A 2-stage elastic net algorithm for estimation of sparse networks with heavy-tailed data

Davide Bernardini, Sandra Paterlini and Emanuele Taufer

Department of Economics and Management, University of Trento, Trento, Italy

ABSTRACT
We propose a new 2-stage procedure that relies on the elastic net penalty to estimate a network based on partial correlations when data are heavy-tailed. The new estimator allows us to consider the LASSO penalty as a special case. Extensive simulation analysis shows that the 2-stage estimator performs best for heavy-tailed data and it is also robust to distribution misspecification, both in terms of identification of the sparsity patterns and numerical accuracy. Empirical results on real-world data focus on the estimation of the European banking network during the Covid-19 pandemic. We show that the new estimator can provide interesting insights both for the development of network indicators, such as network strength, to identify crisis periods and for the detection of banking network properties, such as centrality and level of interconnectedness, that might play a relevant role in setting up adequate risk management and mitigation tools.

ARTICLE HISTORY
Received 27 February 2022
Accepted 12 September 2022

KEYWORDS
Network estimation; elastic net penalty; heavy-tailed data; European banking network

1. Introduction
An undirected graphical model is a set of two elements: a graph $G$, which encodes the conditional dependence structure of a set of random variables, and their joint probability distribution $f$ (see [1,2]). So far, most of the existing literature has focused on Gaussian graphical models, where the joint distribution is multivariate Gaussian $\mathcal{N}_p(\mu, \Sigma)$. In such a set-up, the conditional dependence structure, and thus the graph of the Gaussian graphical model, can be retrieved by looking, for example, at the inverse matrix $\Theta$ of the covariance matrix $\Sigma$. In particular, $\theta_{ij} = 0$ implies conditional independence between variable $i$ and $j$ (see [1]).

Several methods have been proposed to estimate $\Theta$ and thus the conditional dependence graph. Typically, a penalized estimator is used to recover the sparsity pattern and then reconstruct the graph of conditional dependencies (see, for example, [3–10]). A widely used penalty is the Least Absolute Shrinkage and Selection Operator (LASSO) penalty proposed by Tibshirani [11] and based on $\ell_1$-norm. For example, Meinshausen and Bühlmann [8] introduced a conditional LASSO-penalized regression approach. Friedman
et al. [5] proposed the graphical LASSO (glasso) algorithm. It adds an element-wise $\ell_1$-norm penalty to the multivariate Gaussian log-likelihood function in order to estimate a sparse $\Theta$. More recently, Bernardini et al. [12] and Kovács et al. [13] introduced independently penalized estimators of the precision matrix based on the elastic net penalty (see [14]). Bernardini et al. [12] suggested three different alternative approaches and investigate their performance through simulations, while Kovács et al. [13] focused on a direct modification of the glasso algorithm augmented with target matrices. The elastic net penalty extends the LASSO penalty by combining together $\ell_1$-norm (LASSO) and $\ell_2$-norm. Thus, LASSO is a special case of the elastic net penalty when the weight given to $\ell_2$-norm is zero.

This paper proposes a new 2-stage estimator of the precision matrix that considers the case where the joint distribution is a multivariate t-Student. In the following, we thus introduce the 2-stage t-Student elastic net ($2Stelnet$) estimator, which relies on the Expectation–Maximization (EM) algorithm (see [15]). We build both upon the tlasso algorithm by Finegold and Drton [16] and the 2-stage graphical elastic net estimator ($2Sgelnet$) in Bernardini et al. [12]. As for $2Sgelnet$, the new proposed estimator includes both LASSO and elastic net cases, since the former is a special case of the latter. When the distribution of data is not Gaussian, estimators based on Gaussian distribution may lead to poor estimates because the assumption about the underlying distribution is violated. Thus, extensions to other multivariate distributions are useful to address such situations.

Note that for the t-Student distribution, a zero element in $\Theta$ does not necessarily imply conditional independence, but only a zero partial correlation. Thus a partial correlation matrix is estimated. Monte Carlo simulations show that the proposed estimator ($2Stelnet$) based on the t-Student distribution leads, in general, to better estimates in a wide range of situations where the data exhibit heavier tails than in the Gaussian case and in the presence of model misspecification. Fat tails are, for example, a well-known fact of financial time series (see [17]). Furthermore, economics and finance have recently focused on the importance of several network structures (see, for example, [18–22]). There is, consequently, a growing interest in broadening the set of tools available to estimate network structures. For example, financial networks have received renewed attention in the aftermath of the recent global financial crisis. Several empirical studies have focused their attention on the estimation and analysis of these networks; see, for example, Barigozzi and Brownlees [23], Bilio et al. [24], Dermier et al. [25] and Torri et al. [26]. In this paper, we rely on $2Stelnet$ to estimate the relationships among a large set of important European banks for the period 2018–2020. In order to improve our understanding of the systemic risk of the European banking network and assess the impact of the Covid-19 pandemic, we track some common network statistics and the evolution of the average intensity of the connections. We also identify the most central banks by looking at different centrality measures. Finally, we carry out an exploratory simulation study to map the effects of shock in the estimated network.

This paper is organized as follows. Section 2 describes the $2Stelnet$ estimator. Section 3.1 describes the set-up of our simulation analysis and Section 3.2 reports the results obtained. Finally, in Section 4, we estimate the European banking network during the period of 2018–2020 and analyse its characteristics.
2. t-Student graphical models

Let \( X = [X_1, \ldots, X_p]^\top \) be a \( p \)-dimensional random vector with joint multivariate t-Student distribution \( t_p(\mu, \Psi^{-1}, \nu) \) where \( \mu \) is the mean vector, \( \Psi^{-1} \) is the positive definite scatter, or dispersion, matrix and \( \nu \) are the degrees of freedom. The covariance matrix \( \Sigma \) of \( X \) and its inverse, the precision matrix \( \Theta \), are then equal to \( \Sigma = \frac{\nu}{\nu - 2} \Psi^{-1} \) and \( \Theta = \frac{\nu - 2}{\nu} \Psi \), with \( \nu > 2 \). Our goal is to estimate a sparse precision matrix \( \Theta \), from which to retrieve a sparse graph whose weights are partial correlations between couples of variables. Partial correlations are obtained by properly scaling the off-diagonalelements in \( \Theta \). Let \( \theta_{jk} \) be the element in the \( j \)th row and \( k \)th column of \( \Theta \), the partial correlation \( p_{jk} \) between components \( X_j \) and \( X_k \) is then equal to:

\[
p_{jk} = -\frac{\theta_{jk}}{\sqrt{\theta_{jj} \theta_{kk}}}.
\]

Thus, we build a graph \( \mathcal{G}(V, E) \) with the set of nodes \( V = \{1, \ldots, p\} \) representing the elements in \( X \) and the set of edges \( E \subseteq V \times V \) and \( (j, k) \in E \) if \( p_{jk} \neq 0 \), where \( (j, k) \) represents the edge between elements \( X_j \) and \( X_k \). Note that, differently from Gaussian graphical models where the distribution of \( X \) is multivariate Gaussian, here \( p_{jk} = 0 \) does not necessarily imply that elements \( X_j \) and \( X_k \) are conditionally independent (see [27]). Nonetheless, it can be shown that, in this case, if nodes \( j \) and \( k \) are separated by a subset of nodes \( C \subseteq \{h \mid h \in V \land h \neq j, k\} \) in the graph \( \mathcal{G} \), then the elements \( X_j \) and \( X_k \) are conditionally not correlated given the subset \( C \) (see [16]). For elliptical distributions, conditional and partial correlations are equivalent (see [27]).

Following the scale-mixture representation of a multivariate t-Student distribution as in Finegold and Drton [16], we have that

\[
X = \mu + \frac{Y}{\sqrt{\tau}} \sim t_p(\mu, \Psi^{-1}, \nu).
\]

where \( Y \sim \mathcal{N}_p(0, \Psi^{-1}) \) and \( \tau \sim \Gamma\left(\frac{\nu}{2}, \frac{v}{2}\right) \). Thus, the multivariate \( p \)-dimensional t-Student distribution can be seen as a mixture of a multivariate \( p \)-dimensional Gaussian distribution with an independent univariate gamma distribution. By properly exploiting this representation, it is possible to rely on the EM algorithm [15] to estimate the parameters of the multivariate t-Student distribution. Following closely the \textit{lasso} procedure proposed by Finegold and Drton [16], we introduce a similar EM algorithm to produce a sparse estimate of \( \Theta \). Differently from the \textit{lasso} that uses LASSO, or \( \ell_1 \)-norm, penalty (see [11]) to induce sparsity, we propose an approach that utilizes the \textit{elastic net} penalty, a linear combination of \( \ell_1 \)-norm and \( \ell_2 \)-norm [14], to do a penalized estimation of \( \Theta \). In fact, we use the 2SGelnet by Bernardini et al. [12], instead of relying on \textit{glasso} by Friedman et al. [5]. The core idea behind it is to estimate a sparse \( \Psi \), the precision matrix of the multivariate Gaussian in the mixture, since its elements are proportional to the elements of \( \Theta \).

2.1. The EM algorithm

Let \( x_1, \ldots, x_n \) be \( n \) \( p \)-vectors of observations drawn from \( t_p(\mu, \Psi^{-1}, \nu) \) distribution, realizations of \( X \). The random variable \( \tau \) in the mixture (2) is considered the hidden, or latent, variable whose value is updated given the current estimate of the parameters.
and the observed data. Let also \( \tau_i \) be the value of the latent variable \( \tau \), associated with observation \( x_i \). As in the lasso of Finegold and Drton [16], we also assume that the degrees of freedom \( v \) are known in advance. This simplifies the procedure, but \( v \) can be also treated as an unknown parameter and thus estimated. As far as the estimation of the degrees of freedom is concerned, one could consider endogenous estimation by modifying the EM algorithm, see, for example, Liu and Rubin [28], Liu [29], Dogru et al. [30] or Hasannasab et al. [31] where the unknown value of the degrees of freedom \( v \) is iteratively estimated with a proper modification of the EM algorithm, with an additional procedure in the M-step in which the estimate of \( v \) is updated. Alternatively, a simpler and viable procedure would be to estimate \( v \) with an external procedure, see, for example, Pascal et al. [32] and then plug in the obtained estimate in Algorithm 1.

The EM algorithm proceeds as follows [16]. At time step \( t + 1 \):

- **Expectation step (E-step)**
  - The expected value of \( \tau \), given a generic vector of observations \( x \) and parameters \( \mu, \Psi \) and \( v \), is:
    \[
    E(\tau \mid X = x) = \frac{v + p}{v + (x - \mu)\top \Psi(x - \mu)}.
    \]
  - Using the current estimates \( \hat{\mu}^{(t)} \) and \( \hat{\Psi}^{(t)} \) of the parameters, we update the estimated value \( \hat{\tau}_i^{(t+1)} \) of \( \tau_i \):
    \[
    \hat{\tau}_i^{(t+1)} = \frac{v + p}{v + (x_i - \hat{\mu}^{(t)})\top \hat{\Psi}^{(t)}(x_i - \hat{\mu}^{(t)})}.
    \]
    for each \( i = 1, \ldots, n \).

- **Maximization step (M-step)**
  - Compute the updates of parameters given the data and \( \hat{\tau}_i^{(t+1)} \):
    \[
    \hat{\mu}^{(t+1)} = \frac{\sum_{i=1}^{n} \hat{\tau}_i^{(t+1)} x_i}{\sum_{i=1}^{n} \hat{\tau}_i^{(t+1)}}.
    \]
    \[
    \hat{\Psi}^{(t+1)} = \hat{\Psi}_{\alpha,\lambda}(\hat{\tau}_i^{(t+1)}, x_1, \ldots, x_n).
    \]
    where \( \hat{\Psi}_{\alpha,\lambda} \) is a penalized estimator of \( \Psi \), whose penalty is controlled by hyperparameters \( \alpha \) and \( \lambda \).

The EM algorithm cycles sequentially through the E and M steps until a convergence criterion is satisfied. Let \( \hat{\psi}_{jk} \) be the element in \( j \)th row and in \( k \)th column of \( \hat{\Psi} \), we stopped the iterations in the algorithm when \( \max_{j,k} |\hat{\psi}_{jk}^{(t+1)} - \hat{\psi}_{jk}^{(t)}| \) is smaller than a given threshold value \( \delta \).

In the following, we discuss the estimator \( \hat{\Psi}_{\alpha,\lambda}(\cdot) \) of \( \Psi \) for (6) based on \texttt{2Sgelnet} of Bernardini et al. [12].

### 2.2. Two-stage t-Student elastic net – \texttt{[2Stelnet]}\[12\]

The estimator \( \hat{\Psi}_{\alpha,\lambda}(\cdot) \) consists in a 2-step procedure. At first, it estimates the sparsity structure of \( \hat{\Psi}^{(t+1)} \) by using conditional regressions with elastic net penalty. This is inspired
by the neighbourhood selection approach of Meinshausen and Bühlmann [8] where the LASSO penalty is used. At each iteration, we first transform the observed data as follows (from (2)):
\[
x_i = (x_i - \hat{\mu}^{(t+1)})\sqrt{\hat{\tau}_i^{(t+1)}}.
\]
Let \(\tilde{X}\) be the \(n\) by \(p\) matrix of transformed observations, such that the \(i\)th row is equal to \(\tilde{x}_i^T\). Let also \(\tilde{X}_k\) be the \(k\)th column of \(\tilde{X}\) and \(\tilde{X}_{-k}\) be \(\tilde{X}\) without the \(k\)th column. We fit \(p\) elastic net penalized regressions using the \(k\)th component of the transformed vectors as the dependent variable and the remaining components as predictors:
\[
\hat{b}_k = \arg\min_{b_k} \left\{ \frac{1}{2n} \| \tilde{X}_k - a_k - \tilde{X}_{-k} b_k \|_2^2 + \lambda [\alpha \| b_k \|_1 + \frac{1}{2} (1 - \alpha) \| b_k \|_2^2] \right\}.
\]
with \(k = 1, \ldots, p\). Note that \(b_k\) are our coefficients of interest, while \(a_k\) is the intercept. Such intercept term can be ignored and excluded from equation (8) if the variables are standardized, which is a common procedure using penalized estimation methods. The parameter \(\lambda\) controls the overall strength of the penalty, while \(\alpha\) controls the balance between \(\ell_1\) and \(\ell_2\) norms. As in Meinshausen and Bühlmann [8], we reconstruct the graph representing the connections (\(p_{jk} \neq 0\)) among the components of \(X\) (and also of \(Y\)). We include in the neighbourhood of the node \(k\) the node \(j\) if the corresponding coefficient of the component \(j\) in \(\hat{b}_k\) is different from 0. Then, through the reconstructed neighbourhoods of all nodes, we produce an estimate \(\hat{E}\) of the edge set \(E\). This procedure can lead to the situation where an edge \((j, k)\) is included in \(\hat{E}\) according to the neighbourhood of \(j\), \(\text{ne}(j)\), but not accordingly to the neighbourhood of \(k\), \(\text{ne}(k)\). To deal with such a situation, we can use two rules (see [8]):

- **AND rule**: edge \((j, k) \in \hat{E}\) if node \(j \in \text{ne}(k) \land \text{node } k \in \text{ne}(j)\).
- **OR rule**: edge \((j, k) \in \hat{E}\) if node \(j \in \text{ne}(k) \lor \text{node } k \in \text{ne}(j)\).

Once estimated \(\hat{E}\), we use it to set the zero elements constraints in the current update \(\hat{\Psi}^{(t+1)}\). In particular, the update \(\hat{\Psi}^{(t+1)}\) is the maximizer of the following constrained optimization problem, with \(\hat{S}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^{(t+1)} (x_i - \hat{\mu}^{(t+1)}) (x_i - \hat{\mu}^{(t+1)})^T\):
\[
\max_{\Psi} \{ \log(\det(\Psi)) - \text{trace}(\hat{S}^{(t+1)} \Psi) \}
\]
subject to
\[
\psi_{jk} = \psi_{kj} = 0 \text{ if edge } (j, k) \notin \hat{E}.
\]
This problem can be rewritten in Lagrangian form as:
\[
\hat{\Psi}^{(t+1)} = \arg\max_{\Psi} \left\{ \log(\det(\Psi)) - \text{trace}(\hat{S}^{(t+1)} \Psi) - \sum_{(j,k) \notin \hat{E}} \gamma_{jk} \psi_{jk} \right\},
\]
where \(\gamma_{ij}\) are Lagrange multipliers having nonzero values for all \(\psi_{ij}\) constrained to 0. To solve this optimization problem (9), we use the algorithm proposed by Hastie et al. [33,
Section 17.3.1, pp. 631–634] to maximize the constrained log-likelihood and produce an estimate $\hat{\Psi}$, given the edge set $\hat{E}$ estimated in the previous step. As with 2Sgelnet, the existence of this estimator of $\Psi$ is not guaranteed for all situations. When the number $n$ of observations is greater than the number $p$ of nodes, the estimator always exists (see [1, Section 5.2.1]). In the other situations, its existence depends on the structure of the estimated edge set $\hat{E}$ (see [1, Section 5.3.2]).

We report here a pseudo-code for the 2Stelnet algorithm:

| Algorithm 1: 2Stelnet (Section 2.2). |
|-------------------------------------|
| Set $\hat{\mu}(0) = \frac{1}{n} \sum_{i=1}^{n} x_i$, $\hat{S}(0) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}(0))(x_i - \hat{\mu}(0))^\top$ and $\hat{\Psi}(0) = (\hat{S}(0))^{-1}$ |
| Set a threshold value $\delta$, set Convergence = FALSE and set $t = 0$. |
| while Convergence == FALSE do |
| • E-Step: |
| - Compute the updated estimate $\hat{\tau}^{(t+1)}_i$ as in [16] |
| • M-Step: |
| - Compute the updated estimate $\hat{\mu}^{(t+1)}$ as in [16] |
| - Transform each observation $x_i$, for $i = 1, \ldots, p$ as in (7) |
| - Estimate $b_k$ using $p$ elastic net regressions as in (8) |
| - Using each $b_k$, reconstruct an estimated edge set $\hat{E}$ |
| - Given $\hat{E}$ compute the updated estimate $\hat{\Psi}^{(t+1)}$ as the maximizer of (9) |
| • Set $t = t + 1$ |
| if $\max_{j,k}(|\hat{\Psi}^{(t+1)}_{j,k} - \hat{\Psi}^{(t)}_{j,k}|) < \delta$ then |
| Convergence = TRUE |
| end if |
| end while |

3. Simulations

3.1. Simulation set-up

We rely on simulations to assess the performances of 2Stelnet and compare it with 2Sgelnet [12] and with the well-known glasso [5] and lasso [16] algorithms. The degrees of freedom are fixed a priori ($\nu = 3$) for lasso and 2Stelnet. For the sake of brevity, we report only the results obtained with AND rule as they are qualitatively similar to the ones obtained using OR rule. There are some specific situations where one rule is better than the other, but we do not find an overall winner. Results for OR rule are available upon request.

We randomly generate four 50 × 50 precision matrices encoding different relationship structures among 50 variables. We consider the following four network topologies: scale-free, core-periphery, random and cluster. Precision matrices embedding the randomly generated structures are reported in Figure 7 of Appendix A. For scale-free, random
and cluster, we use the R package huge (setting $\nu = 0.3$, $\mu = 0.1$) that allows to directly generate precision matrices with a given sparsity structure. Instead, for core-periphery, we use the algorithms in [26] to generate the sparsity pattern, and then we utilize the procedure suggested in huge package to produce the precision matrices. Given a precision matrix $\Theta$, we use it to generate $n$ vectors of observations from the following two multivariate distributions:

- Multivariate normal: $N_p(0, \Theta^{-1})$.
- Multivariate $t$-Student: $t_p(0, \nu - 2\nu / \nu - 1, \nu)$, with $\nu = 3$.

We consider three sample sizes, $n = 100$, 250, 500 and $p = 50$. Thus, for each of the 8 couples network distribution, we have $100 \times 50$, $250 \times 50$ and $500 \times 50$ datasets, respectively. We set up Monte Carlo experiments with 100 runs for each sample size $n$.

The values of the parameters that control the balance and strength of the elastic net penalty of our estimators are the following. We consider the values 0.5 and 1 for $\alpha$, while for $\lambda$ we consider 100 exponentially spaced values between $e^{-6}$ and 2 (6.5 for $tlasso$). Note that for glasso and $tlasso$ we only have the parameter $\lambda$. We search for the optimal value of $\lambda$, given a value of $\alpha$ for 2Sgelnet and 2Stelnet, using BIC criterion (see [34]): $\text{BIC}(\hat{E}, \hat{\Theta}) = -2 \cdot \log \text{Lik}(\hat{\Theta}) + |\hat{E}| \cdot \ln(n)$. $|\hat{E}|$ is the cardinality of the estimated edge set of the partial correlation graph associated with $\hat{\Theta}$, while logLik($\hat{\Theta}$) is the maximized log-likelihood given our estimate $\hat{\Theta}$.

We compare the estimators by looking both at classification performance and numerical accuracy of the estimates for optimal $\lambda$. We study their behaviour in different settings by testing them on all couples network distribution. We use the F1-score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ as a measure of classification performance, where

- Precision $= \frac{\text{True positives}}{\text{True positives} + \text{False Positives}}$.
- Recall $= \frac{\text{True positives}}{\text{True positives} + \text{False Negatives}}$.

Here, a true positive is a correctly identified edge, while a true negative is a missing edge which is correctly excluded from the edge set. A false positive is a missing edge which is erroneously included in the edge set, while a false negative is an edge which is wrongly excluded. The closer this measure is to 1, the better the classification. The F1-score is a better measure than accuracy when there is an imbalance among classes (here existence or absence of an edge), as it happens in our simulation set-up.

In order to assess the numerical accuracy of the estimates, we compute the Frobenius distance between the theoretical and estimated partial correlation matrices, $\mathbf{P}$ and $\hat{\mathbf{P}}$, respectively,

$$\|\mathbf{P} - \hat{\mathbf{P}}\|_F = \sqrt{\sum_{j=1}^{p} \sum_{k=1}^{p} |p_{jk} - \hat{p}_{jk}|^2}.$$  

These matrices can be easily obtained through a proper scaling of the theoretical and estimated precision matrices, $\Theta$ and $\hat{\Theta}$, as in (1). We follow a common convention and set the diagonal elements $p_{jj} = \hat{p}_{jj} = 0$, with $j = 1, 2, \ldots, p$. When the Frobenius distance is 0, the
estimate is exactly equal to the true value. Thus, the smaller the Frobenius distance, the better the estimate is from the numerical accuracy point of view.

### 3.2. Simulation results

In the following, we report the box plots for both the performance measures (i.e. left column: $F_1$-score; right column: Frobenius distance) and compare the estimators on 100 Monte Carlo runs for each combination of distribution and network. Box plots are grouped by sample size as reported in the $x$-axis (i.e. from left: $n = 100, 250, 500$). We test the performance of glasso, lasso, 2Sgelnet and 2Stelnet. For the 2-stage estimators, we use the AND rule as a procedure to produce an estimate of the edge set. When $\alpha = 1$, we have a pure LASSO penalty (see (8)). For the sake of brevity, analysing the simulation results, we call the 2-stage estimators with $\alpha = 1$ 2StGlasso and 2StLasso, while 2Sgelnet and 2Stelnet refers to the case where $\alpha = 0.5$. It is immediately clear how the behaviour of lasso when $n = 100$ is quite different from other estimators and from the performance of lasso with $n = 250, 500$. This is due to the fact that the optimal model selected using the BIC criterion with the smallest sample size is often the null model (or close to it), that is the model with zero edges or a diagonal precision matrix. The observed behaviour suggests that the BIC criterion with lasso and a small sample is problematic, which to our knowledge, was not discussed previously. This is not the case with the proposed estimator 2Stelnet (and 2StLasso). In Figures 1 and 2 are reported the boxplots of the two performance measures considered for the cases with multivariate normal and multivariate t-Student data, respectively. Each row in the figures corresponds to a different network structure, from top to bottom: scale-free, random, cluster and core-periphery.

Looking at the classification performance as quantified by the $F_1$-score, in Figures 1 and 2, we observe the following:

- 2Stelnet and 2StLasso with $\nu$ equal to 3 perform quite well in all the situations analysed through simulations. When data are from t-Student distribution, they are the best estimators if we look at median. When data are from multivariate normal, they perform similar to 2Sgelnet and 2Sglasso suggesting a good degree of robustness to distributional misspecification. Thus, this choice (i.e. $\nu = 3$) may be a good one from the practical point of view when the degrees of freedom of the data are unknown.
- On the contrary, 2Sgelnet and 2Sglasso are not so robust when data are from t-Student distribution.
- While 2-stage estimators increase their classification performances as the sample size $n$ grows, this is not always true for glasso and lasso.
- The value of $\alpha$, and thus the kind of penalty used, is not particularly relevant, at least in the simulation set-up considered. Depending on the network, one value of $\alpha$ can slightly outperform the other in median terms, but differences seem negligible, and thus a particular winner does not emerge. This is true both for the procedure that assumes normality and for the one that assumes t-Student distribution.
- Some network structures are more difficult to identify. For example, when we look at $F_1$-score, the core-periphery topology is the most difficult to estimate, followed by the cluster structure.
Figure 1. Normal distribution – Performance measures (F1-score, left column; Frobenius distance, right column) – From top to bottom: scale-free, random, cluster, core-periphery.
**Figure 2.** t-Student distribution ($\nu = 3$) – Performance measures (F1-score, left column; Frobenius distance, right column) – From top to bottom: scale-free, random, cluster, core-periphery.
From the numerical accuracy point of view, results show that:

- Both $2Stlenet$ and $2Stlasso$ show relatively good performances, especially with larger sample sizes outperforming $lasso$.
- While $2Stlenet$ and $2Stlasso$ are robust to normal data, on the contrary, $2Sgelnet$ and $2Sglasso$ are not robust to t-Student data. In fact, looking at the Frobenius distance, they perform the worst with such data. Again, these results suggest that setting the $\nu = 3$ when using $2Stlenet$ and $2Stlasso$ may lead to robust estimates even when this value does not match the degrees of freedom of the data (i.e. equal to $\infty$ for the Gaussian distribution).
- Again, a clear-cut winner between elastic net penalty and pure LASSO penalty in the 2-stage estimators does not emerge, neither for the ones based on Gaussian distribution, nor for the ones based on t-Student. In general, the performances observed are quite similar.

### 3.3. The role of correlation

In this subsection, we report the results of additional simulations with highly correlated data. More specifically, to investigate the role of different values of $\alpha$ in the elastic net penalty, we run a Monte-Carlo simulation generating 100 datasets of 100 observations with blocks of highly correlated variables from a multivariate normal and multivariate t-Student (with $\nu = 3$) distribution. To generate such data, we use a block-diagonal structure for the covariance matrix of the multivariate normal and for the covariance matrix of the multivariate normal in the mixture for the t-Student distribution. We set all its non-zero off-diagonal elements equal to 0.9 and the diagonal elements equal to 1. Differently, from previous simulations, we investigate a smaller network of 30 variables instead of 50, which are clustered in 3 fully connected and disjoints groups of 10 variables each. Hence, each synthetic dataset contains 100 observations of 30 variables.

The boxplots in Figure 3 show how the choice of $\alpha$ may be relevant in a highly correlated scenario. In particular, setting $\alpha = 0.5$ (i.e. $2Stelnnet$) instead of $\alpha = 1$ (i.e. $2Stlasso$) leads to notably better results both in terms of classification performance (i.e. F1-score, figure on the left) and numerical accuracy (i.e. Frobenius distance, figure on the right). These results are consistent with the idea that elastic net penalty is able to overcome some limitations.

![Figure 3](image-url) Performance measures with highly correlated variables.
of the LASSO penalty, such as the seemingly random selection among highly correlated variables (see [14]).

4. **European banking network**

In this section, we use the 2Stelnet estimator to reconstruct the European banking network in the period 2018–2020 by using daily stock prices of 36 large European banks, similarly to [26]. The period considered includes the Covid-19 pandemic, thus it is also possible to assess its impact on the evolution of this banking network and extend [26] to include the recent pandemic crisis. To deal with autocorrelation and heteroskedasticity, two common characteristics of financial time series, we fit an AR(1)-GARCH(1,1) model for each of the 36 time series of log-returns. Then, we use the residuals to estimate the partial correlation network both for the single years (i.e. 2018, 2019, 2020) and for the period 2018–2020, using a rolling window of one year with shifts of one month. The estimated correlation matrix of the residuals suggests that there is a moderate to a high level of correlation between couples of banks, with an average value of 0.57 and 75% of the pairwise correlations above 0.48. Furthermore, fitting a multivariate t-Student on the residuals of the full dataset (i.e. 2018–2020) gives an estimate of $\hat{\nu} = 4.67$. Hence, we set the value of $\alpha = 0.5$, because of high correlations, and $\nu = 3$, due to its robustness shown in simulations and its closeness to the estimated value. A sequence of 100 exponentially spaced values between $e^{-6}$ and 1.5 is considered for $\lambda$. We select its best value using the BIC criterion. Figure 8 in Appendix C shows the estimated networks using the Fruchterman–Reingold layout [35]. Furthermore, in Table 2 (Appendix B) we also reported the comparison of different estimators from the computational time point of view. They refer to the times (in seconds) needed for a full BIC calibration procedure using the whole dataset.

Common network measures (mean values) are reported in Table 1, along with the total number of edges detected. We observe an increase in all the measures from 2018 to 2020. The mean values of eccentricity and distance give us an indication about the velocity of transmission of a shock in the network. The former is the minimum number of steps needed to reach the farthest node, in terms of steps, from a specific node, while the latter is the length of the minimum path between two different nodes in the graph. Both values in all three years are quite small, suggesting that a shock could potentially diffuse pretty quickly in the system. The average values of individual clustering coefficients (see [36]) show a tendency of nodes to cluster together and are higher than the case of random graphs (0.19, 0.19 and 0.21, respectively, in 2018, 2019 and 2020). This feature, together with the low values of distance, are two core characteristics of small-world graphs [36]. The mean value of the degree is the average number of banks connected to a given bank. In our case, we detect, on average, seven banks linked to each bank. Figure 9 in Appendix C shows the degree distributions in all three years, which provides a more detailed picture of individual degrees. In 2018, the distribution was more symmetric, while in 2019 and 2020 becomes multi-modal. In fact, there is a large number of banks with a number of connections below and above the average, but few banks with a number of connections around the mean value. Also, in the last column of Table 1 the total number of edges detected is reported. They are the 19.7%, 20% and 21.3% of 630 possible edges, in 2018, 2019 and 2020, respectively. Furthermore, we also report the average strength of the nodes in each estimated network. This measure is a useful proxy of the intensity of the connections in
Table 1. Network measures for the estimated networks in years 2018–2019–2020.

| Year | Degree | Eccentricity | Distance | Clustering | Strength | N°Edges |
|------|--------|--------------|----------|------------|----------|---------|
| 2018 | 6.89   | 3.53         | 2.15     | 0.46       | 0.83     | 124     |
| 2019 | 7.00   | 3.69         | 2.24     | 0.47       | 0.87     | 126     |
| 2020 | 7.44   | 4.08         | 2.29     | 0.48       | 0.91     | 134     |

Figure 4. Strength evolution with a rolling window of 1-year length.

The graph, which here are the relations among banks represented by partial correlations. Thus, a rise in this measure suggests that the intensity of connections in the inter-bank network increased in the time period considered.

This trend is also confirmed by the evolution of network strength evaluated using a rolling window of one year and reported in Figure 4. Notice the effect of the Covid-19 pandemic on the average network strength, with a sharp rise between February 2020 and April 2020 and the subsequent stabilization on a higher level for the remaining part of 2020. This trend could not only identify the presence, but also suggests the persistence of a potential crisis period. Furthermore, in order to check our results, we run the 2Stelnet algorithm for different values of $\alpha$ and $\nu$; the results are in Figure 10 (Appendix C). Note that once $\alpha$ is fixed, the three different solutions with $\nu$ equal to 3, 15 and 30 show similar behaviour. With $\alpha = 0.25$ we have some higher variability in the Covid-19 period due to the fact that in this case the model estimates more weak relationships and these tend to vary by changing the degrees of freedom. Given the objective of the analysis of this dataset, i.e. estimate a sparse network, the indication obtained from the simulations, the values of correlations and the degrees of freedom estimated from the data, it appears that the proposed solution, with $\nu = 3$ and $\alpha = 0.5$ provides an accurate picture of the network.

A network representation of the banking system also allows detecting the most important banks by looking at their position in the network instead of their market capitalization, for example. Thus, we compute three possible centrality measures and look at the first five banks according to each measure in order to detect the most important banks in the network. Values are reported in Figure 5 (banks ordered by country and degree of centrality in 2018). Degree centrality identifies the most important banks according to the number of connections a specific bank has. Note that, for this measure, there are more than five banks in the first five positions. In fact, if we consider the first five largest values of degree, it could happen that there are multiple banks having the same number of connections. According
to degree centrality, Credit Agricole, Banco Santander, Société Generale, ING Group and UniCredit are among the five most important banks in all the three years considered, while BNP Paribas, Intesa San Paolo, Commerzbank and UBS Group appear twice within the first five. Strength centrality is based on the strength of each node in the network. Thus, it considers not only the number of connections, but also their importance. According to strength, only BNP Paribas appears three times within the first five positions in 2018, 2019 and 2020. Credit Agricole, Banco Santander and Nordea Bank are instead included in two out of three years. The last measure is eigenvector centrality. It considers the importance, in terms of the number of connections, of the neighbours of a node to quantify the importance of the node itself. Thus, from this point of view, the importance of a bank depends on the importance of other banks connected to that specific bank. Looking at eigenvector centrality, ING Group is the only bank that is among the first five banks in all the three years considered. Instead, BNP Paribas, Credit Agricole, Banco Santander and Société Generale appear two times in the five most important banks. Summing up, we observe that BNP

**Figure 5.** Centrality measures for 2018–2020.
Figure 6. Shock diffusion simulation in the 2020 estimated network. Final steady-states for all banks in the network after a unitary positive shock in Nordea Bank and Mt. Paschi.

Paribas, Credit Agricole and Banco Santander are listed at least twice among the first five banks according to all the three centrality measures considered. This can suggest that they play an important role in the interbank network, at least in the time period considered.

Finally, we investigate the effects of a shock in the banking network. We consider the network estimated in 2020, where all nodes are connected within a single component. The most and the least important banks, according to strength centrality, are Nordea Bank and Monte Paschi di Siena, respectively. We simulate the diffusion of a positive unitary shock in these two banks with the approach used by Anufriev and Panchenko [37] and observe how different the effects are on the whole banking network. Given the estimated partial correlation matrix in 2020, \( \hat{\mathbf{P}}_{2020} \), the direct, second and high-order effect of the shocks can be evaluated as follows:

\[
\mathbf{s}_i^\infty = \mathbf{e}_i + \hat{\mathbf{P}}_{2020} \mathbf{e}_i + \hat{\mathbf{P}}_{2020}^2 \mathbf{e}_i + \cdots = \sum_{t=0}^{\infty} \hat{\mathbf{P}}_{2020}^t \mathbf{e}_i = (\mathbf{I} - \hat{\mathbf{P}}_{2020})^{-1} \mathbf{e}_i. \tag{11}
\]

where \( \mathbf{e}_i \) is the initial vector of shocks and \( \mathbf{s}_i^\infty \) is final steady-state induced by \( \mathbf{e}_i \). In our simulation, \( \mathbf{e}_i \) is a vector of 0s with 1 in the \( i \)th position corresponding to a positive unitary shock in the \( i \)th bank (from Table 3 in Appendix C, \( \mathbf{e}_{\text{Nordea}} = \mathbf{e}_{31} \) and \( \mathbf{e}_{\text{Mt.Paschi}} = \mathbf{e}_{24} \)). The convergence of the sum in (11) is guaranteed when the spectral radius of \( \hat{\mathbf{P}}_{2020} \) is smaller than 1 (see Anufriev and Panchenko [37]). This is verified for our estimated partial correlation matrix \( \hat{\mathbf{P}}_{2020} \). In Figure 6, we report the final steady-states of the two positive unitary shocks considered. It is clear how the whole system reacts differently to the two different vectors of initial shocks. An individual shock in a more central bank could have a much larger impact on the entire network. In our simulation, the final overall effects, measured as the sums of final steady-states of each bank, are 142.18 and 33.15 for initial shock in Nordea and Monte Paschi, respectively. That is, the overall effects of a shock are more than four times larger if it hits the most central bank.

5. Conclusions

We introduce a new 2-stage estimator that brings together the \textit{lasso} of Finegold and Drton [16] and the \textit{2Sgelnet} of Bernardini et al. [12]. This procedure is more flexible than \textit{2Sgelnet} and it is more suitable for situations when data exhibit heavy tails and in the presence of
model misspecification. The proposed estimator relies on the elastic net penalty and also allows to consider the LASSO penalty as a special case. Exploiting the scale-mixture representation of the multivariate t-Student distribution, we use the EM algorithm to estimate the precision matrix $\Theta$. A sparse estimate is produced in the M-step through a 2-stage procedure using elastic net penalty. By running Monte-Carlo simulations, the proposed estimator is compared with glasso, tlasso and 2Sgelnet by looking both at classification performances and numerical accuracy of the optimal models selected using the BIC criterion. Four network topologies and two multivariate distributions have been considered. Simulation analysis has also been performed for hub, small-world and band networks and for t-Student with $\nu = 20$ and contaminated normal distributions. It leads to qualitatively similar results, which are available in Appendix D. We observe that 2Stelnet performs quite well with respect to the other estimators considered, especially with the largest sample sizes. The results also suggest that the choice of $\nu = 3$ for 2Stelnet leads to an estimator which performs remarkably well even if there is a mismatch between the real distribution of data and the one assumed by the estimator (i.e. multivariate t-Student with 3 degrees of freedom). Furthermore, additional simulations with highly correlated data suggest that the elastic net penalty (i.e. $\alpha = 0.5$) leads to better results than LASSO (i.e. $\alpha = 1$).

Despite the good behaviour of 2Stelnet in low-dimensional settings ($n > p$), severe limitations can arise in high-dimensional situations ($n < p$) where the existence and uniqueness of the estimator in the second stage (9) is not guaranteed.

Finally, an empirical application is proposed where 2Stelnet is used to estimate the European banking network from the stock prices of a large set of European banks. We show the impact of the Covid-19 pandemic on network strength, which is an indicator of potential crisis periods. Different centrality measures are also used to detect the most central banks in the network. To conclude our empirical analysis, we evaluate the effects of a shock in the most and least central banks, according to strength, by using the 2020 partial correlation network; not surprisingly, we found much larger effects if the shock hits the most central bank, suggesting that the degree of interconnectedness plays an important role and needs to be taken into account when setting up adequate risk management and risk mitigation tools.

**Notes**

1. Only existing paths are considered for average distance. The eccentricity of individual nodes is set to 0.
2. For weighted graphs, it is common also to consider the weights directly and thus calculate a weighted degree. Here strength corresponds to this weighted degree measure.
3. Note that in a partial correlation network positive and negative values could average out. One can use absolute values instead, as suggested in [37]. Given the small percentage of negative partial correlation, 2%, 10% and 8% in 2018, 2019 and 2020, respectively, we leave it for further research.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).
ORCID

Sandra Paterlini  http://orcid.org/0000-0003-4269-4496
Emanuele Taufer  http://orcid.org/0000-0001-5140-9912

References

[1] Lauritzen S. Graphical models. New York, NY: Oxford University Press; 1996.
[2] Koller D, Friedman N. Probabilistic graphical models: principles and techniques. Cambridge, MA: MIT Press; 2009.
[3] Banerjee O, El Ghaoui L, d’Aspremont A. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. J Mach Learn Res. 2008;9:485–516.
[4] Cai T, Liu W, Luo X. A constrained L1 minimization approach to sparse precision matrix estimation. J Am Stat Assoc. 2011;106(494):594–607.
[5] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9(3):432–441.
[6] Lee S, Sobczyk P, Bogdan M. Structure learning of Gaussian Markov random fields with false discovery rate control. Symmetry. 2019;11:1311.
[7] Mazumder R, Hastie T. The graphical lasso: new insights and alternatives. Electron J Stat. 2012;6:2125–2149.
[8] Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. Ann Stat. 2006;34(3):1436–1462.
[9] Yuan M. High dimensional inverse covariance matrix estimation via linear programming. J Mach Learn Res. 2010;11:2261–2286.
[10] Zhou S, Rütimann P, Xu M, et al. High-dimensional covariance estimation based on gaussian graphical models. J Mach Learn Res. 2011;12:2975–3026.
[11] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B (Methodol). 1996;58(1):267–288.
[12] Bernardini D, Paterlini S, Taufer E. New estimation approaches for graphical models with elastic net penalty. Econ Stat. 2022. In Press.
[13] Kovács S, Ruckstuhl T, Obrist H, et al. Graphical elastic net and target matrices: fast algorithms and software for sparse precision matrix estimation; 2021. Available from: arXiv:2101.02148.
[14] Zou H, Hastie T. Regularization and variable selection via the elastic-net. J R Stat Soc B. 2005;67:301–320.
[15] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B (Methodol). 1977;39:1–38.
[16] Finegold M, Drton M. Robust graphical modeling of gene networks using classical and alternative t-distributions. Ann Appl Stat. 2011;5:1057–1080.
[17] Cont R. Empirical properties of asset returns: stylized facts and statistical issues. Quant Finance. 2001;1:223–236.
[18] Jackson MO. Networks and economic behavior. Annu Rev Econ. 2009;1:489–511.
[19] Carvalho VM, Tahbaz-Salehi A. Production networks: a primer. Annu Rev Econ. 2019;11:635–663.
[20] Diebold FX, Yilmaz K. Financial and macroeconomic connectedness: a network approach to measurement and monitoring. New York, NY: Oxford University Press; 2015.
[21] Jackson MO, Pernoud A. Systemic risk in financial networks: a survey; 2020. Available from: SSRN: http://dx.doi.org/10.2139/ssrn.3651864.
[22] Bardoscia M, Barucca P, Battiston S, et al. The physics of financial networks. Nat Rev Phys. 2021;3:490–507.
[23] Barigozzi M, Brownlees C. Nets: network estimation for time series. J Appl Econom. 2019;34:347–364.
[24] Billio M, Getmansky M, Lo AW, et al. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J Financ Econ. 2012;104:535–559.
[25] Demirer M, Diebold FX, Liu L, et al. Estimating global bank network connectedness. J Appl Econom. 2018;33:1–15.
[26] Torri G, Giacometti R, Paterlini S. Robust and sparse banking network estimation. E J Oper Res. 2018;270(1):51–65.
[27] Baba K, Shibata R, Sibuya M. Partial correlation and conditional correlation as measures of conditional independence. Aust N Z J Stat. 2004;46(4):657–664.
[28] Liu C, Rubin DB. ML estimation of the t distribution using EM and its extensions, ECM and ECME. Stat Sin. 1995;5:19–39.
[29] Liu C. ML estimation of the multivariate t distribution and the EM algorithm. J Multivar Anal. 1997;63:296–312.
[30] Doğru FZ, Bulut YM, Arslan O. Double reweighted estimators for the parameters of the multivariate distribution; 2017. Available from: arXiv:1707.01130.
[31] Hasannasab M, Hertrich J, Laus F, et al. Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the student t distribution. Numer Algorithms. 2021;87:77–118.
[32] Pascal F, Ollila E, Palomar DP. Improved estimation of the degree of freedom parameter of multivariate t-distribution. In: 2021 29th European Signal Processing Conference (EUSIPCO); 2021. p. 860–864.
[33] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2009.
[34] Foygel R, Drton M. Extended Bayesian information criteria for gaussian graphical models. In: Advances in Neural Information Processing Systems (NIPS 2010); Vol. 23; 2010. p. 604–612.
[35] Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw – Pract Exp. 1991;11:1129–1164.
[36] Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. Nature. 1998;393: 440–442.
[37] Anufriev M, Panchenko V. Connecting the dots: econometric methods for uncovering networks with an application to the Australian financial institutions. J Bank Finance. 2015;61:241–255.
A Synthetic precision matrices

Figure 7. Precision matrices of simulated data (diagonal elements ignored)

B Computational times

Table 2. Computational times (in seconds) for the estimation of the European banking network in the period 2018-2020. Values reported in the table are relative to the whole BIC calibration procedure used to find the optimal value of $\lambda$ from 100 exponentially spaced values between $e^{-6}$ and 1.5. Column name indicates the underlying distribution assumed by estimators (glasso, 2Sglasso, 2Sgelnet for normal, lasso, 2Stlasso, 2Stelnet for t-Student with different degrees of freedom).

|                | Normal | t-Stud (v = 3) | t-Stud (v = 15) | t-Stud (v = 30) |
|----------------|--------|----------------|-----------------|-----------------|
| glasso/lasso    | 10.62  | 993.79         | 619.72          | 378.92          |
| 2Sglasso/2Stlasso | 5.48   | 1049.42        | 645.55          | 364.72          |
| 2Sgelnet/2Stelnet | 5.54   | 1515.49        | 773.76          | 838.57          |
### Table 3. List of banks with KS-tests ($\alpha = 0.05$) on residuals (R: rejected - NR: not rejected)

| Bank                          | ShortName     | Normal | t-Student |
|-------------------------------|---------------|--------|-----------|
| HSBC Holdings PLC             | HSBC          | R      | NR        |
| BNP Paribas SA                | BNP Paribas   | R      | NR        |
| Credit Agricole SA            | Cr. Agricole  | R      | NR        |
| Banco Santander SA            | B. Santander  | R      | NR        |
| Societe Generale SA           | Soc. General  | R      | NR        |
| Barclays PLC                  | Barclays      | R      | NR        |
| Lloyds Banking Group PLC      | Lloyds        | R      | NR        |
| ING Groep NV                  | ING Group     | R      | NR        |
| UniCredit SpA                 | Unicredit     | R      | NR        |
| Natwest Group PLC             | Natwest       | R      | NR        |
| Intesa Sanpaolo SpA           | Intesa SP     | R      | NR        |
| Banco Bilbao Vizcaya Argentaria SA | B. Bilbao VA | R      | NR        |
| Standard Chartered PLC        | Std. Chartered| R      | NR        |
| Danske Bank A/S               | Danske Bank   | R      | NR        |
| Commerzbank AG                | Commerzbank   | R      | NR        |
| Svenska Handelsbanken AB      | Svenska       | R      | NR        |
| KBC Groep NV                  | KBC Group     | R      | NR        |
| Skandinaviska Enskilda Banken AB | Skandinaviska | R      | NR        |
| Dnb ASA                       | Dnb           | R      | NR        |
| Erste Group Bank AG           | Erste Group   | R      | NR        |
| Swedbank AB                   | Swedenbank    | R      | NR        |
| Banco de Sabadell SA          | B. Sabadell   | R      | NR        |
| Raiffeisen Bank International AG | Raiffeisen   | R      | NR        |
| Banca Monte dei Paschi di Siena SpA | Mt. Paschi  | R      | NR        |
| Bank of Ireland Group PLC     | B. Ireland    | R      | NR        |
| AIB Group plc                 | AIB Group     | R      | NR        |
| Jyske Bank A/S                | Jyske Bank    | R      | NR        |
| National Bank of Greece SA    | N.B. of Greece| R      | NR        |
| Komercni Banka as             | Komercni      | R      | NR        |
| Banco BPM SpA                 | Banco BPM     | R      | NR        |
| Nordea Bank Abp               | Nordea Bank   | R      | NR        |
| Deutsche Bank AG              | Deutsche Bank | R      | R         |
| UBI Banca SPA                 | UBI           | R      | NR        |
| Credit Suisse Group           | Credit Suisse | R      | NR        |
| UBS Group AG                  | UBS Group     | R      | NR        |
| Natixis                       | Natixis       | R      | NR        |
Figure 8. Estimated partial correlation networks
Figure 9. Degree distribution (median: red line - mean: blue dashed line)

Figure 10. Evolution of strength measure for 2Stelnet estimators assuming different degrees of freedom ($\nu = 3, 15, 30$) and different values of $\alpha = 0.25, 0.5, 0.75, 1$. 
**D Online supplement**

Additional distributions considered (with $p = 50$):

- Multivariate t-Student: $t_p(0, \frac{v^{-2} \Theta^{-1}}{v}, v)$, with $v = 20$.
- Contaminated normal: $\mathcal{N}_p(0, \Theta^{-1}) \cdot Ber + \mathcal{N}_p(0, \text{diag}(\Theta^{-1})) \cdot (1 - Ber)$, with $Ber \sim \text{Bernoulli}(pd = 0.85)$.

Additional networks considered:

**Figure 11.** Precision matrices of simulated data (diagonal elements ignored)
Figure 12. Normal distribution - Performance measures ($F_1$-score, left column; Frobenius distance, right column) - From top to bottom: hub, small-world, band
Figure 13. t-Student distribution $v = 3$ - Performance measures ($F_1$-score, left column; Frobenius distance, right column) - From top to bottom: hub, small-world, band
Figure 14. t-Student distribution $(v = 20)$ - Performance measures (F₁-score, left column; Frobenius distance, right column) - From top to bottom: scale-free, random, cluster, core-periphery
Figure 15. t-Student distribution $v = 20$ - Performance measures ($F_1$-score, left column; Frobenius distance, right column) - From top to bottom: hub, small-world, band
Figure 16. Contaminated normal - Performance measures (F1-score, left column; Frobenius distance, right column) - From top to bottom: scale-free, random, cluster, core-periphery
Figure 17. Contaminated normal - Performance measures (F1-score, left column; Frobenius distance, right column) - From top to bottom: hub, small-world, band