$

At this step, a larger λ_0 often leads to detected subnetworks with higher proportion of more informative edges and smaller sizes whereas a smaller λ_0 often produces larger networks including more informative edges in G. The iterations of the above two steps implement the optimization of 1.

$\boldsymbol{W}$ matrix calculation

Let $z_{i,j}$ be the Fisher's Z transformed correlation coefficient $\widehat{R}_{i,j}$, for example. We can simply let $w_{i,j} = z_{i,j}$ or further transform it to the probability scale. Assume that sample correlation coefficients for all edges follow a mixture distribution $z_{i,j} \sim \pi_0 f_0(z_{i,j}) + \pi_1 f_1(z_{i,j})$ where $\pi_0 + \pi_1 = 1$ (Efron, 2004; Wu *et al*, 2006; Efron, 2007). f_1 represents the distribution of correlations corresponding to the component of connected edges $z_{i,j}|(\delta_{i,j} = 1) \sim f_0(z_{i,j})$, and f_0 for the unconnected edges $z_{i,j}|(\delta_{i,j} = 0) \sim f_1(z_{i,j})$. We adopt the empirical Bayes method to obtain $\hat{\pi}_0, \hat{\pi}_1, \hat{f}_0, \hat{f}_1$, (Efron, 2007) and then $w_{i,j}$ is the posterior probability that $z_{i,j}$ is from the non-null component.

Convenient thresholding value calculation

We calculate the thresholding values for edges inside-networks or outside-networks separately, and these cut-offs can be linked to the overall local fdr value. Therefore, the computation is more straightforward by using the following cut-offs.

An edge inside networks z^{in} is truly connected if $fdr^{in}(z^{in}) \leqslant 1/(T+1)$, where T is the threshold. Equivalently if $\frac{f_1(z)}{f_0(z)} \geqslant T\frac{\pi_0^{in}}{\pi_1^{in}}$, we consider the edge is connected by using the fact below.

$$\begin{aligned} &\frac{\pi_1^{in} f_1(z^{in})}{\pi_0^{in} f_0(z)} = (1 - fdr^{in}(z^{in}))/fdr^{in}(z^{in}) \geqslant T \\ &\Rightarrow \frac{f_1(z^{in})}{f_0(z^{in})} \geqslant T\frac{\pi_0^{in}}{\pi_1^{in}} \end{aligned} \tag{3}$$

The above cut-off can be linked to $fdr^{all}(z^{in})$ by using the fact that

$$\begin{aligned} &\frac{f_1(z^{in})}{f_0(z^{in})} \geqslant T\frac{\pi_0^{in}}{\pi_1^{in}} \\ \Rightarrow &\frac{f_1(z^{in})}{f_0(z^{in})}\frac{\pi_1^{all}}{\pi_0^{all}} = (1 - fdr^{all}(z^{in}))/fdr^{all}(z^{in}) \geqslant T\frac{\pi_0^{in}}{\pi_1^{in}}\frac{\pi_1^{all}}{\pi_0^{all}} \\ \Rightarrow &fdr^{all}(z^{in}) = \frac{1}{T\frac{\pi_0^{in}}{\pi_1^{in}}\frac{\pi_1^{all}}{\pi_0^{all}} + 1} \end{aligned} \tag{4}$$

For example, if $T = 4, \hat{\pi}_0^{in}/\hat{\pi}_1^{in} = 0.1$ and $\hat{\pi}_1^{all}/\hat{\pi}_0^{all} = 0.1$, then $fdr^{all}(z^{in})$=0.96 and many edges inside networks are estimated as connected because the distribution of inside network edges suggests that most edges inside are correlated.

Similarly, for edges outside networks

$$fdr^{all}(z^{out}) = \frac{1}{T\frac{\pi_0^{out}}{\pi_1^{out}}\frac{\pi_1^{all}}{\pi_0^{all}} + 1}.$$

Given $T = 4, \hat{\pi}_0^{out}/\hat{\pi}_1^{out} = 40$ and $\hat{\pi}_1^{all}/\hat{\pi}_0^{all} = 0.1$, then $fdr^{all}(out)$=0.06 and most edges outside the network are thresholded by using a more stringent cut-off.

In practice, we use the thresholds of $1/(T\hat{\theta}^{in}/\hat{\theta}^{all} + 1)$ and $1/(T\hat{\theta}^{out}\hat{\theta}^{all} + 1)$ for edges inside and outside networks based on the overall local *fdr* value.

By using these overall fdr based thresholds, the computation is more efficient. In addition, we can observe how topological structure and edge distributions can jointly impact the decision making process of the correlation matrix thresholding. When the data shows no network structure, for instance, $\pi_0^{in}/\pi_1^{in} = \pi_0^{out}/\pi_1^{out} = 10$ and $\pi_1^{all}/\pi_0^{all} = 10$, then $fdr^{all}(in) = fdr^{all}(out) = 0.2$. The network guided thresholding rule becomes the universal thresholding rule.

NICE algorithm

The detailed NICE algorithm is described in Algorithm 1.

Algorithm 1 NICE algorithm

```
procedure NICE–ALGORITHM
    The weight matrix W = g(R̂);
    Calculate the Laplacian matrix L=D-W
    for cluster number C = 2 : |V| − 1 do
        Compute the first C eigenvectors [u_2, ··· , u_C] of L, with eigenvalues ranked from the smallest;
        Let U = [u_2^T, ··· , u_C^T] be a |V| × C matrix containing all C − 1 eigenvectors;
        Perform K-means clustering algorithm on U with number of clusters of C to cluster |V| nodes into C networks;
        Calculate criterion 2 for each C.
    end for
    Select C that maximizes criterion 2 and perform permutation tests on the large community networks.
    Implement the topological structure oriented thresholding strategies for covariance entries inside and outside networks (see details in 2.2)
end procedure
```

Proof of Theorem 1

Proof. Applying the universal decision rule with z_0 as threshold:

$$E(\#FP) = m\int_{z_0}^{\infty} \frac{\pi_0 f_0(z)}{f(z)} f(z)dz = m\pi_0 F_0(z_0) = m\omega\pi_0^{in}F_0(z_0) + m(1-\omega)\pi_0^{out}F_0(z_0) \quad (5)$$

where $\int_{z_0}^{\infty} f = F(z_0)$ and m is the total number of edges $m = |E|$.

For edges in communities:

$$E^{in}(\#FP) = \omega m \int_{z_{in}}^{\infty} \frac{\pi_0^{in} f_0^{in}(z)}{f^{in}(z)} f^{in}(z)dz = \omega m \pi_0^{in} F_0^{in}(z_{in}) \tag{6}$$

For edges outside communities:

$$E^{out}(\#FP) = (1-\omega)m \int_{z_{0,out}}^{\infty} \frac{\pi_0^{out} f_0^{out}(z)}{f^{out}(z)} f^{out}(z)dz = (1-\omega)m\pi_0^{out} F_0^{out}(z_{0,out}) \tag{7}$$

where $z_{0,in} < z_0 < z_{0,out}$, and $F_0^{out}(z) = F_0^{in}(z) = F_0(z)$. There we expect $E(\#FP)$ (5)$>$ $E^{in}(\#FP) + E^{out}(\#FP)$ (6 + 7) if

$$\begin{aligned} &-\omega m \pi_0^{in}(F_0(z_{in}) - F_0(z_0)) + (1-\omega)m\pi_0^{out}(F_0(z_0) - F_0(z_{0,out})) > 0, \\ &\Leftrightarrow \frac{F_0(z_0) - F_0(z_{0,out})}{F_0(z_{0,in}) - F_0(z_0)} > \frac{\omega \pi_0^{in}}{(1-\omega)\pi_0^{out}} \end{aligned} \tag{8}$$

We further calculate the expected number of true positive (TP) edges using universal threshold and in/out communities to evaluate the power of our network based thresholding.

Applying the universal decision rule with z_0 as threshold:

$$E(\#TP) = m \int_{z_0}^{\infty} \frac{\pi_1 f_1(z)}{f(z)} f(z)dz = m\pi_1 F_0(z_0) = m\omega\pi_1^{in} F_1(z_0) + m(1-\omega)\pi_1^{out} F_1(z_0) \tag{9}$$

For edges in communities:

$$E^{in}(\#\ TP) = \omega m \int_{z_{in}}^{\infty} \frac{\pi_1^{in} f_1^{in}(z)}{f^{in}(z)} f^{in}(z)dz = \omega m \pi_1^{in} F_1^{in}(z_{in}) \tag{10}$$

For edges outside communities:

$$E^{out}(\#\ TP) = (1-\omega)m \int_{z_{out}}^{\infty} \frac{\pi_1^{out} f_1^{out}(z)}{f^{out}(z)} f^{out}(z)dz = (1-\omega)m\pi_1^{out} F_1^{out}(z_{out}) \tag{11}$$

where $z_{0,in} < z_1 < z_{0,out}$, and $F_1^{out}(z) = F_1^{in}(z) = F_1(z)$. There we expect $E(\#TP)$ (9)$<$ $E^{in}(\#TP) + E^{out}(\#TP)$ (10 + 11) (i.e. $E(\#FN) > E^{in}(\#FN) + E^{out}(\#FN)$) if

$$
\begin{aligned}
&-\omega m\pi_0^{in}(F_1(z_{in}) - F_1(z_0)) + (1-\omega)m\pi_1^{out}(F_1(z_0) - F_1(z_{0,out})) < 0, \\
&\Leftrightarrow \frac{F_1(z_0) - F_1(z_{0,out})}{F_1(z_{in}) - F_1(z_0)} < \frac{\omega\pi_1^{in}}{(1-\omega)\pi_1^{out}}
\end{aligned}
\tag{12}
$$

Condition 1 is generally true and permutation tests ensure the communities have large proportions of highly correlated edges. Numerous empirical experiments and results further verify this assumption. If the assumption of network induced correlation matrix is true, the C selection procedure chooses the parameter to optimize the objective function of step one that reduce false positive findings and improve power simultaneously.

[Figure 1 about here.]

References

Chen, S., Kang, J., Wang, G. (2015). An empirical Bayes normalization method for connectivity metrics in resting state fMRI. *Frontiers in neuroscience*, **9**, 316-323.

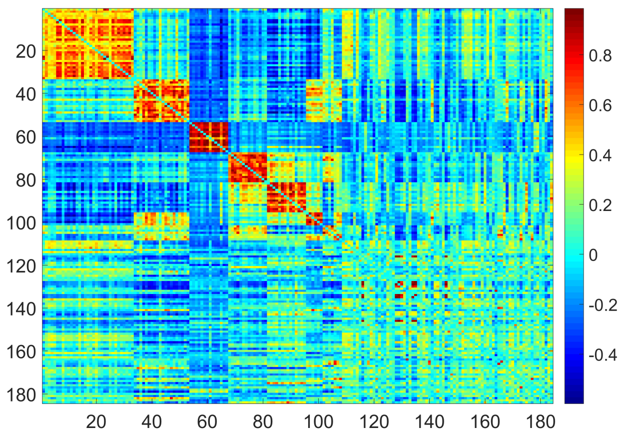
Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association*, **99**, 96-104.

Efron, B. (2007). Size, power and false discovery rates. *The Annals of Statistics*, **35**(4), 1351-1377.

Wu, B., Guan, Z., Zhao, H. (2006). Parametric and nonparametric FDR estimation revisited. *Biometrics*, **62**(3), 735-744.

Figure 1: *Glasso* results for Example Data 1: it shows that *Glasso* may false negatively regularize edges to zero in networks (with the sparsity assumption).

(a) Correlation heatmap in the order of detected communities



(b) *glasso* results