

Robustly fitting Gaussian graphical models: the R-package robFitConGraph

Daniel Vogel*, Stuart J. Watt, and Anna Wiedemann

April 8, 2022

Abstract

This paper gives a tutorial-style introduction to the R-package robFitConGraph, which provides a robust goodness-of-fit test for Gaussian graphical models. Its use is demonstrated at a data example on music performance anxiety, which also illustrates *why* one would want to fit a Gaussian graphical model – and why one should do so robustly. The underlying theory is briefly explained.

Keywords: covariance selection model, deviance test, M-estimator, music performance anxiety, partial correlation

1 Introduction

The first two Sections 1 *Introduction* and 2 *A Case Study* are intended for a general audience, assuming neither deeper familiarity with graphical models nor robustness. Section 3 *Background and Theory* discusses some aspects in detail.

Gaussian graphical modeling

Graphical models are an important tool for analyzing the dependence structure of several variables. Gaussian graphical models are employed for continuously distributed data, where a multivariate normal, or Gaussian, distribution is adequate.

A graph encodes the dependence structure of a random vector $\mathbf{X} = (X_1, \dots, X_p)$ as follows: The nodes represent the individual variables, and an edge between two nodes represents partial correlation, or the absence of an edge zero partial correlation. Partial correlation is a measure for conditional dependence between two variables, conditional on all remaining variables, which, in a very loose sense, can be understood as a measure for dependence *not* explained by joint dependence on other variables. As soon as more than two variables are considered, conditional dependence is arguably more important than marginal dependence: It suggests and helps to verify causal relationships. While ice cream sales and criminal assault rates per day – at any given place in the temperate climate zone –

*Corresponding author: daniel.vogel@tu-dortmund.de

are certainly positively correlated, their conditional independence given a suitable mediator variable, such as outside mean temperature, provides evidence for why this may be so.

Graphical modeling refers to the statistical task of finding an appropriate graph that describes the dependence structure of a given data set, i. e., identifying all zero partial correlations. The aim is to find a parsimonious graph, i. e., one with few edges, which does not contradict the data. A full graph, with all edges present, means no restriction and contains no structural information. A completely empty graph, with no edges at all, means all variables are independent.

Let Σ denote the covariance matrix of \mathbf{X} . Throughout, we assume that all second moments of \mathbf{X} are finite, and that furthermore Σ is positive definite and thus may be inverted to yield the *concentration matrix* or *precision matrix* $K = \Sigma^{-1}$. The assumption of strict positive definiteness is a mild one and is equivalent to the probability mass not being concentrated on a lower-dimensional affine linear subspace of \mathbb{R}^p . We make the same assumption on the data points $\mathbb{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$, and hence the sample covariance matrix $\hat{\Sigma}_n$ is positive definite and the sample concentration matrix $\hat{K}_n = \hat{\Sigma}_n^{-1}$ exists.

The basis for Gaussian graphical modeling is the following characterization: The variables X_i and X_j are partially uncorrelated given all remaining $p - 2$ variables if and only if $K_{i,j} = 0$. The *partial correlation* $p_{i,j}$ of X_i and X_j given the remaining components of \mathbf{X} is defined as the correlation between the residuals of X_i and X_j when regressing both on the remaining components. Some matrix calculus yields that

$$p_{i,j} = -\frac{K_{i,j}}{\sqrt{K_{i,i}K_{j,j}}}, \quad (1)$$

see, e. g., Whittaker (1990, Chapter 5). Thus the pairwise partial correlations are obtained from the inverse covariance matrix K in a very similar fashion as the pairwise correlations are obtained from the covariance matrix Σ itself. The only difference is the minus sign. Consequently, an absent edge in the graph means a zero entry in K , and finding the graph for given data comes down to finding the zero-pattern in the inverse of the true covariance matrix Σ . Three basic sub-tasks of graphical modeling can be identified:

(T1) Finding an appropriate graph.

(T2) Determining if a given graph fits the data.

(T3) Estimating the (remaining) partial correlations under a given graph structure.

Whether (T1) or (T2) is considered more important may be debatable and mainly depends on whether one pursues an explorative or an inferential analysis. Task (T3) may appear of lesser importance, but it is intrinsically linked to (T2). We briefly outline how these tasks are approached, in reverse order, starting in with (T3).

Task (T3) We require some mathematical notation. Define a graph $G = (V, E)$ as a set of vertices $V = \{1, \dots, p\}$ and a set of undirected edges $E \subseteq \{\{i, j\} : i, j = 1, \dots, p, i < j\}$. Let \mathcal{S}_p denote the set of all symmetric $p \times p$ matrices and \mathcal{S}_p^+ the set of all positive definite, symmetric $p \times p$ matrices. For any graph $G = (V, E)$, let $\mathcal{S}_p^+(G)$ be the set of matrices $A \in \mathcal{S}_p^+$ with zero entries at off-diagonal positions specified by G , i. e., $A_{i,j} = 0$

for all $i, j = 1, \dots, p$, $i \neq j$, with $\{i, j\} \notin E$. We call any set of p -dimensional probability measures with the common property that they possess a concentration matrix $K \in \mathcal{S}_p^+(G)$ a *covariance selection model* induced by G . We call a covariance selection model consisting of all regular p -variate Gaussian distributions a *Gaussian graphical model* and denote it by $N_p(G)$, i. e., $N_p(G) = \{N_p(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^p, \Sigma^{-1} \in \mathcal{S}_p^+(G)\}$. The maximum-likelihood estimator $\hat{\Sigma}_G$ of Σ within the parametric family $N_p(G)$ is given as the solution to

$$\hat{\Sigma}_G = \arg \min_{\Sigma^{-1} \in \mathcal{S}_p^+(G)} \left\{ \log \det \Sigma + \frac{1}{n} \sum_{i=1}^n \text{trace} \left(\hat{\Sigma}_n \Sigma^{-1} \right) \right\}. \quad (2)$$

This optimization problem leads to the estimation equations

$$\begin{cases} \left[\hat{\Sigma}_G \right]_{i,j} = \left[\hat{\Sigma}_n \right]_{i,j} & \text{for } \{i, j\} \in E \text{ or } i = j, \\ \left[\hat{\Sigma}_G^{-1} \right]_{i,j} = 0 & \text{for } \{i, j\} \notin E \text{ and } i \neq j. \end{cases}$$

The solution $\hat{\Sigma}_G$ depends on the data only through the sample covariance matrix $\hat{\Sigma}_n$. This approach is due to Dempster (1972), and the optimization problem (2) has since been thoroughly studied (e. g. Speed and Kiiveri, 1986). Algorithms to compute $\hat{\Sigma}_G$ for arbitrary graphs G can be found, e. g., in Lauritzen (1996, Chapter 5) or Hastie et al. (2009, Chapter 17). For decomposable graphs G , there is also an explicit solution, i. e., $\hat{\Sigma}_G$ can be computed in a finite number of steps. For details see also Lauritzen (1996, Chapter 5). In R, these algorithms are implemented in the function `fitConGraph` in the package `ggm` (Marchetti et al., 2020). With a graph-constrained¹ estimate $\hat{\Sigma}_G$ for the covariance matrix available, the remaining non-zero partial correlations are computed from $\hat{\Sigma}_G$ as unconstrained partial correlations are computed from $\hat{\Sigma}_n$ by virtue of (1).

Task (T2) Within the parametric framework described above, the likelihood-ratio test for testing G against the full model is given by the test statistic

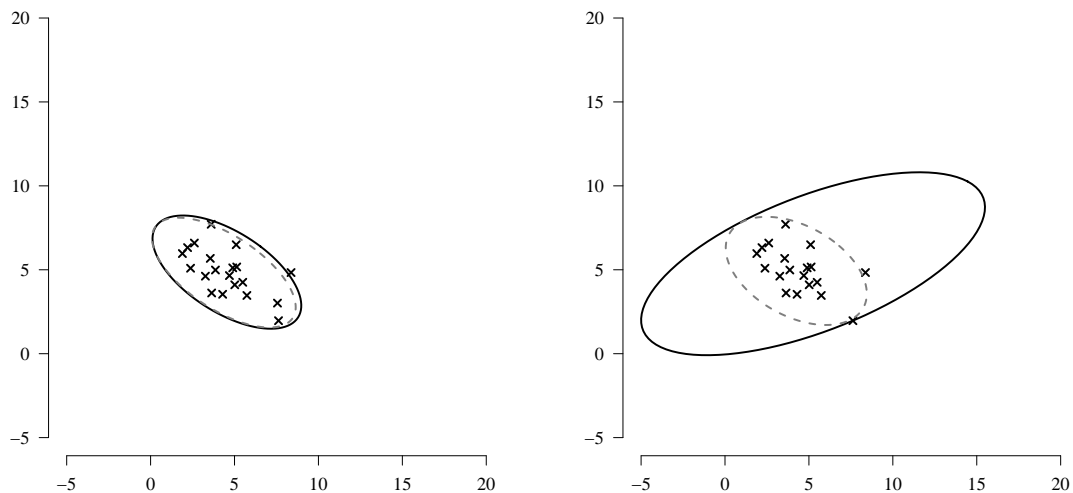
$$D_n^\Sigma(G) = n \left(\log \det \hat{\Sigma}_G - \log \det \hat{\Sigma}_n \right), \quad (3)$$

which, under the null hypothesis that G is the true graph, converges to a χ_q^2 distribution as $n \rightarrow \infty$, where q is the number of missing edges in G . The quantity D_n^Σ is also called *deviance* and this likelihood-ratio test hence *deviance test*. It simultaneously tests the absence of all edges not in G avoiding any multiple-testing problems. The deviance is also returned by `fitConGraph`.

Task (T1) While statistical theory provides rather precise and unambiguous solutions to tasks (T3) and (T2), this is not the case for (T1), which is already reflected by the phrasing of *finding an appropriate graph* rather than *finding the best fitting graph*. Deciding on an appropriate graph may also be influenced by interpretability aspects and relevant domain knowledge. A multitude of approaches exist. A basic idea, also initiated by Dempster

¹i. e., it obeys the zero-pattern in the inverse induced by G

Figure 1: Robust vs non-robust scatter estimator



(1972) is the iterative application of the deviance test. For instance, one starts with the full graph, then removes one or several edges (with small absolute partial correlations), and keeps the new candidate graph if the deviance test accepts it. This may be iterated until no further edge removal leads to an accepted graph. The opposite search direction is also possible: One starts with the empty graph and successively adds edges until a graph is obtained which is accepted by the deviance test. Elaborate model search strategies have been proposed (e. g. Edwards and Havránek, 1985; Drton and Perlman, 2008).

Many other model selection approaches use L_1 -regularization and are aimed at finding sparse graphs in high-dimensional settings. For instance, Meinshausen and Bühlmann (2006) propose a node-wise LASSO-regression. Yuan and Lin (2007) and Friedman et al. (2008) add an L_1 -penalty for $K = \Sigma^{-1}$ to the optimization problem (2). Within this framework, the regularization parameter must be chosen, usually by means of cross validation.

Robustness

Robustness in general terms is the property of a statistical method to yield sensible results if its assumptions are violated. In a more specific sense, it means insensitivity to outliers. Starting with the pioneering work of Huber (1964) and Hampel (1971, 1974), robust statistics has evolved into a large research area, see, e. g., the textbooks by Maronna et al. (2006) or Huber and Ronchetti (2009). For our purposes, the important fact to note is that the sample covariance matrix $\hat{\Sigma}_n$ is not robust. In Figure 1, the left-hand panel shows a small data set of 20 two-dimensional observations. The black ellipse visualizes the sample covariance matrix, i. e., the 95% probability ellipse of the thus fitted normal model. In the right-hand panel, one single observation has been moved from the center to the upper right corner. The covariance estimate has tremendously changed, suggesting even a positive rather than a negative correlation. The dashed curve, in contrast, represents an alternative,

robust estimator of multivariate scatter (a t_3 M-estimator, see below) and is little altered by the outlier.

With the sample covariance matrix $\hat{\Sigma}_n$ being the main ingredient of essentially all graphical modeling tasks, they all inherit its lack of robustness. The good news is: Robust alternatives exist. The work on robust multivariate location and scatter estimation as been originated by Maronna (1976), who developed Huber’s M-estimation approach for the multivariate setting. Since then, many proposals have been made. Here we consider only one rather simple and easy-to-compute robust scatter estimator, which is hence popular with practitioners: the t_ν M-estimator. It is an M-estimator, whose loss function stems from the maximum-likelihood estimator within the elliptical t_ν -model. The parameter ν is referred to as *degrees of freedom* and may be any positive real number. The smaller the ν , the heavier-tailed the t_ν distribution and, consequently, the more outlier-resistant the corresponding M-estimator. The parameter ν is usually not inferred from the data, but selected by the data analyst. A common choice is $\nu = 3$.

The t_ν M-estimator of scatter \hat{S}_n , along with the corresponding estimate of location $\hat{\boldsymbol{\mu}}_n$, is defined as the solution to the optimization problem

$$(\hat{\boldsymbol{\mu}}_n, \hat{S}_n) = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^p, S \in \mathcal{S}_p^+} \left[\sum_{i=1}^n \rho_{\nu,p} \left\{ (\mathbf{x}_i - \boldsymbol{\mu})^\top S^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} + n \log \det S \right], \quad (4)$$

with $\rho_{\nu,p}(x) = (\nu + p) \log(1 + x/\nu)$. This yields the estimation equations

$$\begin{cases} 0 = \sum_{i=1}^n \psi_{\nu,p}(\hat{r}_i) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n), \\ \hat{S}_n = n^{-1} \sum_{i=1}^n \psi_{\nu,p}(\hat{r}_i) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)^\top, \end{cases}$$

where $\psi_{\nu,p}(x) = \rho'_{\nu,p}(x) = (\nu + p)/(\nu + x)$ and $\hat{r}_i = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)^\top \hat{S}_n^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)$. The t_ν M-estimator can be computed, e. g., by a fixed-point algorithm. Implementations in R can be found, e. g., in the functions `cov.trob` from the package MASS (Venables and Ripley, 2002), `tM` from the package ICS (Nordhausen et al., 2008), and `MVTMLE` from the package fastM (Dümbgen et al., 2016, 2018). The latter uses a partial Newton-Raphson algorithm.

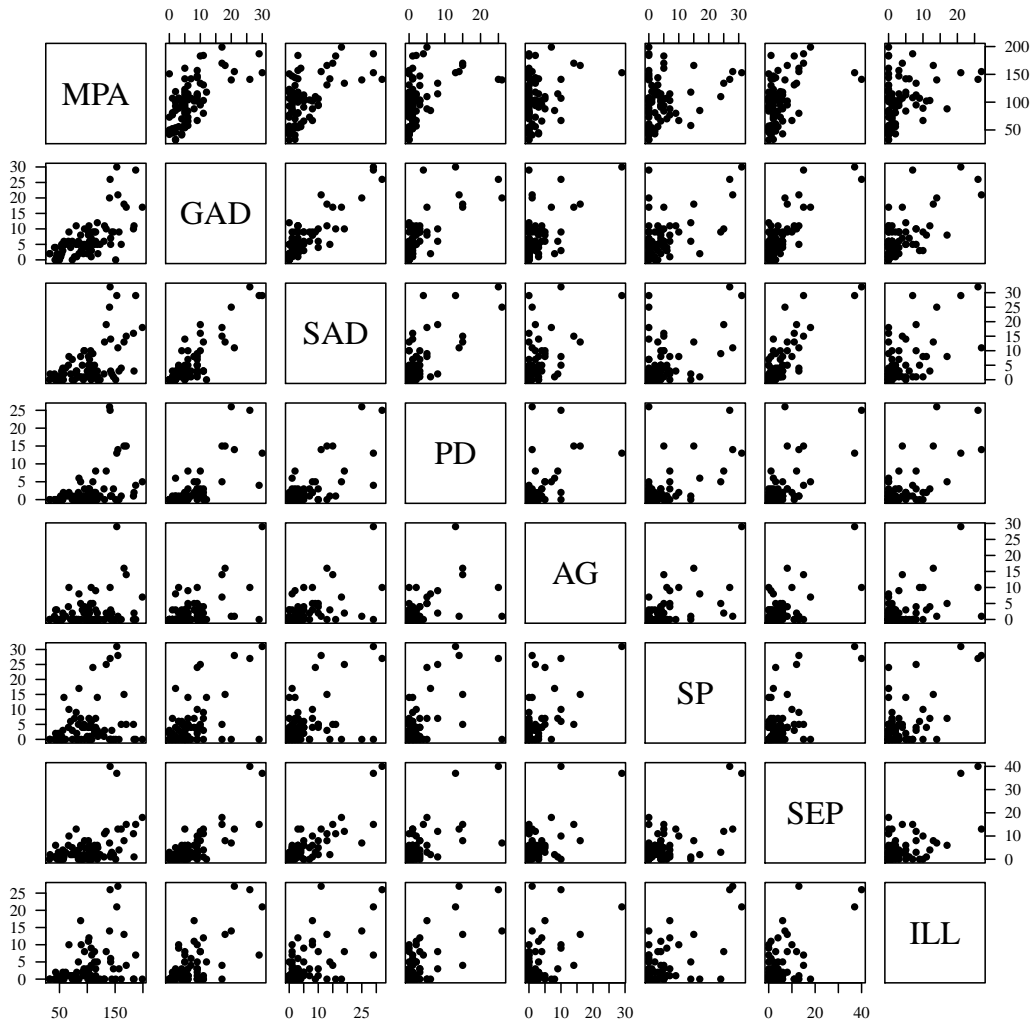
With robust alternatives \hat{S}_n being available, an intuitive path to a robust analysis is the plug-in approach: First solving (4) and then plugging the thus obtained estimate \hat{S}_n instead of $\hat{\Sigma}_n$ into (2) to obtain a graph-constrained robust estimate \hat{S}_G and the corresponding robust (pseudo-)deviance

$$D_n^S(G) = n \left(\log \det \hat{S}_G - \log \det \hat{S}_n \right). \quad (5)$$

This is accomplished in the package `robFitConGraph` by the function of the same name. Plug-in robustifications for the L_1 -regularization methods have equally been proposed (Öllerer and Croux, 2015; Tarr et al., 2016). Alternatively, Finegold and Drton (2011) regularize the elliptical t_ν -log-density.

Before providing further details on the function `robFitConGraph` and the underlying theory in Section 3, its use shall be demonstrated at a data example on music performance anxiety.

Figure 2: Pairwise scatter plots of the anxieties data



2 A Case Study: Music Performance Anxiety

The fear about one’s ability to perform a specific task, such as giving a presentation or sitting an exam, affects almost everyone. Pressure can be particularly high in certain professions where performing in front of others is an integral part of day-to-day life. While some levels of stress and anxiety are normal and actually help us to achieve optimal performance, severe levels of stress and anxiety are debilitating and can develop into a disorder. Professional musicians are often exposed to extreme pressure where maintaining top quality performances are not just essential to keeping their job, but to progress in their careers. Music performance anxiety (MPA) can be understood as a continuum ranging from low to high anxiety levels. The latter poses a serious problem to the profession and is the subject of ongoing clinical research, see, e. g. Fernholz et al. (2019) for a recent review.

MPA is often considered to be a form of social anxiety which, loosely speaking, is the

overwhelming fear of social situations (e. g. Kenny, 2011; Cox and Kenardy, 1993; Nicholson et al., 2015; Dobos et al., 2019). This is underlined by the fact that the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) now acknowledges evidence of individuals suffering exclusively from performance anxiety as a distinct sub-type of social anxiety disorder. However, some researchers and clinicians have questioned this description, as MPA is a complex phenomenon caused by the interaction of many different factors, and the fear of social judgment is not necessarily always the main problem.

Wiedemann et al. (2022) analyzed a data set consisting of eight numerical variables measured at $n = 82$ students at German music colleges: music performance anxiety (MPA), agoraphobia (AG), generalized anxiety disorder (GAD), panic disorder (PD), separation anxiety disorder (SEP), specific phobia (SP), social anxiety disorder (SAD) as well as illness anxiety disorder (ILL). Each variable is a summary score from a self-assessment inventory with Likert-scale items. A higher value signifies a higher severity of the condition. MPA was assessed using the German version of the Kenny Music Performance Anxiety Inventory (K-MPAI; Kenny, 2009) translated by Spahn et al. (2016). All other anxieties were assessed using the German translation of the disorder-specific anxiety measures (Beesdo-Baum et al., 2012; Lebeau et al., 2012) for the dimensional anxiety scales of the DSM-5. The data set is also included in the package `robFitConGraph` as `anxieties`.

Figure 2 shows the pairwise scatter plots of the data set. While most participants score low on most anxiety scales – as we would expect and hope –, there are a few very high values in all variables. The normality assumption can be seen to be violated in a similar manner as in Figure 1 with the same implications for any sample-covariance-based analysis. For instance, removing the outlier in the variable AG reduces the Pearson correlation between AG and SEP from 0.612 to 0.407. Computing the correlation coefficient from a t_3 M-estimator, we obtain the value 0.410, which reduces to 0.350 when removing said outlier. A robust analysis is highly recommended for this data. All results reported in the following are based on a t_3 M-estimator of scatter. It yields the correlation coefficients given in Table 1. They can also be obtained by the function `robFitConGraph` by supplying the full model as adjacency matrix.

```
> p <- ncol(anxieties)
> Shat <- robFitConGraph(X = anxieties,
+                       amat = matrix(1, ncol = p, nrow = p),
+                       df_est = 3)$Shat
> cov2cor(Shat)
```

All variables are positively correlated. This is neither surprising nor uncommon for multivariate data. The corresponding partial correlations, as given in Table 2, shed further light on the dependence structure of the variables.

```
> Phat <- -cov2cor(solve(Shat))
> diag(Phat) <- 1
> Phat
```

Many of the partial correlations are near zero, suggesting conditional independences, i. e., their association may be fully mediated by other variables in the data set.

Table 1: Pairwise correlations computed from t_3 M-estimate

	MPA	GAD	SAD	PD	AG	SP	SEP	ILL
MPA	.	0.62	0.37	0.43	0.17	0.26	0.38	0.32
GAD	0.62	.	0.66	0.65	0.40	0.35	0.60	0.47
SAD	0.37	0.66	.	0.50	0.51	0.36	0.64	0.42
PD	0.43	0.65	0.50	.	0.51	0.44	0.48	0.46
AG	0.17	0.40	0.51	0.51	.	0.49	0.41	0.36
SP	0.26	0.35	0.36	0.44	0.49	.	0.37	0.34
SEP	0.38	0.60	0.64	0.48	0.41	0.37	.	0.29
ILL	0.32	0.47	0.42	0.46	0.36	0.34	0.29	.

Table 2: Pairwise *partial* correlations computed from t_3 M-estimate

	MPA	GAD	SAD	PD	AG	SP	SEP	ILL
MPA	.	0.44	-0.05	0.05	-0.14	0.08	0.05	0.05
GAD	0.44	.	0.33	0.34	-0.02	-0.06	0.19	0.15
SAD	-0.05	0.33	.	-0.05	0.25	0.00	0.37	0.13
PD	0.05	0.34	-0.05	.	0.24	0.14	0.09	0.16
AG	-0.14	-0.02	0.25	0.24	.	0.30	0.04	0.07
SP	0.08	-0.06	0.00	0.14	0.30	.	0.11	0.12
SEP	0.05	0.19	0.37	0.09	0.04	0.11	.	-0.12
ILL	0.05	0.15	0.13	0.16	0.07	0.12	-0.12	.

Inferential analysis: MPA and social anxiety

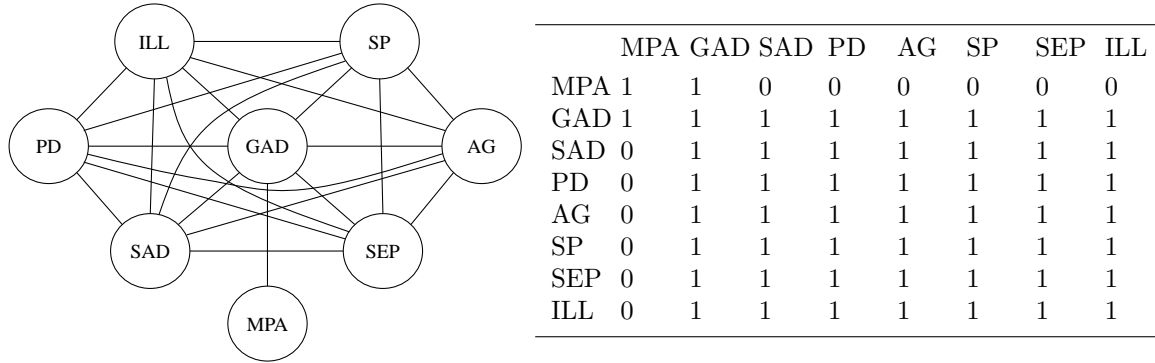
One hypothesis examined by Wiedemann et al. (2022) is whether MPA is primarily related to a social anxiety disorder (SAD). Based on the positive correlation of 0.37, a short-sighted analysis may deduce a strong connection between MPA and SAD, which in light of other, even larger correlations, is right away questionable. As we will see, the data in fact carries no evidence for a particularly strong connection between MPA and SAD.

For that purpose one tests the hypothesis that MPA and SAD are conditionally independent given all remaining six variables, i. e., testing a graphical model with only one missing edge between MPA and SAD by means a robust pseudo-deviance test. If this test rejects, there is evidence for a strong link between MPA and SAD. Their dependence can then not be fully explained by their associations with other anxiety types. However, this hypothesis is *not* rejected (a p-value of 0.66), and we find no such evidence in the data.

One can even further test the stronger hypothesis that MPA is, given GAD only, conditionally independent of all other six specific anxiety scales. This is done by testing the graph in Figure 3, which can equivalently be expressed by its adjacency matrix on the right-hand side of Figure 3.

Almost unnoticed we make here use of a non-trivial result, which is the real merit and the real beauty of graphical models: Each absent edge in the graph denotes a conditional independence of two individual variables given the respective remaining six variables. This

Figure 3: The graph encoding the hypothesis MPA is conditionally independent of all other variables given GAD (left) and the corresponding adjacency matrix (right).



is indeed equivalent to MPA and (SAD, PD, AG, SP, SEP, ILL) being conditionally independent given GAD only, because, in the graph, GAD separates MPA from (SAD, PD, AG, SP, SEP, ILL). This is known as the equivalence between the local Markov property and the global Markov property. For details see Lauritzen (1996, Chapter 3).

The pseudo-deviance test based on the t_3 M-estimator for the graph in Figure 3 is carried out as follows:

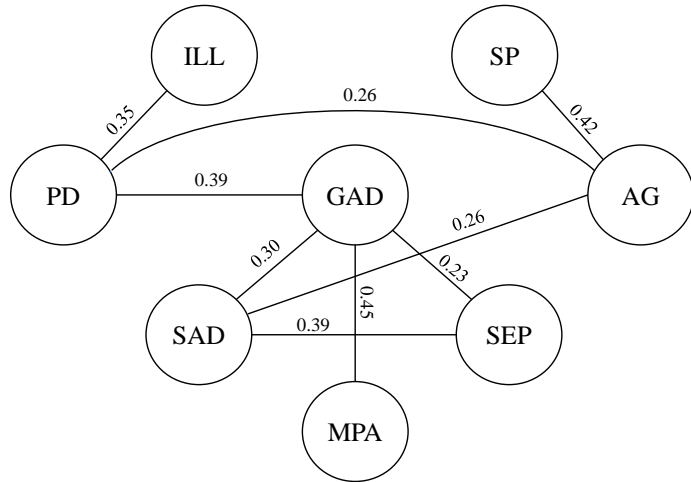
```
> amat <- matrix(1, ncol = p, nrow = p)
> amat["MPA", c("SAD", "PD", "AG", "SP", "SEP", "ILL")] <- 0
> amat[c("SAD", "PD", "AG", "SP", "SEP", "ILL"), "MPA"] <- 0
> robFitConGraph(X = anxieties, amat = amat, df_est = 3)$pval
[1] 0.881159
```

With a p-value of 0.88, this hypothesis is also *not* rejected. There is no evidence against the null hypothesis of MPA being conditionally independent of the specific anxiety types given generalized anxiety (GAD). Accepting the null hypothesis, GAD is fully sufficient for predicting MPA. When already knowing a person's GAD score, i. e., how anxious the person generally is, additionally knowing any of their specific anxieties scores provides no further information about their MPA score. So this analysis does not support the hypothesis that MPA is foremost related to social anxiety.

The analysis is complemented by testing if MPA and GAD are conditionally independent given the remaining six variables. This hypothesis *is* rejected with a p-value below 0.01, which corresponds to the partial correlation of 0.44 in Table 2.² So there is a strong connection between MPA and GAD, this connections is not mediated by any of the other variables, and GAD explains the connection between MPA and the remaining variables. Altogether, GAD is the link between MPA and the other anxieties.

²For a single missing edge, the deviance test can indeed be expressed in terms of corresponding partial correlation only.

Figure 4: A fitting graph with 9 edges based on a t_3 scatter estimator with p-value 0.45.



Explorative analysis

Naturally, the question arises which other edges may be removed. For that purpose we remove all edges which have absolute partial correlation below 0.15 (Table 2), corresponding to an individual p-value above 0.2. The resulting graph is depicted in Figure 4, which is not rejected by the pseudo-deviance test with a p-value of 0.45.

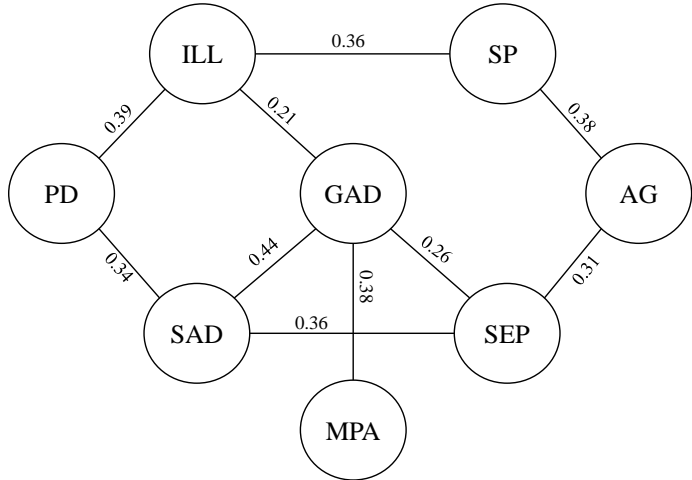
The partial correlations along the edges in Figure 4 are fitted under the graph, i. e., they are estimated taking the graph structure into account. They are different from the partial correlations given in Table 2.

```
> amat <- abs(Phat) > 0.15
> Shat_G <- robFitConGraph(X = anxieties,
+                          amat = amat, df_est = 3)$Shat
> Phat_G <- -cov2cor(solve(Shat))
> diag(Phat_G) <- 1
```

The information learned from such a fitted graph is, e. g., that GAD is central within this set of variables: It has many edges to other vertices, it has much predictive power about the other variables. If one were to retain only a single variable (as a very simple dimension reduction approach, say), GAD would be a natural candidate based on this “vertex degree criterion”.

The fitted graph in Figure 4 with 9 edges (out of 28 possible) is one parsimonious graph that fits the data. Generally, there is no well-defined *most parsimonious*, *best-fitting* graph, as obviously fit and parsimony are contradicting goals. However, one may elaborate upon the initial search by checking if other (equally or more) parsimonious graphs may fit as well. For instance, increasing the initial partial correlation threshold to 0.17, say, which results in a further removal of the edge ILL–PD, leads to a p-value of 0.01 and hence should be rejected. Alternatively, removing the “next smallest edge” from the candidate graph in Figure 4, which is GAD–SEP, results in a p-value of 0.17, which is still acceptable.

Figure 5: A fitting graph with 10 edges based on the sample covariance matrix with p-value 0.51.



So far, we have used `robFitConGraph` with `df_est = 3`, which is also the default setting in case `df_est` is not specified.

Generally, the results are not very sensitive to variations in ν . A smaller value of ν downweights outliers more strongly. In the present example, taking $\nu = 1$, we find the hypothesis of Figure 3 equally accepted with a p-value of 0.58. The explorative graph in Figure 4 is accepted with a p-value of 0.26.

The classical analysis

After having cautioned its use at the beginning, we close this section by remarking that a classical sample-covariance-based analysis leads to qualitatively the same findings. The classical deviance test gives a p-value of 0.63 for the hypothesized graph of Figure 3. An explorative analysis analogous to the one above leads to the graph in Figure 5 with a p-value of 0.51, which is also communicated by Wiedemann et al. (2022). The graph is different, but it equally shows the centrality of GAD as well as the marginality of MPA within the set of variables. It is generally advisable to try varying parameter settings and different methods, robust and non-robust, in any statistical analysis. Conclusions unanimously obtained by several methods may be considered even more trustworthy.

3 Background and theory

The centerpiece of the package `robFitConGraph` is the function by the same name. For a given data set \mathbb{X}_n and a given graph G , it simultaneously provides a robust graph-constrained matrix \hat{S}_G and the p-value of the corresponding pseudo-deviance test. But so far, the actual worth of the function `robFitConGraph` has not become apparent: It appears, its plug-in functionality is equally achieved by a combination of, say, `cov.trob` and

`fitConGraph`, as the latter takes the sample covariance matrix as input. However, there are two good reasons for `robFitConGraph`, which both require a slightly deeper foray into statistical theory.

- (1) The limit distribution of the pseudo-deviance $D_n^S(G)$ does not follow strictly a χ_q^2 -distribution under the graph G , but a χ_q^2 -variate multiplied by a constant σ_1 . This constant is required to obtain p-values and depends on the data dimension p and the degrees of freedom ν of the estimator. It is computed by the function `find_sigma1`.
- (2) The plug-in estimator \hat{S}_G is just one approach to a graph-constrained robust scatter estimator. The function `robFitConGraph` also provides an alternative approach, referred to as the *direct estimator* and denoted by \tilde{S}_G below.

Another good reason is speed. Being implemented in C++, `robFitConGraph` is considerably faster than the combination of `cov.trob` and `fitConGraph`. A runtime comparison is given by Watt (2019).

The Constant σ_1

Recall the class of all p -dimensional, continuous, elliptical distributions, i. e., distributions possessing a p -dimensional Lebesgue density f of the form

$$f(\mathbf{x}) = \det(S)^{-\frac{1}{2}} g\{(\mathbf{x} - \boldsymbol{\mu})^\top S^{-1}(\mathbf{x} - \boldsymbol{\mu})\} \quad (6)$$

for some $\boldsymbol{\mu} \in \mathbb{R}^p$, $S \in \mathcal{S}_p^+$ and $g : [0, \infty) \rightarrow [0, \infty)$, such that f integrates to 1. Let $E_p(\boldsymbol{\mu}, S, g)$ denote the distribution described by (6). The univariate function g is called the *elliptical generator* and S the *scatter* or *shape matrix* of $E_p(\boldsymbol{\mu}, S, g)$. It is proportional to the covariance matrix if $E_p(\boldsymbol{\mu}, S, g)$ has second moments. In the present paper, we only consider two examples of elliptical distributions: the normal distribution with

$$g_N(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x}{2}\right), \quad 0 \leq x,$$

and the elliptical t_ν -distribution with

$$g_{\nu,p}(x) = c_{\nu,p} \left(1 + \frac{x}{\nu}\right)^{-(\nu+p)/2}, \quad 0 \leq x.$$

where the normalizing constant $c_{\nu,p}$ is given in the Technical Appendix at the end of the paper. Assume the independent observations $\mathbb{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ to stem from an elliptical distribution $E_p(\boldsymbol{\mu}, S, g)$ and let \hat{V}_n be an *arbitrary* $p \times p$ scatter matrix estimator fulfilling two fairly natural conditions:

- (1) \hat{V}_n is affine equivalent, i. e., $\hat{V}_n(\mathbb{X}_n A^\top + \mathbf{1}_n \mathbf{b}^\top) = A \hat{S}_n(\mathbb{X}_n) A^\top$ for any $\mathbf{b} \in \mathbb{R}^p$ and full rank $A \in \mathbb{R}^{p \times p}$, where $\mathbf{1}_n$ is the n -vector consisting of ones.
- (2) \hat{V}_n is \sqrt{n} -consistent and asymptotically normal at $E_p(\boldsymbol{\mu}, S, g)$, i. e., there exist matrices $V \in \mathcal{S}_p$ and $W \in \mathcal{S}_{p^2}$ such that

$$\sqrt{n} \text{vec}\{\hat{V}_n(\mathbb{X}_n) - V\} \rightarrow N_{p^2}(\mathbf{0}, W)$$

in distribution as $n \rightarrow \infty$.

Tyler (1982) showed that under these two conditions, V and W take on specific forms: $V = \eta S$ for some $\eta > 0$, and

$$W = 2\eta^2\sigma_1\mathcal{M}_p(S \otimes S) + \eta^2\sigma_2\text{vec}(S)\text{vec}(S)^\top, \quad (7)$$

where \otimes is the Kronecker product, $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/p$ are scalar constants independent of $\boldsymbol{\mu}$ and S , and \mathcal{M}_p is a fixed $p^2 \times p^2$ matrix defined in the Technical Appendix. The latter formula greatly simplifies the asymptotic efficiency comparison of any two affine equivariant scatter estimators at elliptical distributions. In case of \hat{V}_n being an elliptical M-estimator, with a general loss function ρ instead of $\rho_{\nu,p}$ in (4), Tyler (1983) gives the following expressions for the scalars η , σ_1 and σ_2 : Letting $\psi(x) = \rho'(x)$, $\phi(x) = x\psi(x)$, and $R = (\mathbf{X} - \boldsymbol{\mu})^\top S^{-1}(\mathbf{X} - \boldsymbol{\mu})$ for $\mathbf{X} \sim E_p(\boldsymbol{\mu}, S, g)$, the scalar η is the solution to $\mathbf{E}\{\phi(R/\eta)\} = p$. Letting further

$$\gamma_1 = \frac{\mathbf{E}\{\phi^2(R/\eta)\}}{p(p+2)}, \quad \gamma_2 = \frac{1}{p}\mathbf{E}\left\{\frac{R}{\eta}\phi'\left(\frac{R}{\eta}\right)\right\},$$

the scalars σ_1 and σ_2 are

$$\sigma_1 = \frac{(p+2)^2\gamma_1}{(2\gamma_2+p)^2}, \quad \sigma_2 = \gamma_2^{-2}\left[\gamma_1 - 1 - \frac{2\gamma_1(\gamma_2-1)\{p+(p+4)\gamma_2\}}{(2\gamma_2+p)^2}\right].$$

Tyler (1983) further showed that the asymptotic variance of any scale-invariant, continuously differentiable function h of \hat{V}_n only depends σ_1 and not on σ_2 . A *scale-invariant* function $h : \mathcal{S}_p \rightarrow \mathbb{R}$ satisfies $h(\alpha V) = h(V)$ for any $\alpha > 0$. The pseudo-deviance test statistic (5) is such a scale-invariant function of the scatter estimator \hat{S}_n . Dependence is an inherently scale-free concept. So this equally applies to any aspect of multivariate scatter that quantifies dependence in one way or another, may it be correlations, partial correlations, canonical correlations, principal components, etc.

In the package `robFitConGraph`, the functions `find_eta` and `find_sigma1` compute the scalars η and σ_1 , respectively, for t_{ν_1} M-estimators in case the data stem from a normal or an elliptical t_{ν_2} distribution. Note that the degrees of freedom ν_1 of the estimator loss function $\rho_{\nu_1,p}$ and the degrees of freedom of the population distribution ν_2 can be generally different. If they coincide, the t_{ν_1} M-estimator is a maximum-likelihood estimator.

The value of ν_1 is specified by `df_est` and ν_2 by `df_data`. For both, `Inf` is allowed, which corresponds to the sample covariance matrix and the normal distribution, respectively. For `find_sigma1`, the input `df_est = 0` is also allowed. This corresponds to Tyler's distribution-free M-estimator of scatter (Tyler, 1987). In this case $\sigma_1 = 1 + 2/p$ regardless of the elliptical population distribution. There is no t_0 distribution, hence `df_data = 0` is not allowed.

The value σ_1 can be given directly to `robFitConGraph` via the optional argument `sigma1`. If none is provided, `robFitConGraph` calls `find_sigma1` with `df_data = Inf` (i. e., assuming Gaussian data) and its argument `df_est` being passed on to `find_sigma1`. The latter is also optional with the default `df_est = 3`.

The Direct vs. the Plug-in Estimate

Instead of solving (4) and (2) sequentially, an alternative approach is to directly solve the optimization problem

$$(\tilde{\boldsymbol{\mu}}_G, \tilde{S}_G) = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^p, S^{-1} \in \mathcal{S}_p^+(G)} \left[\sum_{i=1}^n \rho_{\nu,p} \left\{ (\mathbf{x}_i - \boldsymbol{\mu})^\top S^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} + n \log \det S \right], \quad (8)$$

with $\rho_{\nu,p}$ as in (4), leading to the estimation equations

$$\begin{cases} 0 = \sum_{i=1}^n \psi_{\nu,p}(\tilde{r}_i)(\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_G), \\ \left[\tilde{S}_G \right]_{j,k} = \left[n^{-1} \sum_{i=1}^n \psi_{\nu,p}(\tilde{r}_i)(\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_G)(\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_G)^\top \right]_{j,k}, & \text{for } \{j, k\} \in E \text{ or } j = k, \\ \left[\tilde{S}_G^{-1} \right]_{j,k} = 0, & \text{for } \{j, k\} \notin E \text{ and } j \neq k, \end{cases} \quad (9)$$

where $\tilde{r}_i = (\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_G)^\top \tilde{S}_G^{-1} (\mathbf{x}_i - \tilde{\boldsymbol{\mu}}_G)$ and, as before, $\psi_{\nu,p}(x) = \rho'_{\nu,p}(x)$.

The estimator \tilde{S}_G shall be called the *direct estimator*, which is short for *direct graph-constrained t_ν M-estimator*, and is an alternative to the plug-in graph-constrained t_ν M-estimator. The direct estimator is invoked by setting the option `direct = TRUE` or `plug_in = FALSE`. In case of conflicting specifications, `plug_in` has priority, and a message will be displayed.

Contrary to the plug-in estimator \hat{S}_G , the direct estimator \tilde{S}_G is not a function of the corresponding unconstrained estimate \hat{S}_n alone.

One main theoretical result of Vogel and Tyler (2014) is the asymptotic equivalence of \hat{S}_G and \tilde{S}_G under elliptical population distributions. Hence the limiting distribution of the pseudo-deviance and the constant σ_1 are the same in both cases.

It may be argued that this asymptotic equivalence result favors the plug-in estimator: Considering the elliptical distribution (6) with the fixed generator $g_{\nu,p}$, one defines the elliptical- t_ν graphical model $E_p(g_{\nu,p}, G)$ analogously to the Gaussian graphical model $N_p(G)$. The direct graphical t_ν M-estimator $(\tilde{\boldsymbol{\mu}}_G, \tilde{S}_G)$ with the corresponding loss function $\rho_{\nu,p}$ is then the maximum-likelihood estimator within this parametric family.³ So direct graphical M-estimators generalize maximum-likelihood estimators, which are known to be first-order efficient. Plug-in estimators are popular with practitioners as they are easily applied, fast to compute, and now it turns out that they also possess desirable asymptotic properties. Direct graphical M-estimators are generally harder to compute. They may be solved by a double-loop, iteratively reweighted least-squares algorithm, where the Gaussian model fit is nested into the M-estimation loop.

The other half of the story is that direct graphical M-estimators can be substantially more efficient in small samples as is demonstrated by simulations in Vogel and Tyler (2014).

³The parameters of interest are $\boldsymbol{\mu}$ and S . The degrees of freedom ν is held fixed.

Ellipticity vs. Normality

Multivariate data containing outliers may be modeled conceptually in two different ways: Either by a corrupted Gaussian distribution, i. e., a few observations are erroneous and stem from a different, outlier-generating distribution, or by a heavy-tailed elliptical distribution, which generates outliers itself. Using the t_ν M-estimator implicitly suggests the latter viewpoint. However, two issues arise:

- (1) Many data sets exhibit features such as the anxiety data set: It is clearly not normal as it contains outliers, but it is clearly not elliptical either as it is skewed. Hence we do adopt the viewpoint of corrupted normal data. In the analysis, however, we apply outlier-resistant methods which have been derived from considerations in the elliptical t_ν model. We work on the plausible, but not formalized assumption that they provide outlier-resistance regardless of the outliers being scattered symmetrically around the center or not.
- (2) An absent edge in a Gaussian graphical model, i. e., a zero partial correlation, has the interpretation of conditional independence, a notion we have repeatedly used in Section 2. Under ellipticity, an absent edge has the slightly weaker interpretation of conditional uncorrelatedness. This is occasionally mentioned as a limitation of elliptical graphical models. However, this limitation is largely void. The conclusions of any statistical analysis are always a combination of the information contained in the data and the modeling assumptions. When performing a purely linear analysis based on partial correlations, concluding conditional independences due the normality assumption certainly falls into the latter category.

Technical Appendix

The normalizing constant in the elliptical t_ν -density is

$$c_{\nu,p} = \frac{\Gamma\{(\nu + p)/2\}}{(\nu\pi)^{p/2}\Gamma(\nu/2)},$$

where Γ is the gamma function, see e. g. Bilodeau and Brenner (1999, Chapter 13).

The Kronecker product $A \otimes B$ of two matrices $A, B \in \mathbb{R}^{p \times p}$ is defined as the $p^2 \times p^2$ matrix with entry $a_{i,j}b_{k,l}$ at position $((i-1)p+k, (j-1)p+l)$. For a matrix $A = (\mathbf{a}_1, \dots, \mathbf{a}_p) \in \mathbb{R}^{p \times p}$, the notation $\text{vec}(A)$ means the p^2 -vector obtained by stacking the columns of A , i. e., $\text{vec}(A) = (\mathbf{a}_1^\top, \dots, \mathbf{a}_p^\top)^\top$. The matrix \mathcal{M}_p in (7) is defined as

$$\mathcal{M}_p = \frac{1}{2} (\mathcal{I}_{p^2} + \mathcal{K}_p),$$

where \mathcal{I}_{p^2} denotes the $p^2 \times p^2$ identity matrix and

$$\mathcal{K}_p = \sum_{i=1}^p \sum_{j=1}^p \mathbf{e}_i \mathbf{e}_j^\top \otimes \mathbf{e}_j \mathbf{e}_i^\top,$$

where $\mathbf{e}_1, \dots, \mathbf{e}_p$ denote the Euclidean basis vectors in \mathbb{R}^p . The matrix \mathcal{K}_p , commonly referred to as the *commutation matrix*, is orthogonal and corresponds to the transpose

operator $\mathcal{K}_p \text{vec}(A) = \text{vec}(A^\top)$. The idempotent matrix \mathcal{M}_p is called the *symmetrization matrix* since it maps $\text{vec}(A)$ to $\text{vec}(A + A^\top)/2$.

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. Arlington, VA: American Psychiatric Publishing, fifth edition.
- Beesdo-Baum, K., Klotsche, J., Knappe, S., Craske, M. G., LeBeau, R. T., Hoyer, J., Strobel, A., Pieper, L., and Wittchen, H.-U. (2012). Psychometric properties of the dimensional anxiety scales for DSM-V in an unselected sample of german treatment seeking patients. *Depression and Anxiety*, 29(12):1014–1024.
- Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer Texts in Statistics. New York: Springer.
- Cox, W. J. and Kenardy, J. (1993). Performance anxiety, social phobia, and setting effects in instrumental music students. *Journal of Anxiety Disorders*, 7(1):49–60.
- Dempster, A. P. (1972). Covariance Selection. *Biometrics*, 28:157–175.
- Dobos, B., Piko, B. F., and Kenny, D. T. (2019). Music performance anxiety and its relationship with social phobia and dimensions of perfectionism. *Research Studies in Music Education*, 41(3):310–326.
- Drton, M. and Perlman, M. D. (2008). A SINful approach to Gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200.
- Dümbgen, L., Nordhausen, K., and Schuhmacher, H. (2016). New algorithms for M-estimation of multivariate scatter and location. *Journal of Multivariate Analysis*, 144:200–217.
- Dümbgen, L., Nordhausen, K., and Schuhmacher, H. (2018). *fastM: Fast Computation of Multivariate M-Estimators*. R package version 0.0-4.
- Edwards, D. and Havránek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, 72:339–351.
- Fernholz, I., Mumm, J. L., Plag, J., Noeres, K., Rotter, G., Willich, S. N., Ströhle, A., Berghöfer, A., and Schmidt, A. (2019). Performance anxiety in professional musicians: a systematic review on prevalence, risk factors and clinical treatment effects. *Psychological Medicine*, 49(14):2287–2306.
- Finegold, M. and Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t -distributions. *Annals of Applied Statistics*, 5(2A):1057–1080.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

- Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, 42:1887–1896.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383–393.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, second edition.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101.
- Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, second edition.
- Kenny, D. T. (2009). The factor structure of the revised Kenny music performance anxiety inventory. In *International Symposium on performance science*, pages 37–41. Association Européenne des Conservatoires Utrecht, The Netherlands.
- Kenny, D. T. (2011). *The Psychology of Music Performance Anxiety*. Oxford University Press.
- Lauritzen, S. L. (1996). *Graphical models*. Oxford Statistical Science Series. Oxford University Press.
- Lebeau, R. T., Glenn, D. E., Hanover, L. N., Beesdo-Baum, K., Wittchen, H.-U., and Craske, M. G. (2012). A dimensional approach to measuring anxiety for DSM-5. *International Journal of Methods in Psychiatric Research*, 21(4):258–272.
- Marchetti, G. M., Drton, M., and Sadeghi, K. (2020). *ggm: Graphical Markov Models with Mixed Graphs*. R package version 2.5.
- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 4:51–67.
- Maronna, R. A., Martin, D. R., and Yohai, V. J. (2006). *Robust statistics: Theory and methods*. Wiley Series in Probability and Statistics. Chichester: Wiley.
- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34(3):1436–1462.
- Nicholson, D. R., Cody, M. W., and Beck, J. G. (2015). Anxiety in musicians: On and off stage. *Psychology of Music*, 43(3):438–449.
- Nordhausen, K., Oja, H., and Tyler, D. E. (2008). Tools for exploring multivariate data: The package ICS. *Journal of Statistical Software*, 28(6):1–31.
- Öllerer, V. and Croux, C. (2015). Robust high-dimensional precision matrix estimation. In Nordhausen, K. and Taskinen, S., editors, *Modern Nonparametric, Robust and Multivariate Methods*, pages 325–350. Berlin, Heidelberg: Springer-Verlag.

- Spahn, C., Walther, J.-C., and Nusseck, M. (2016). The effectiveness of a multimodal concept of audition training for music students in coping with music performance anxiety. *Psychology of Music*, 44(4):893–909.
- Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Annals of Statistics*, 14:138–150.
- Tarr, G., Müller, S., and Weber, N. C. (2016). Robust estimation of precision matrices under cellwise contamination. *Computational Statistics & Data Analysis*, 93:404–420.
- Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika*, 69:429–436.
- Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, 70:411–420.
- Tyler, D. E. (1987). A Distribution-Free M -Estimator of Multivariate Scatter. *Annals of Statistics*, 15(1):234 – 251.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.
- Vogel, D. and Tyler, D. E. (2014). Robust estimators for non-decomposable elliptical graphical models. *Biometrika*, 101(4):865–882.
- Watt, S. J. (2019). Algorithms for data analysis. B.Sc. thesis, University of Aberdeen.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester, etc.: Wiley.
- Wiedemann, A., Vogel, D., Voss, C., and Hoyer, J. (2022). How does music performance anxiety relate to other anxiety disorders? *Psychology of Music*, 50:204–2017.
- Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35.