Community Detection in Partial Correlation Network Models Online Appendix

Christian Brownlees[†] Guðmundur Stefán Guðmundsson[†] Gábor Lugosi[†]

April 1, 2020

1 Additional Simulation Results

1.1 Blockbuster Illustration

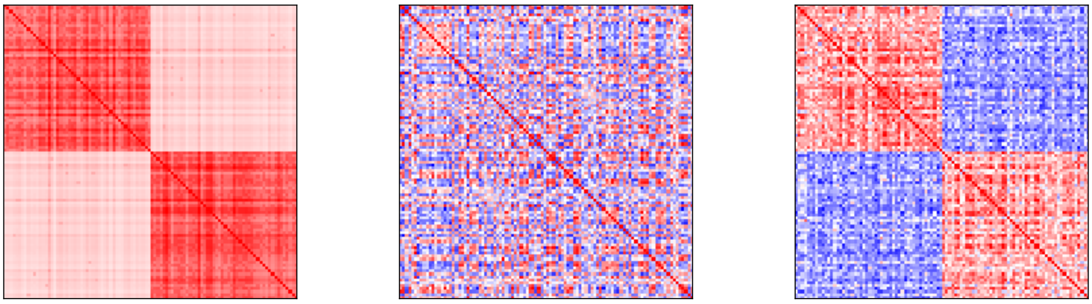
This section contains an illustration of the Blockbuster algorithm on simulated data. We draw a single sample from the SBPCM with $R = 1$ common factors and $k = 2$ communities for $T = 500$ and $n = 100$. We draw a GSBM where each community has size n/k. The edge probabilities are set to $p_s = p = 0.5$ for all $s = 1, \ldots, k$ and $q_{vr} = q = 0.01$ for all $v, r = 1, \ldots, k$, $v \neq r$. The network-dependence parameter ϕ is set to 20 while the network variance σ^2 is 1. We draw $[\boldsymbol{\Theta}]_{ii}$ from a power law distribution $f(x) = \alpha x_m^\alpha / x^{\alpha+1}$ for $x \geq x_m$ with $x_m = 0.75$ and $\alpha = 2.5$. The edge-weights are drawn uniformly in the interval $[0.3, 1]$. We draw the data identically and independently from a multivariate Gaussian with covariance matrix given as in (9), where the factor loadings $\boldsymbol{q}$ are generated from a standard normal.

The first panel of Figure OA-1 displays a heatmap of the correlation matrix of the panel conditional on the factor, with the series ordered by the true community partition. The second and third panels show the corresponding sample correlation matrix when the series are randomly shuffled and when they are ordered by the estimated community

e-mail: christian.brownlees@upf.edu; gsgudmundsson@econ.au.dk; gabor.lugosi@upf.edu.

partition. The figure shows that the Blockbuster algorithm detects the communities accurately. In particular, in this specific instance there are no misclustered series.

Figure OA-1: Heatmaps of simulated data



The first panel displays a heatmap of the correlation matrix (conditional on the factor) with the series ordered by the true community partition. The second and third panels show the sample correlation matrix when the series are randomly shuffled and ordered by the estimated community partition. The data are simulated with $n = 100$, $T = 500$ and $k = 2$.

1.2 Blockbuster Performance and Pre-filtering

The Blockbuster algorithm is designed for stationary strongly mixing data and we have established performance bounds for it that do not depend on specific assumptions on the time series model that may have generated the data. However, we point out that performance improvements may be attainable if one exploits knowledge of the specific time series structure of the data.

In this section we consider the same DGP as in the simulation exercise of Section 3. Instead of carrying out community detection by applying the Blockbuster algorithm to the raw time series as in Section 3, we consider a modified procedure. We estimate univariate AR(1) models for each series by OLS and then apply Blockbuster to the panel of AR(1) residuals.

Table OA-1: Hit Ratio of Blockbuster: Pre-filtering

	Panel A: $\phi = 5$					Panel B: $\phi = 50$				
	$T = 50$	100	200	500	1 000	50	100	200	500	1 000
p/q	$n = 50$					$n = 50$				
0.25/0.01	64.0%	81.4%	86.1%	95.2%	95.7%	81.5%	91.8%	92.3%	94.0%	95.1%
0.50/0.01	78.0%	96.1%	99.5%	99.9%	100.0%	97.0%	99.6%	99.7%	100.0%	100.0%
0.25/0.05	41.4%	44.6%	49.0%	57.9%	61.8%	42.2%	48.8%	53.9%	61.1%	64.5%
	$n = 100$					$n = 100$				
0.25/0.01	48.6%	64.1%	83.7%	93.4%	96.1%	68.7%	85.0%	94.8%	94.2%	95.1%
0.50/0.01	60.5%	84.2%	97.8%	98.6%	99.9%	89.9%	98.3%	99.7%	99.9%	99.9%
0.25/0.05	33.5%	35.9%	39.2%	47.2%	52.9%	34.3%	36.5%	41.2%	54.3%	59.5%
	$n = 200$					$n = 200$				
0.25/0.01	37.1%	47.2%	66.2%	89.1%	94.5%	49.1%	71.8%	85.7%	94.4%	95.9%
0.50/0.01	44.8%	62.0%	89.9%	98.0%	99.7%	70.7%	90.6%	98.4%	99.8%	99.9%
0.25/0.05	28.9%	30.0%	31.2%	35.5%	43.6%	29.4%	30.3%	33.3%	39.5%	49.4%

We report the results of this exercise in Table OA-1. These results can be compared to the ones in Table 1 in the main text. The table shows that applying Blockbuster to the OLS residuals systematically improves performance in the vast majority of cases, although the improvements are typically minor. Overall, the exercise shows that incorporating finer time-series information into the block detection procedure may enhance the performance of the algorithm.

1.3 Blockbuster Performance and Weakly Correlated Factors

In order to establish the consistency of the Blockbuster algorithm in the presence of factors in Section 2.4, we assume that the low-rank structure induced by the factors is orthogonal to the one generated by the communities. This assumption is made on the grounds of simplicity and may be violated in practice. In particular, it may be reasonable to assume that there exits some weak form of correlation between the factors and the communities. In this section we explore how the presence of weak correlation between factors and communities affects clustering performance. We consider the same DGP as in the simulation exercise of Section 3 with the exception that in this exercise the factor loadings of the model depend on the community membership indicator. In particular, the factor loading of series i is drawn from a normal whose variance depends on the community of series i. Specifically, the variances of the loadings of the series in community 1, 2, 3, 4 and 5 are respectively 0.500, 0.875, 1.250, 1.625, and 2.000.

Table OA-2: Hit Ratio of Blockbuster: Weakly Correlated Factors

	Panel A: $\phi = 5$					Panel B: $\phi = 50$				
	$T = 50$	100	200	500	1 000	50	100	200	500	1 000
p/q	$n = 50$					$n = 50$				
0.25/0.01	64.8%	80.6%	88.6%	94.6%	93.7%	84.5%	90.8%	94.3%	93.6%	95.2%
0.50/0.01	82.5%	94.8%	96.5%	99.7%	99.8%	97.9%	99.5%	99.8%	99.6%	100.0%
0.25/0.05	41.2%	46.3%	50.9%	57.2%	60.8%	43.8%	50.9%	54.2%	61.4%	63.9%
	$n = 100$					$n = 100$				
0.25/0.01	49.4%	64.8%	83.3%	94.3%	95.3%	74.3%	87.7%	93.0%	96.1%	95.1%
0.50/0.01	62.9%	86.2%	96.2%	99.6%	99.8%	93.3%	99.0%	99.6%	99.9%	99.9%
0.25/0.05	34.4%	35.5%	40.0%	47.4%	54.4%	37.1%	38.9%	45.6%	53.4%	59.7%
	$n = 200$					$n = 200$				
0.25/0.01	37.1%	47.0%	66.0%	89.0%	94.3%	57.3%	78.2%	89.7%	96.0%	96.5%
0.50/0.01	43.0%	66.0%	88.0%	98.6%	99.6%	85.5%	95.5%	99.1%	99.8%	99.9%
0.25/0.05	29.2%	30.4%	33.0%	36.6%	43.6%	30.8%	32.6%	34.9%	40.7%	50.8%

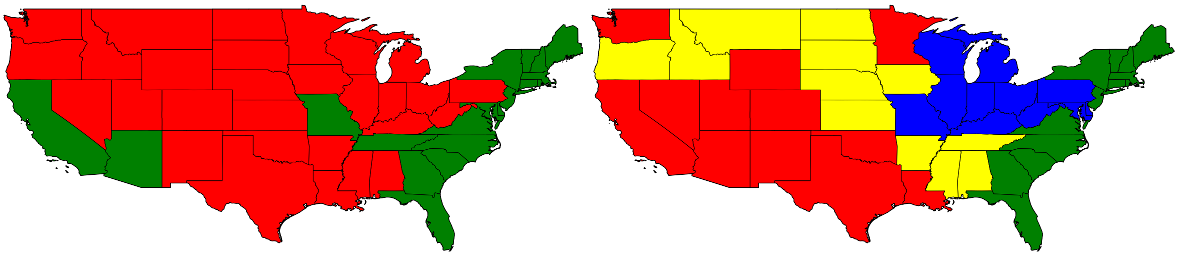
We report the results of this exercise in Table OA-2. The results of this table can be compared to the ones in Table 1 in the paper. The table shows that in case of weak correlation between the factors and the communities the performance of Blockbuster deteriorates, but the losses are overall fairly small.

2 Additional Empirical Results

2.1 Estimation Results for Alternative Number of Communities

We run the Blockbuster algorithm with the number of communities set to two and four. The results are reported in Figure OA-2. When the number of communities is set to two, the algorithm partitions the U.S. into East Coast states together with California, Arizona, Missouri and Tennessee, and a residual cluster containing all remaining states. When the number of communities is set to four, in comparison to the baseline case, the oil-producing and agricultural states community gets split into into two separate clusters and California and Arizona are absorbed into the cluster containing oil-producing states. Note that the community corresponding to the East Coast is relatively stable across different choices of the number of clusters.

Figure OA-2: U.S. REAL ACTIVITY CLUSTERING ($k = 2$ AND 4)

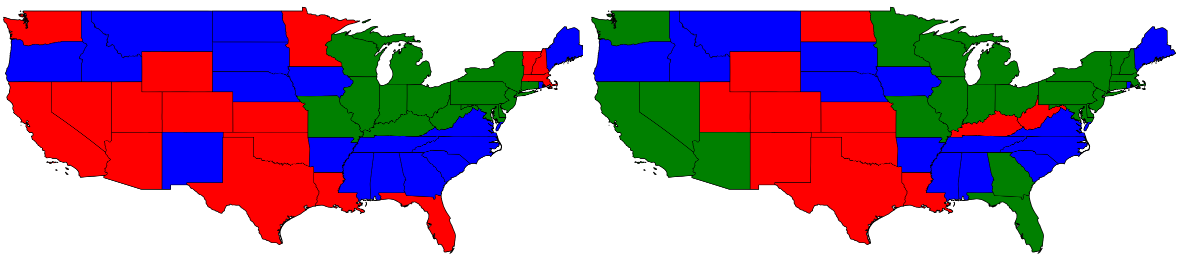


The figure displays the communities detected by Blockbuster when the number of communities is 2 (left) and 4 (right).

2.2 Robustness of the Community Detection Results over Different Sub-samples

In this section we analyse the stability of the community detection results over different sub-samples. In particular, we apply Blockbuster to the sub-sample that starts at the beginning of the sample and ends at observation $\lfloor T/2 + s(T/8) \rfloor$ for $s = 0, 1, 2, 3$. We assume that the number of communities k is 3 and that the data has one factor (R is 1). Setting $s = 2$ or $s = 3$ yields identical results to the ones based on the entire sample. Figure OA-3 shows the results for $s = 0$ and $s = 1$. As one may expect, for low values of s the community assignments differ slightly, however, overall the discrepancies are moderate.

Figure OA-3: Stability of Community Detection Results

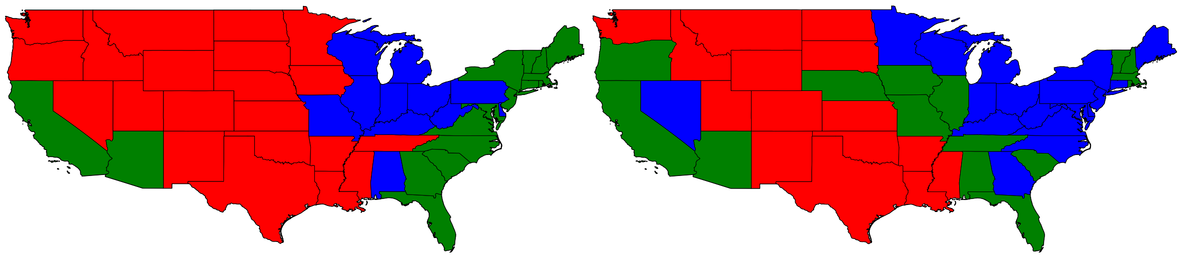


$s = 0$ $\qquad$ $s = 1$

2.3 Community Detection based on the GLASSO algorithm

In this section we compare the Blockbuster community detection results to the ones obtained by applying spectral clustering to the estimate of the Laplacian of the partial correlation network obtained by the GLASSO algorithm. In particular, we estimate the partial correlation network of the panel via the GLASSO algorithm after filtering out the first factor (estimated by PCA) from the data. We choose the shrinkage parameter of the GLASSO via cross-validation. Finally, we apply the classic spectral clustering algorithm to the Laplacian of the estimated partial correlation network. Figure (OA-4) compares the estimated community partition of the US states when the number of communities is set to 3 by Blockbuster (left) and GLASSO (right). Overall, the results are close to what is obtained by applying the Blockbuster algorithm to the sample covariance matrix.

Figure OA-4: Community Detection Comparison



Blockbuster on Sample Covariance Matrix

Spectral Clustering on Laplacian of Partial Correlation Network (GLASSO)

2.4 Alternative Out-of-Sample Loss Function Results

In this section we report the results of the forecasting exercise based on two alternative loss functions. In particular, we have carried out the same forecasting exercise described in Section 4 with the only modification that out-of-sample forecasts are now evaluated on the basis of the Square loss, a popular loss function for covariance matrices (Pourahmadi, 2013, Section 2.2), and the Quasi-likelihood loss function, a loss function proposed by Andrew Patton (Patton and Sheppard, 2009; Patton, 2011). The two losses are respectively

Table OA-3: U.S. Real Activity Forecasting: Square and Quasi-Likelihood Losses

k	Square Loss			Quasi-Likelihood Loss		
	SCM	LW	POET	SCM	LW	POET
2	3.710%	23.682%	1.641%	9.476%	1.137%	5.383%
3	−3.406%	18.041%	−3.024%	9.083%	0.708%	6.977%
4	−4.203%	17.409%	−4.102%	9.013%	0.631%	9.375%

The table reports the relative gain of the Blockbuster covariance estimator over the sample covariance estimator, the POET estimator and the (linear) Ledoit and Wolf shrinkage estimator in terms of the Square and Quasi-Likelihood losses for different choices of the number of communities k.

defined as

$$L_{SL} = \text{tr}(\mathbf{\Sigma}_B^{-1}\widehat{\mathbf{\Sigma}}_C - \mathbf{I}_n)^2,$$

$$L_{QL} = \log|\widehat{\mathbf{\Sigma}}_C^{-1}| + \text{tr}(\widehat{\mathbf{\Sigma}}_C^{-1}\mathbf{\Sigma}_B).$$

The results are reported in Table OA-3. In the case of the Square loss, the table shows that for $k = 2$ the Blockbuster procedure achieves the best performance and improves upon all other benchmarks. For $k > 2$ the evidence is mixed. In particular, while Blockbuster appears to always be able to improve over the linear Ledoit and Wolf estimator, however it performs worse than the sample covariance and POET benchmarks. For the Quasi-Likelihood loss, the table shows that the Blockbuster procedure performs better than the other benchmarks and in particular it performs better than the sample covariance matrix.

2.5 Additional Tables

Table OA-4: U.S. Real Activity Correlation Matrix Block Averages ($k = 3$)

Community	Red	Blue	Green
Red	0.143	−0.078	−0.107
Blue	−0.078	0.207	−0.076
Green	−0.107	−0.076	0.222

The table shows the averages of the elements of the correlation matrix (conditional on the factor), within the estimated community blocks on the diagonal and within the across-community blocks on the off-diagonal, for $k = 3$.

3 Population Lemma for the Common Factor Model

Lemma OA-1. *Let* $\mathcal{G} \sim \text{GSBM}(\mathbf{Z}, \mathbf{B}, \mathbf{\Theta}, \mathbf{W})$ *be a Generalised Stochastic Block Model as in* Definition 1 *and* $\mathcal{K}_\epsilon$ *its population precision matrix. Let* $\mathcal{K}^{-1} = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^{R} \boldsymbol{q}_r \boldsymbol{q}_r'$ *be the population analogue of* (9), $\mathcal{U}$ *the matrix of its* $(R+1)$*-th to* $(R+k)$*-th bottom eigenvectors and* $\mathcal{X}$ *the row-normalised counterpart of* $\mathcal{U}$.

Then $\lambda_i(\mathcal{K}) = 1/\left(\sigma^2/(1+\phi) + \|\boldsymbol{q}_i\|^2\right)$ *for* $i = 1, \ldots R$, $\lambda_i(\mathcal{K}) \in [1/\sigma^2, (1+\phi)/\sigma^2)$ *for* $i = R+1, \ldots, R+k$ *and* $\lambda_i(\mathcal{K}) = (1+\phi)/\sigma^2$ *for all* $i = R+k+1, \ldots, n$*. Furthermore, there exists a* $k \times k$ *orthonormal matrix* $\mathbf{V}$ *such that* $\mathcal{X} = \mathbf{Z}\mathbf{V}$.

Proof. We first find the eigenvalues and eigenvectors of $\mathcal{K}^{-1} = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^{R} \boldsymbol{q}_r \boldsymbol{q}_r'$. We then show that $\mathcal{U} = \mathcal{U}_\epsilon$, which allows us to apply Lemma 1 and finish the proof.

We proceed by induction and consider $R = 1$ with $\mathcal{K}_1^{-1} = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \boldsymbol{q}_1 \boldsymbol{q}_1'$ first. We then find the eigenvectors of $\mathcal{K}^{-1} \equiv \mathcal{K}_R^{-1} = \mathcal{K}_{R-1}^{-1} + \boldsymbol{q}_R \boldsymbol{q}_R'$ given those of $\mathcal{K}_{R-1}^{-1}$. The eigenvectors $\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i = 1, \ldots, n$ form a basis in $\mathbb{R}^n$, so we may write

$$\boldsymbol{q}_1 = \sum_{i=1}^{n} \gamma_{i,1} \boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}),$$

where $\gamma_{i,1}$ are scalars. By Assumption 3, we have $\boldsymbol{q}_1'[\mathcal{U}_\epsilon]_{\bullet i} = 0$ which implies $\gamma_{i,1} = 0$ for $i = 1, \ldots, k$, so that $\boldsymbol{q}_1 = \gamma_{k+1,1} \boldsymbol{u}_{k+1}(\mathcal{K}_{\boldsymbol{\epsilon}}) + \ldots + \gamma_{n,1} \boldsymbol{u}_n(\mathcal{K}_{\boldsymbol{\epsilon}})$.

We guess and verify the eigenvectors and eigenvalues of $\mathcal{K}_1$ using the eigenvalue equation $\mathcal{K}_1^{-1} \boldsymbol{u}_i(\mathcal{K}_1) = \lambda_i(\mathcal{K}_1)^{-1} \boldsymbol{u}_i(\mathcal{K}_1)$ and ensure they are mutually orthogonal. We begin with the bottom eigenvalue and eigenvector of $\mathcal{K}_1$ and guess $\lambda_1(\mathcal{K}_1)^{-1} = \lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} + \|\boldsymbol{q}_1\|^2$ and $\boldsymbol{u}_1(\mathcal{K}_1) = \boldsymbol{q}_1/\|\boldsymbol{q}_1\|$. Then $\mathcal{K}_1^{-1} \boldsymbol{q}_1 = (\mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \boldsymbol{q}_1 \boldsymbol{q}_1') \boldsymbol{q}_1 = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{q}_1 + \|\boldsymbol{q}_1\|^2 \boldsymbol{q}_1 = \left(\lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} + \|\boldsymbol{q}_1\|^2\right) \boldsymbol{q}_1$ where the last equality follows from

$$\mathcal{K}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{q}_1 = \sum_{i=k+1}^{n} \gamma_{i,1} \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} \boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \sum_{i=k+1}^{n} \gamma_{i,1} \lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} \boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} \boldsymbol{q}_1,$$

as $\lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i = k+1, \ldots, n$ from Lemma 1. Dividing through by $\|\boldsymbol{q}_1\|$ we have the eigenvector. Next we consider the second to the $k+1$-th bottom eigenvalues and eigenvectors of $\mathcal{K}_1$. We postulate $\lambda_{1+i}(\mathcal{K}_1)^{-1} = \lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1}$ and $\boldsymbol{u}_{1+i}(\mathcal{K}_1) = \boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i =$

$1,\ldots,k$. Consider $\mathcal{K}_1^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) + \boldsymbol{q}_1\boldsymbol{q}_1'\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i = 1,\ldots,k$, which follows from the fact that the bottom k eigenvectors of $\mathcal{K}_{\boldsymbol{\epsilon}}^{-1}$ are orthogonal to $\boldsymbol{q}_1$ and shows that those are the desired eigenvalues and eigenvectors.

It remains to find the last $n-k-1$ eigenvectors of $\mathcal{K}_1$. Let $\boldsymbol{u}_n(\mathcal{K}_1) = \gamma_{n-1,1}\boldsymbol{u}_n(\mathcal{K}_{\boldsymbol{\epsilon}}) - \gamma_{n,1}\boldsymbol{u}_{n-1}(\mathcal{K}_{\boldsymbol{\epsilon}})$ and

$$\boldsymbol{u}_{n-i}(\mathcal{K}_1) = \gamma_{n,1}\boldsymbol{u}_n(\mathcal{K}_{\boldsymbol{\epsilon}}) + \gamma_{n-1,1}\boldsymbol{u}_{n-1}(\mathcal{K}_{\boldsymbol{\epsilon}}) + \ldots + \gamma_{n-i,1}\boldsymbol{u}_{n-i}(\mathcal{K}_{\boldsymbol{\epsilon}}) - \left(\frac{\sum_{j=n-i}^{n}\gamma_{j,1}^2}{\gamma_{n-i-1,1}}\right)\boldsymbol{u}_{n-i-1}(\mathcal{K}_{\boldsymbol{\epsilon}}),$$

for all $i = 1,\ldots n-k-2$. These vectors are orthogonal to $\boldsymbol{q}_1$, as the last term always cancels out all the others. They are also orthogonal to each other by a similar argument. Returning to the eigenvalue equation, we have $\mathcal{K}_1^{-1}\boldsymbol{u}_i(\mathcal{K}_1) = \left(\mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \boldsymbol{q}_1\boldsymbol{q}_1'\right)\boldsymbol{u}_i(\mathcal{K}_1) = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{u}_i(\mathcal{K}_1) = \lambda_n\left(\mathcal{K}_{\boldsymbol{\epsilon}}\right)^{-1}\boldsymbol{u}_i(\mathcal{K}_1)$ for $i = k+2,\ldots,n$. Dividing through by the norm delivers the remaining eigenvectors of $\mathcal{K}_1$, all with eigenvalue $\lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1}$.

Now assume that we know the eigenvectors $\boldsymbol{u}_i(\mathcal{K}_{R-1})$ of $\mathcal{K}_{R-1}$. We look for the eigenvectors of $\mathcal{K}^{-1} \equiv \mathcal{K}_R^{-1} = \mathcal{K}_{R-1}^{-1} + \boldsymbol{q}_R\boldsymbol{q}_R'$ and note that $\mathcal{K}_{R-1}^{-1} = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^{R-1}\boldsymbol{q}_r\boldsymbol{q}_r'$. Proceeding as before, we begin with the bottom R eigenvalues and eigenvectors of $\mathcal{K}$ and guess $\lambda_i(\mathcal{K})^{-1} = \lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} + \|\boldsymbol{q}_i\|^2$ and $\boldsymbol{u}_1(\mathcal{K}) = \boldsymbol{q}_i/\|\boldsymbol{q}_i\|$ for $i = 1,\ldots R$. We have

$$\mathcal{K}^{-1}\boldsymbol{q}_i = \left(\mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^{R}\boldsymbol{q}_r\boldsymbol{q}_r'\right)\boldsymbol{q}_i = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{q}_i + \|\boldsymbol{q}_i\|^2\boldsymbol{q}_i = \left(\lambda_n(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1} + \|\boldsymbol{q}_i\|^2\right)\boldsymbol{q}_i,$$

for all $i = 1,\ldots,R$ as $\boldsymbol{q}_i'\boldsymbol{q}_v = 0$ for all $v = 1,\ldots,R$, $v \neq i$. We next postulate $\lambda_{R+i}(\mathcal{K})^{-1} = \lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1}$ and $\boldsymbol{u}_{R+i}(\mathcal{K}) = \boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i = 1,\ldots,k$. As before

$$\mathcal{K}^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) + \sum_{r=1}^{R}\boldsymbol{q}_r\boldsymbol{q}_r'\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}) = \lambda_i(\mathcal{K}_{\boldsymbol{\epsilon}})^{-1}\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}}),$$

for $i = 1,\ldots,k$. Finally, as $\mathcal{K}$ is a rank-one update of $\mathcal{K}_{R-1}$ we may apply a similar logic to before: Notice that $\boldsymbol{u}_i(\mathcal{K}_{R-1})$ form a basis for $\mathbb{R}^n$ and write

$$\boldsymbol{q}_R = \sum_{i=R+k}^{n}\gamma_{i,R}\boldsymbol{u}_i(\mathcal{K}_{R-1}).$$

Then let $\boldsymbol{u}_n(\mathcal{K}) = \gamma_{n-1,R}\boldsymbol{u}_n(\mathcal{K}_{R-1}) - \gamma_{n,R}\boldsymbol{u}_{n-1}(\mathcal{K}_{R-1})$ and

$$\boldsymbol{u}_{n-i}(\mathcal{K}) = \gamma_{n,R}\boldsymbol{u}_n(\mathcal{K}_{R-1}) + \ldots + \gamma_{n-i,R}\boldsymbol{u}_{n-i}(\mathcal{K}_{R-1}) - \left(\frac{\sum_{j=n-i}^{n}\gamma_{j,R}^2}{\gamma_{n-i-1,R}}\right)\boldsymbol{u}_{n-i-1}(\mathcal{K}_{R-1}),$$

for all $i = 1, \ldots n-k-R-1$. We then have $\mathcal{K}^{-1}\boldsymbol{u}_i(\mathcal{K}) = \left(\mathcal{K}_{R-1}^{-1} + \boldsymbol{q}_R\boldsymbol{q}_R'\right)\boldsymbol{u}_i(\mathcal{K}) = \mathcal{K}_{R-1}^{-1}\boldsymbol{u}_i(\mathcal{K}) = \lambda_n\left(\mathcal{K}_{R-1}\right)^{-1}\boldsymbol{u}_i(\mathcal{K})$ for all $i = 1, \ldots n-k-R-1$. Notice that by induction $\lambda_n\left(\mathcal{K}_{R-1}\right)^{-1} = \lambda_n\left(\mathcal{K}_{\boldsymbol{\epsilon}}\right)^{-1}$. After normalising, we have the last $n-k-R$ eigenvectors, all with eigenvalue $\lambda_n\left(\mathcal{K}_{\boldsymbol{\epsilon}}\right)^{-1}$. We have thus shown that $\mathcal{U} = \mathcal{U}_\epsilon$ and may apply Lemma 1 to finish the proof. □

4 Auxiliary Results

This appendix contains some results that are used in Section 2. In Theorem OA-1 we extend the random graph concentration result of Theorem 3.1 in Oliveira (2009) to allow for weighted graphs.

Theorem OA-1. *Consider a random undirected and weighted graph on n vertices $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ where $(i,j) \in \mathcal{E}$ with probability $p_{i,j} = p_{j,i}$ independently and the edge weights $W_{ij} \in \mathcal{W}$ are independent random variables supported on the interval $[\alpha_{ij}, \beta_{ij}]$ with $\beta_{ij} > 0$, mean $\mu > 0$ and variance σ_{ij}^2 for all i,j. Let $\mathbf{L}$ be the Laplacian matrix corresponding to $\mathcal{G}$ and $\mathcal{L}$ be its population analogue. Let $\bar{d}_{min} = \min_i \bar{d}_i$ be the minimum expected degree of $\mathcal{G}$.*

If $\bar{d}_{min} = \Omega\left(\log(n)\right)$, then

$$\|\mathbf{L} - \mathcal{L}\| = O\left(\sqrt{\frac{\log(n)}{\bar{d}_{min}}}\right),$$

with high probability.

Proof. We follow Oliveira (2009). We begin by controlling the following sum of independent zero-mean random variables

$$d_i - \bar{d}_i = \sum_{j=1}^{n} \left(B_{ij} W_{ij} - p_{ij}\mu_{ij} \right).$$

Notice that $|B_{ij}W_{ij} - p_{ij}\mu_{ij}|$ takes its maximum at one of the following: $p_{ij}\mu_{ij}$, $|\alpha_{ij}|+p_{ij}\mu_{ij}$ or $\beta_{ij} - p_{ij}\mu_{ij}$. Hence

$$B_{ij}W_{ij} - p_{ij}\mu_{ij} \leq |\alpha_{ij}| + \beta_{ij} \leq \max_{ij} |\alpha_{ij}| + \max_{ij} \beta_{ij}.$$

as the the eigenvalues of $\mathbf{A}_{ij}$ are in $\{-1, 0, 1\}$. Next consider the variance statistic

$$\sum_{j=1}^{n} \mathbb{E}\left[(B_{ij}W_{ij} - p_{ij}\mu_{ij})^2 \right] = \sum_{j=1}^{n} \mathrm{Var}\left(B_{ij}W_{ij} \right).$$

We require the variance of the composite random variable $B_{ij}W_{ij}$, where B_{ij} and W_{ij} are independent. We have

$$\mathrm{Var}\left(B_{ij}W_{ij}\right) = \mathrm{Var}\left(B_{ij}\right) \mathbb{E}\left[W_{ij}\right]^2 + \mathrm{Var}\left(W_{ij}\right) \mathbb{E}\left[B_{ij}\right]^2 + \mathrm{Var}\left(B_{ij}\right) \mathrm{Var}\left(W_{ij}\right)$$

$$= p_{ij}(1-p_{ij})\mu_{ij}^2 + \sigma_{ij}^2 p_{ij}^2 + p_{ij}(1-p_{ij})\sigma_{ij}^2 = p_{ij}(1-p_{ij})\mu_{ij}^2 + p_{ij}\sigma_{ij}^2$$

$$\leq p_{ij}\left(\mu_{ij}^2 + \sigma_{ij}^2\right) = p_{ij}\mu_{ij}\left(\mu_{ij} + \frac{\sigma_{ij}^2}{\mu_{ij}}\right) \leq p_{ij}\mu_{ij}\left(\max_{ij} \mu_{ij} + \max_{ij} \sigma_{ij}^2\right),$$

using $\mathbb{E}\left[B_{ij}\right] = p_{ij}$ and $\mathrm{Var}\left(B_{ij}\right) = p_{ij}(1-p_{ij})$. It follows that

$$\sum_{j=1}^{n} \mathrm{Var}\left(B_{ij}W_{ij}\right) \leq \left(\max_{ij} \mu_{ij} + \max_{ij} \sigma_{ij}^2\right) \bar{d}_i,$$

as $\bar{d}_i = \sum_j p_{ij}\mu_{ij}$.

Define $\nu = \max_{ij} \mu_{ij} + \max_{ij} \sigma_{ij}^2 + \max_{ij} |\alpha_{ij}| + \max_{ij} \beta_{ij}$, fix $c > 0$ and assume $n^{-c} \leq \delta \leq 1/2$. We may then apply Corollary 7.1. from Oliveira (2009) with $n = 1$ to obtain for all $r > 0$ and all i

$$\mathbb{P}\left(\left|d_i - \bar{d}_i\right| \geq r\right) = \mathbb{P}\left(\left|\frac{d_i}{\bar{d}_i} - 1\right| \geq \frac{r}{\bar{d}_i}\right) = \mathbb{P}\left(\left|\frac{d_i}{\bar{d}_i} - 1\right| \geq t\right) \leq 2e^{-\frac{t^2\bar{d}_i^2}{8s^2+4Mt\bar{d}_i}},$$

where we took $r = t\bar{d}_i$. Using the arguments from the previous paragraph, we set $M = \nu$ and $s^2 = \nu \bar{d}_i$ and take $t = 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}} \leq 2$, where the inequality follows as $\bar{d}_{min} > C\log(n)$ and we may choose C high enough. Then

$$\mathbb{P}\left(\left|\frac{d_i}{\bar{d}_i} - 1\right| \leq 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}}\right) \geq 1 - \delta. \tag{OA-1}$$

By the mean value theorem

$$\left|\sqrt{1+x} - 1\right| = \left(\frac{1}{2\sqrt{1+\gamma}}\right)|x| \leq \max_{\gamma \in [-3/4, 3/4]} \left(\frac{1}{2\sqrt{1+\gamma}}\right)|x| = |x|,$$

with $x \in [-3/4, 3/4]$ and the last equality follows as $\gamma = -3/4$ yields the maximum. Take $x = \frac{d_i}{\bar{d}_i} - 1$ and notice that we may choose C to make $\bar{d}_{min} > C\log(n)$ large enough for (OA-1) to imply $\left|\frac{d_i}{\bar{d}_i} - 1\right| \leq 3/4$. It follows that

$$\left|\sqrt{\frac{d_i}{\bar{d}_i}} - 1\right| \leq \left|\frac{d_i}{\bar{d}_i} - 1\right| \leq 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}},$$

with probability greater than $1 - \delta$.

Define $\mathbf{T} = \mathbf{D}^{-1/2}$ and $\mathcal{T} = \mathcal{D}^{-1/2}$. Notice that $\|\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\| = \max_i \left|\sqrt{\frac{d_i}{\bar{d}_i}} - 1\right|$ as this is a diagonal matrix, which yields $\|\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\| \leq 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}}$ with probability greater than $1 - \delta$. We define the intermediate operator $\mathcal{M} = \mathbf{I}_n - \mathcal{T}\mathbf{A}\mathcal{T}$ and note that it satisfies $\mathcal{M} = \mathbf{I}_n - (\mathcal{T}\mathbf{T}^{-1})(\mathbf{I}_n - \mathbf{L})(\mathcal{T}\mathbf{T}^{-1})$ as $\mathbf{L} = \mathbf{I}_n - \mathbf{TAT}$. To compare $\mathcal{L}$ with $\mathbf{L}$, we bound the distance of each from $\mathcal{M}$. We begin with $\|\mathcal{M} - \mathbf{L}\|$ and write

$$\begin{aligned}
\|\mathcal{M} - \mathbf{L}\| &= \left\|\left(\mathcal{T}\mathbf{T}^{-1}\right)(\mathbf{I}_n - \mathbf{L})\left(\mathcal{T}\mathbf{T}^{-1}\right) + (\mathbf{I}_n - \mathbf{L})\left(\mathcal{T}\mathbf{T}^{-1}\right) - (\mathbf{I}_n - \mathbf{L})\left(\mathcal{T}\mathbf{T}^{-1}\right) - (\mathbf{I}_n - \mathbf{L})\right\| \\
&= \left\|\left(\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\right)(\mathbf{I}_n - \mathbf{L})\left(\mathcal{T}\mathbf{T}^{-1}\right) + (\mathbf{I}_n - \mathbf{L})\left(\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\right)\right\| \\
&\leq \left\|\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\right\|\left\|\mathcal{T}\mathbf{T}^{-1}\right\| + \left\|\mathcal{T}\mathbf{T}^{-1} - \mathbf{I}_n\right\| \\
&\leq 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}}\left(1 + 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}}\right) + 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}}
\end{aligned}$$

$$\leq 10\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}},$$

where we used the fact that $\|\mathbf{I}_n - \mathbf{L}\| \leq 1$ (Chung, 1997).

We now control $\|\mathcal{M} - \mathcal{L}\|$. Let B_{ij} for all $1 \leq i, j \leq n$ be independent Bernoulli variables that take the value 1 with probability p_{ij} and 0 otherwise. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ be the standard basis for $\mathbb{R}^n$. Define for all $1 \leq i, j \leq n$ the $n \times n$ matrices

$$\mathbf{A}_{ij} = \begin{cases} \boldsymbol{e}_i \boldsymbol{e}_j' + \boldsymbol{e}_j \boldsymbol{e}_i', & i \neq j, \\ \boldsymbol{e}_i \boldsymbol{e}_i', & i = j. \end{cases}$$

Then we may write the object of interest as the sum

$$\mathcal{M} - \mathcal{L} = \sum_{i \leq j} \mathbf{Y}_{ij},$$

where

$$\mathbf{Y}_{ij} = \frac{B_{ij} W_{ij} - p_{ij} \mu_{ij}}{\sqrt{\bar{d}_i \bar{d}_j}} \mathbf{A}_{ij}$$

are mean-zero independent random matrices. As the eigenvalues of $\mathbf{A}_{ij}$ are in $\{-1, 0, 1\}$, the eigenvalues of $\mathbf{Y}_{ij}$ are in

$$\left\{ \frac{\pm (B_{ij} W_{ij} - p_{ij} \mu_{ij})}{\sqrt{\bar{d}_i \bar{d}_j}}, \frac{\pm p_{ij} \mu_{ij}}{\sqrt{\bar{d}_i \bar{d}_j}}, 0 \right\},$$

and thus $\|\mathbf{Y}_{ij}\| \leq \nu / \sqrt{\bar{d}_i \bar{d}_j} \leq \nu / \bar{d}_{min}$ by similar arguments to before. Notice that

$$\mathbf{A}_{ij}^2 = \begin{cases} \boldsymbol{e}_i \boldsymbol{e}_i' + \boldsymbol{e}_j \boldsymbol{e}_j', & i \neq j, \\ \boldsymbol{e}_i \boldsymbol{e}_i', & i = j. \end{cases}$$

The variance statistic is

$$\sum_{1 \leq i \leq j \leq n} \mathbb{E}\left[\mathbf{Y}_{ij}^2\right] = \sum_{1 \leq i \leq j \leq n} \operatorname{Var}(B_{ij} W_{ij}) \frac{\mathbf{A}_{ij}^2}{\bar{d}_i \bar{d}_j}$$

$$= \sum_{i=1}^{n} \operatorname{Var}(B_{ii}W_{ii}) \frac{\boldsymbol{e}_i \boldsymbol{e}_i'}{\bar{d}_i^2} + \sum_{1 \le i < j \le n} \operatorname{Var}(B_{ij}W_{ij}) \frac{(\boldsymbol{e}_i \boldsymbol{e}_i' + \boldsymbol{e}_j \boldsymbol{e}_j')}{\bar{d}_i \bar{d}_j}$$

$$= \sum_{i=1}^{n} \frac{1}{\bar{d}_i} \left(\sum_{j=1}^{n} \frac{\operatorname{Var}(B_{ij}W_{ij})}{\bar{d}_j} \right) \boldsymbol{e}_i \boldsymbol{e}_i'.$$

Using $\operatorname{Var}(B_{ij}W_{ij}) \le p_{ij}\mu_{ij}\nu$ from before and the fact that this is a diagonal matrix, we obtain

$$\lambda_n \left(\sum_{1 \le i \le j \le n} \mathbb{E}\left[\mathbf{Y}_{ij}^2\right] \right) = \max_i \frac{1}{\bar{d}_i} \left(\sum_{j=1}^{n} \frac{\operatorname{Var}(B_{ij}W_{ij})}{\bar{d}_j} \right)$$

$$\le \nu \max_i \frac{1}{\bar{d}_i} \left(\sum_{j=1}^{n} \frac{p_{ij}\mu_{ij}}{\bar{d}_j} \right) \le \frac{\nu}{\bar{d}_{min}} \max_i \frac{1}{\bar{d}_i} \left(\sum_{j=1}^{n} p_{ij}\mu_{ij} \right) = \frac{\nu}{\bar{d}_{min}},$$

as $\sum_j p_{ij}\mu_{ij} = \bar{d}_i$.

We may thus set $M = s^2 = \nu/\bar{d}_{min}$ and apply Corollary 7.1 from Oliveira (2009), which yields for all $t > 0$

$$\mathbb{P}\left(\left\| \sum_{1 \le i \le j \le n} \mathbf{Y}_i \right\| \ge t \right) \le 2n e^{-\frac{t^2 \bar{d}_{min}}{\nu(8+4t)}}.$$

Take $t = 4\sqrt{(\nu \log(2n/\delta))/\bar{d}_{min}}$. When we bounded the degrees, we had ensured $t \le 3/4 \le 2$. Hence

$$\mathbb{P}\left(\|\mathcal{M} - \mathcal{L}\| \le 4\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}} \right) \ge 1 - \delta,$$

and it follows by the triangle inequality that

$$\mathbb{P}\left(\|\mathbf{L} - \mathcal{L}\| \le 14\sqrt{\frac{\nu \log(2n/\delta)}{\bar{d}_{min}}} \right) \ge 1 - \delta.$$

□

Theorem OA-2 extends the matrix concentration inequality results for matrices with independent sub-Gaussian rows of Theorem 5.39 from Vershynin (2012). The rows are allowed to be strongly mixing with generalised exponential tails. A key result that we make use of is Theorem 1 (and Remark 1) of Merlevède, Peligrad, and Rio (2011), which is a Bernstein-type concentration inequality for strongly mixing random variables. We consider sample covariance matrices in particular.

Theorem OA-2. *Consider a $T \times n$ matrix $\mathbf{Y}$, whose rows are observations of centred n-dimensional stationary random vectors $[\mathbf{Y}]_{t\bullet}$ that satisfy Assumption 2. Let their covariance matrix be $\mathbf{\Sigma}$ and define the sample covariance matrix as $\widehat{\mathbf{\Sigma}} = \frac{1}{T}\mathbf{Y}'\mathbf{Y}$.*

If $T = \Omega\left(n^{2/\gamma-1}\right)$, we have

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| = O\left(\sqrt{\frac{n}{T}}\|\mathbf{\Sigma}\|\right),$$

with high probability where $C > 0$ is an absolute constant.

Proof. We begin by working with the isotropic version of the random vectors, $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t$. For those, we adopt a similar strategy to Vershynin (2012) and proceed in three steps. First we discretise the unit sphere using a net which allows us to approximate the spectral norm of the quantity of interest. We then apply the mixing concentration inequality of Merlevède *et al.* (2011) to control the spectral norm for every vector on the unit sphere. Next, we take the union bound over all such vectors to evaluate the spectral norm. At this point we have a concentration inequality for the isotropic vectors. In the fourth step, we translate that statement to the non-isotropic vectors $\boldsymbol{Y}_t$.

Step 1: Approximation. Apply Lemma 5.4 of Vershynin (2012) with a 1/4-net $\mathcal{N}_{1/4} \equiv \mathcal{N}$ to obtain

$$\left\|\frac{1}{T}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_n\right\| \leq \frac{1}{1-1/2}\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\boldsymbol{x}'\left(\frac{1}{T}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_n\right)\boldsymbol{x}\right|$$

$$= 2\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\frac{1}{T}\boldsymbol{x}'\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{x}\right| = 2\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right|,$$

and let

$$\epsilon = C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}},$$

for some $r \geq 0$ where $C > 0$ is a constant. We first wish to establish

$$\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right| \leq \frac{\epsilon}{2},$$

with high probability.

Step 2: Concentration. Fix a vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$ and notice that $\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2$ may be written as a sum so that

$$\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1 = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{Y}_t'\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right)^2 - 1 \equiv \frac{1}{T}\sum_{t=1}^{T} Z_t,$$

where $Z_t \equiv \left(\mathbf{Y}_t'\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right)^2 - 1$. Notice that $\mathbb{E}\left[\left(\mathbf{Y}_t'\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right)^2\right] = \|\boldsymbol{x}\|^2 = 1$ as $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t$ are isotropic vectors, so Z_t are centred. Assumption 2 implies

$$\mathbb{P}\left(\left|\left(\mathbf{Y}_t'\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right)^2 - 1\right| > s\right) \le c_3 e^{-\left(s/c_2'\right)^{\gamma_2/2}},$$

which is the required tail behaviour for the sequence $\{Z_t\}$ (See Lemmas 7 and 8 of Gudmundsson, 2018). We may then apply Theorem 1 of Merlevède *et al.* (2011) to the sequence $\{Z_t\}$. For any $T \ge 4$, there exist positive constants C_1, C_2, C_3 and C_4 that depend only on c_1, c_2, c_3, γ_1 and γ_2 such that for any $\epsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right| \ge \frac{\epsilon}{2}\right) = \mathbb{P}\left(\left|\sum_{t=1}^{T} Z_t\right| \ge \frac{T\epsilon}{2}\right) \le \mathbb{P}\left(\sup_{r\le T}\left|\sum_{t=1}^{r} Z_t\right| \ge \frac{T\epsilon}{2}\right)$$

$$\begin{aligned}&\le T\exp\left(-\frac{(T\epsilon/2)^\gamma}{C_1}\right) + \exp\left(-\frac{(T\epsilon/2)^2}{C_2\left(1+TV\right)}\right)\\ &+ \exp\left(-\frac{(T\epsilon/2)^2}{C_3 T}\exp\left(\frac{(T\epsilon/2)^{\gamma(1-\gamma)}}{C_4\left(\log\left(T\epsilon/2\right)\right)^\gamma}\right)\right),\end{aligned} \tag{OA-2}$$

where V is finite and γ is as in Assumption (2).

We begin by looking at the first term of (OA-2). Substituting $\epsilon = C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}}$ in, we obtain

$$T\exp\left(-\frac{(T\epsilon/2)^\gamma}{C_1}\right) = \exp\left(\log(T) - \frac{\left(T\left(C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}}\right)\right)^\gamma}{2^\gamma C_1}\right)$$

$$= \exp\left(\log(T) - \frac{1}{2^\gamma C_1}\left(C\sqrt{Tn} + r\sqrt{T}\right)^\gamma\right) \le \exp\left(\log(T) - \frac{1}{2^\gamma C_1}\left(C^2 Tn + r^2 T\right)^{\gamma/2}\right),$$

as $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$. Notice that

$$\frac{\left(C^2Tn + Tr^2\right)^{\gamma/2}}{C^2n + r^2} = \frac{(Tn)^{\gamma/2}}{n} \frac{\left(C^2 + r^2/n\right)^{\gamma/2}}{C^2 + r^2/n} = T^{\gamma/2} n^{\gamma/2-1} \left(C^2 + \frac{r^2}{n}\right)^{\gamma/2-1}.$$

As $T = \Omega\left(n^{2/\gamma - 1}\right)$, this ratio is not shrinking to zero. Furthermore, we may assume that $T = o\left(e^n\right)$ for all practical purposes, so we may write

$$\exp\left(\log(T) - \frac{1}{2^\gamma C_1}\left(C^2Tn + r^2T\right)^{\gamma/2}\right) \leq \exp\left(-C'\left(C^2n + r^2\right)\right),$$

for some appropriately small constant $C' > 0$.

Plugging $\epsilon = C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}}$ into the second term of (OA-2), we obtain

$$\exp\left(-\frac{\left(T\epsilon/2\right)^2}{C_2\left(1 + TV\right)}\right) = \exp\left(-\frac{1}{4C_2\left(1 + TV\right)}\left(C\sqrt{Tn} + r\sqrt{T}\right)^2\right)$$

$$\leq \exp\left(-\frac{1}{4C_2\left(V + 1/T\right)}\left(C^2n + r^2\right)\right) \leq \exp\left(-C''\left(C^2n + r^2\right)\right),$$

for an appropriate constant $C'' > 0$, where we used the fact that $(a+b)^2 \geq a^2 + b^2$ if $a, b \geq 0$.

We begin by looking at the inner exponential term in the third term of (OA-2)

$$\exp\left(\frac{\left(T\epsilon/2\right)^{\gamma(1-\gamma)}}{C_4\left(\log\left(T\epsilon/2\right)\right)^\gamma}\right) = \exp\left(\frac{1}{2^{\gamma(1-\gamma)}C_4}\left(\frac{\left(C\sqrt{Tn} + r\sqrt{T}\right)^{1-\gamma}}{\log\left(C\sqrt{Tn} + r\sqrt{T} - \log(2)\right)}\right)^\gamma\right).$$

As $\gamma < 1$, we have $(1-\gamma)x^{-\gamma} > x^{-1}$ for all x large enough. This term is thus increasing in T, so we may bound the third term of (OA-2) with

$$\exp\left(-\frac{\left(T\epsilon/2\right)^2}{C_3T}\exp\left(\frac{\left(T\epsilon/2\right)^{\gamma(1-\gamma)}}{C_4\left(\log\left(T\epsilon/2\right)\right)^\gamma}\right)\right) \leq \exp\left(-\frac{\left(T\epsilon/2\right)^2}{C_3T}C^*\right)$$

$$= \exp\left(-\frac{C^*}{4C_3}\left(C\sqrt{n} + r\right)^2\right) \leq \exp\left(-\frac{C^*}{4C_3}\left(C^2n + r^2\right)\right) \leq \exp\left(-C'''\left(C^2n + r^2\right)\right),$$

for appropriate constants $C^*, C''' > 0$.

Finally, let $c = \min\{C', C'', C'''\}$. Taking things together, we obtain from (OA-2)

$$\mathbb{P}\left(\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right| \geq \frac{\epsilon}{2}\right) \leq 3\exp\left(-c\left(C^2 n + r^2\right)\right),$$

for the vector $\boldsymbol{x}$ that we fixed.

Step 3: Union bound. Notice that we may choose the net $\mathcal{N}$ such that its covering number $\mathcal{N}\left(\mathbb{S}^{n-1}, 1/4\right)$ is bounded by $(1 + 2/(1/4))^n = 9^n$ by Lemma 5.2 of Vershynin (2012). We then obtain

$$\mathbb{P}\left(\max_{\boldsymbol{x}\in\mathcal{N}}\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right| \geq \frac{\epsilon}{2}\right) = \mathbb{P}\left(\bigcup_{\boldsymbol{x}\in\mathcal{N}}\left(\left|\frac{1}{T}\left\|\mathbf{Y}\mathbf{\Sigma}^{-1/2}\boldsymbol{x}\right\|^2 - 1\right| \geq \frac{\epsilon}{2}\right)\right)$$

$$\leq 3 \cdot 9^n \exp\left(-c\left(C^2 n + r^2\right)\right) \leq 3e^{-cr^2},$$

where the second inequality follows for C large enough. In other words, if $T = \Omega\left(n^{2/\gamma - 1}\right)$ and n is sufficiently large, we have for every $r \geq 0$

$$\mathbb{P}\left(\left\|\frac{1}{T}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_n\right\| \leq C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}}\right) \geq 1 - 3e^{-cr^2},$$

where $c, C > 0$ are absolute constants.

Step 4: Translation to the non-isotropic random vectors. Notice that

$$\left\|\mathbf{\Sigma}^{1/2}\right\|^2 = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|=1}\left\|\mathbf{\Sigma}^{1/2}\boldsymbol{x}\right\|^2 = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|=1}\boldsymbol{x}\mathbf{\Sigma}\boldsymbol{x} = \|\mathbf{\Sigma}\|.$$

We then have

$$\epsilon\|\mathbf{\Sigma}\| = \epsilon\left\|\mathbf{\Sigma}^{1/2}\right\|^2 \geq \left\|\mathbf{\Sigma}^{1/2}\right\|\left\|\frac{1}{T}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_n\right\|\left\|\mathbf{\Sigma}^{1/2}\right\|$$

$$\geq \left\|\frac{1}{T}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{Y}\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} - \mathbf{\Sigma}\right\| = \left\|\frac{1}{T}\mathbf{Y}'\mathbf{Y} - \mathbf{\Sigma}\right\|,$$

with probability at least $1 - 3e^{-cr^2}$. This shows that if $T = \Omega\left(n^{2/\gamma - 1}\right)$ and n is sufficiently large, we have for every $r \geq 0$

$$\mathbb{P}\left(\left\|\frac{1}{T}\mathbf{Y}'\mathbf{Y} - \mathbf{\Sigma}\right\| \leq \left(C\sqrt{\frac{n}{T}} + \frac{r}{\sqrt{T}}\right)\|\mathbf{\Sigma}\|\right) \geq 1 - 3e^{-cr^2}.$$

As the vectors $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t$ are isotropic, $c, C > 0$ are absolute constants and do not depend on $\|\mathbf{\Sigma}\|$. Finally, take $r = C\sqrt{n}$ to obtain

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| \leq 2C\sqrt{\frac{n}{T}}\|\mathbf{\Sigma}\| \equiv C'\sqrt{\frac{n}{T}}\|\mathbf{\Sigma}\|,$$

with probability at least $1 - 3e^{-cC^2 n} \geq 1 - 3e^{-n}$ for C large enough. □

References

Chung, F. R. K. (1997). *Spectral Graph Theory*. American Mathematical Society.

Gudmundsson, G. S. (2018). Community detection in large vector autoregressions. Working paper.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, **151**, 435–474.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Technical report.

Patton, A. J. (2011). Data-based ranking of realised volatility estimators. *Journal of Econometrics*, **161**(2), 284–303.

Patton, A. J. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer, Berlin.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Inc.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, page 210–268. Cambridge University Press.