Hüsler–Reiss Graphical Models for Multivariate Extremes

Manuel Hentschel

University of Geneva

May 16, 2024

Workshop on graphical models and clustering, Montpellier 2024

Based on

Statistical Inference for Hüsler-Reiss Graphical Models Through Matrix Completions
(Hentschel, M., Engelke, S., and Segers, J., arxiv 2210.14292)

Graphical models for multivariate extremes
(Engelke, S., Hentschel, M., Lalancette, M., and Röttger, F., arxiv 2402.02187)

R package: `graphicalExtremes`
(Engelke, S., Hitz, A., Gnecco, N., and Hentschel, M.)
`https://github.com/sebastian-engelke/graphicalExtremes`

Outline

Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
- Inference on known graph structures
- Structure estimation

Application

Definition

Definition (Graph)

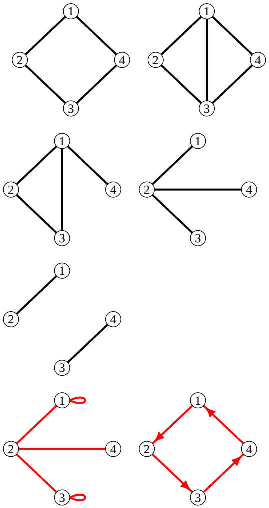
An undirected graph $G = (V, E)$ consists of a finite set of vertices V and a set of edges

$$E \subseteq \{\{i, j\} \mid i, j \in V, i \neq j\}.$$

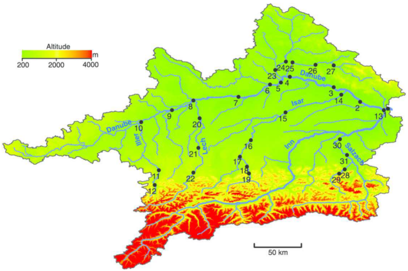
Definition (Graphical model)

A random vector X indexed by V is a graphical model with respect to graph $G = (V, E)$, if it satisfies the Markov property

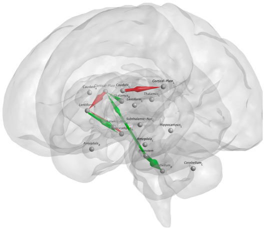
$$\{i, j\} \notin E \Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}.$$

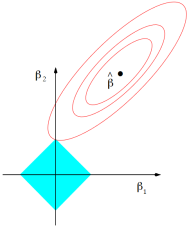


Graphical models – Motivation



- Use of structural (expert) knowledge
- Interpretability
- Sparsity
    - Faster computation
    - Implicit regularization





Gaussian graphical models

- Parametrized in terms of precision matrix $\Theta = \Sigma^{-1}$
- Density

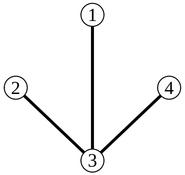
$$p(x) = (2\pi)^{-d/2} |\Theta|^{1/2} \exp(-\tfrac{1}{2}(x-\mu)^\top \Theta (x-\mu))$$

- Conditional independence is encoded as

$$\Theta_{ij} = 0 \iff X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$$

- Popular inference method: "Graphical Lasso"

$$\Theta_{\text{glasso}} = \underset{\Theta \in \mathcal{S}^+}{\text{argmin}} -\ell(\Theta, \hat{\Sigma}) + \lambda \|\Theta\|_1$$



$$\Theta = \begin{pmatrix} 5 & 0 & -4 & 0 \\ 0 & 3 & -2 & 0 \\ -4 & -2 & 8 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

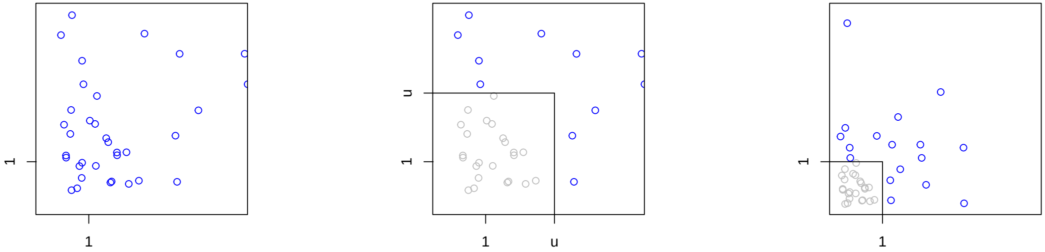
Multivariate generalized Pareto distribution

- Random vector X with standard Pareto univariate margins
- Z is multivariate Pareto if

$$\mathbb{P}(Z \leq z) = \lim_{u \to \infty} \mathbb{P}\left(u^{-1}X < z \,\middle|\, \max_{j=1,\ldots,d} X_j > u\right)$$

- supported on

$$\mathcal{L} = \left\{x \in [0,\infty)^d \,\middle|\, \max_{j=1,\ldots,d} x_j \geq 1\right\}$$



Exponent measure (density)

- Distribution of Z characterized by exponent measure Λ with

$$\mathbb{P}(Z \leq z) = \frac{\Lambda([0, z] \setminus [\mathbf{0}, \mathbf{1}])}{\Lambda(\mathcal{L})}.$$

- We assume that Λ is absolutely continuous:

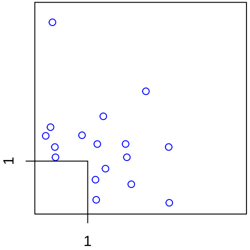
$$\Lambda(A) = \int_A \lambda(x) dx.$$

- Then Z has density

$$f_Z(z) = \frac{\lambda(z)}{\Lambda(\mathcal{L})}.$$

- Important homogeneity properties:

$$\lambda(\alpha x) = \alpha^{-(d+1)} \lambda(x), \qquad \Lambda(\alpha A) = \alpha^{-1} \Lambda(A).$$



Extremal graphical models

- Conditional independence problematic on $\mathcal{L}$.
- For $z_2 < 1$ we must (!) have

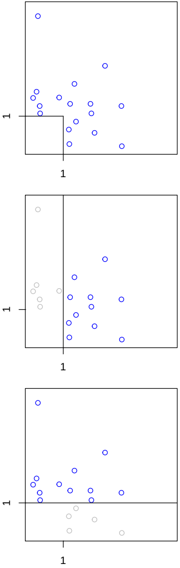
$$Z_1 \not\perp\!\!\!\perp Z_3 \mid Z_2 = z_2$$

- Approach by Engelke and Hitz (2020):
  - Assume positive, continuous exponent measure density λ
  - Let $Z^{(k)} = Z \mid \{Z_k > 1\}$
  - Define

$$Z_A \perp_e Z_B \mid Z_C \iff Z_A^{(k)} \perp\!\!\!\perp Z_B^{(k)} \mid Z_C^{(k)} \quad \forall k \in V$$

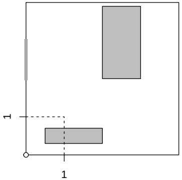
  - Implies factorization of density λ (for $A \cup B \cup C = V$):

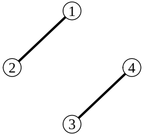
$$\lambda(z) = \frac{\lambda_{A\cup C}(z_{A\cup C})\lambda_{B\cup C}(z_{B\cup C})}{\lambda_C(z_C)}$$



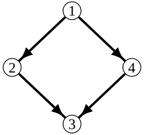
More general definitions

- The previous definition requires connected models with Lebesgue densities
- More general definition in Engelke et al. (2022a)
  - Replaces $\mathcal{L}^{(k)}$ by rectangular test sets
  - Works with general exponent measures
  - Allows disconnected models
  - Compatible with "$\cdot \perp_e \cdot \mid \cdot$"
  - Can be applied to Lévy processes





- (Active research about directed graphical models)



Choice of univariate margins (!!)

- So far, we have considered standard Pareto margins
- For connected Hüsler–Reiss models, standard exponential margins are more convenient
- We can always transform between the two!

$$Y_i \sim \mathrm{Exp}(1) \qquad Z_i \sim \mathrm{Pareto}(1)$$

$$Y_i = \log(Z_i) \qquad Z_i = \exp(Y_i)$$

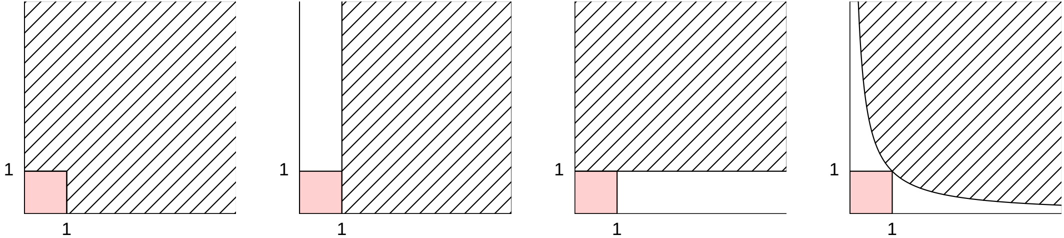
$$f_Y(y) = f_Z(\exp(y)) \prod_{i=1}^{d} \exp(y_i) \qquad f_Z(z) = f_Y(\log(z)) \prod_{i=1}^{d} z_i^{-1}$$

$$\mathcal{L}_Y = \{y \in [-\infty, \infty)^d \mid \max_i y_i \geq 0\} \qquad \mathcal{L}_Z = \{z \in [0, \infty)^d \mid \max_i z_i \geq 1\}$$

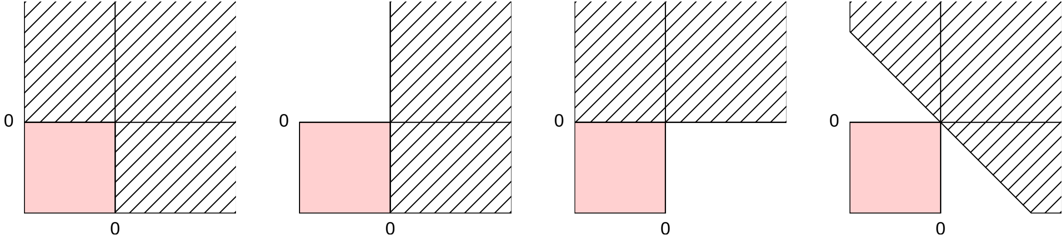
$$f_Y(y + \mathbf{1}\beta) = f_Y(y) \exp(-\beta) \qquad f_Z(\alpha z) = \alpha^{-(d+1)} f_Z(z)$$

Choice of univariate margins (!!)

Support of Z, $Z^{(1)}$, $Z^{(2)}$, ...:



Support of Y, $Y^{(1)}$, $Y^{(2)}$, ...:



Variogram matrices

Definition

A square matrix $\Gamma \in \mathbb{R}^{d \times d}$ is conditionally negative definite if it satisfies

$$\begin{aligned} \Gamma &= \Gamma^\top, \\ \operatorname{diag}(\Gamma) &= \mathbf{0}, \\ v^\top \Gamma v &< 0 \quad \forall \mathbf{0} \neq v \perp \mathbf{1}. \end{aligned}$$

$$\Gamma = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$

Also arise as

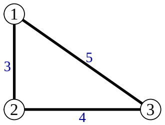
- variogram matrices (for $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$):

$$\Gamma_{ij} = \mathrm{Var}(X_i - X_j) = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}$$

$$\Sigma = \begin{pmatrix} 9.78 & 3.78 & -1.56 \\ 3.78 & 6.78 & 1.44 \\ -1.56 & 1.44 & 12.11 \end{pmatrix}$$

- Euclidean distance matrices:

$$\Gamma_{ij} = |P_i - P_j|^2$$



The Hüsler–Reiss distribution

Definition (Hüsler–Reiss Pareto distribution)

Y is Hüsler–Reiss Pareto distributed with parameter matrix Γ if its exponent measure density satisfies

$$\lambda(y;\Gamma) = \exp(-y_k) \cdot \varphi_{d-1}(\tilde{y}_{\setminus\{k\}};\Sigma^{(k)}), \qquad y \in \mathbb{R}^d, k \in \{1,\dots,d\},$$

where φ_{d-1} denotes the $d-1$-dimensional centered normal density with covariance $\Sigma^{(k)}$ and

$$\Sigma^{(k)}_{ij} = \tfrac{1}{2}(\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}), \qquad i,j \neq k,$$
$$\tilde{y}_i = y_i - y_k + \tfrac{1}{2}\Gamma_{ik}, \qquad i \neq k.$$

Introduced by Hüsler and Reiss (1989) as limit of $\max_{i=1\dots n} X_i$ where $X_i \sim \mathcal{N}(0,\Sigma_n)$ and

$$(1-(\Sigma_n)_{ij})\log n \to \tfrac{1}{4}\Gamma_{ij}$$

A stochastic construction

- Recall the density

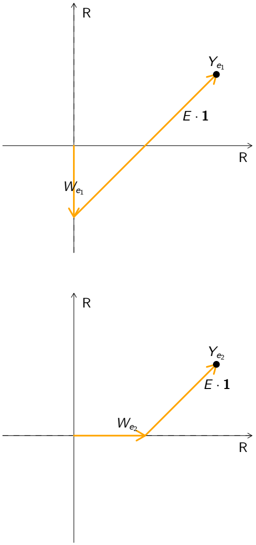
$$\lambda(y;\Gamma) = \exp(-y_k) \cdot \varphi_{d-1}(\tilde{y}_{\setminus\{k\}}; \Sigma^{(k)})$$

- Let $Y \sim \lambda(y;\Gamma)$ and $Y^{(k)} = Y \mid \{Y_k > 0\}$.
- Then:

$$Y^{(k)} \overset{d}{=} W^{(k)} + E\mathbf{1},$$

with $W_k^{(k)} = 0$, $\mu_i^{(k)} = -\frac{1}{2}\Gamma_{ik}$, and

$$E \sim \mathrm{Exp}(1),$$
$$W_{\setminus\{k\}}^{(k)} \sim \mathcal{N}(\mu^{(k)}, \Sigma^{(k)}).$$

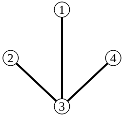


Hüsler–Reiss graphical models

- (Gaussian) precision matrices

$$\Theta^{(k)} = \left(\Sigma^{(k)}\right)^{-1} \in \mathbb{R}^{(d-1)\times(d-1)}$$

$$\Theta^{(k)}_{ij} = 0 \iff W^{(k)}_i \perp\!\!\!\perp W^{(k)}_j \mid W^{(k)}_{\setminus\{i,j\}}$$



$$\Gamma = \begin{pmatrix} 0 & 0.75 & 0.25 & 1.25 \\ 0.75 & 0 & 0.5 & 1.5 \\ 0.25 & 0.5 & 0 & 1 \\ 1.25 & 1.5 & 1 & 0 \end{pmatrix}$$

- Proposition 1 in Engelke and Hitz (2020):

$$\Theta^{(k)}_{ij} = 0 \iff Y_i \perp_e Y_j \mid Y_{\setminus\{i,j\}}$$

- Lemma 1 in Engelke and Hitz (2020):

$$\Theta^{(k)}_{ij} = \Theta^{(k')}_{ij} \qquad \forall i,j \neq k,k'$$

$$\Sigma^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.75 & 0.25 & 0.25 \\ 0 & 0.25 & 0.25 & 0.25 \\ 0 & 0.25 & 0.25 & 1.25 \end{pmatrix}, \quad \Sigma^{(2)} = \begin{pmatrix} 0.75 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 & 1.5 \end{pmatrix}, \ldots$$

$$\Theta^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & -2 & 7 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad \Theta^{(2)} = \begin{pmatrix} 4 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & 7 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \ldots$$

- Definition 3.1 in Hentschel et al. (2022):
  Hüsler–Reiss precision matrix $\Theta \in \mathbb{R}^{d\times d}$ with

$$\Theta_{ij} = \Theta^{(k)}_{ij} \qquad \text{for some } k \neq i,j$$

$$\Theta = \begin{pmatrix} 4 & 0 & -4 & 0 \\ 0 & 2 & -2 & 0 \\ -4 & -2 & 7 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Hüsler–Reiss precision matrix

Proposition

With $P = I - d^{-1}\mathbf{1}\mathbf{1}^\top$, the Hüsler–Reiss precision matrix Θ can be expressed as

$$\Theta = \Sigma^+$$

$$\Sigma = P(-\tfrac{1}{2}\Gamma)P$$

$$\Gamma = \begin{pmatrix} 0 & 0.75 & 0.25 & 1.25 \\ 0.75 & 0 & 0.5 & 1.5 \\ 0.25 & 0.5 & 0 & 1 \\ 1.25 & 1.5 & 1 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.23 & -0.08 & 0.05 & -0.20 \\ -0.08 & 0.36 & -0.02 & -0.27 \\ 0.05 & -0.02 & 0.11 & -0.14 \\ -0.20 & -0.27 & -0.14 & 0.61 \end{pmatrix}$$

$$\Theta = \begin{pmatrix} 4 & 0 & -4 & 0 \\ 0 & 2 & -2 & 0 \\ -4 & -2 & 7 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Intuition:

- Σ is positive semi-definite:

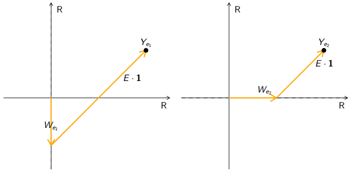
$$v^\top \Sigma v = -\tfrac{1}{2}(Pv)^\top \Gamma (Pv) \geq 0$$

- with kernel $\mathbb{R}\mathbf{1}$:

$$\Sigma\mathbf{1} = (\ldots P)\mathbf{1} = \mathbf{0}$$

- and so is its Moore–Penrose inverse Θ.

Density and stochastic construction



Proposition

The Hüsler–Reiss exponent measure density $\lambda(\cdot;\Gamma)$ can be expressed as

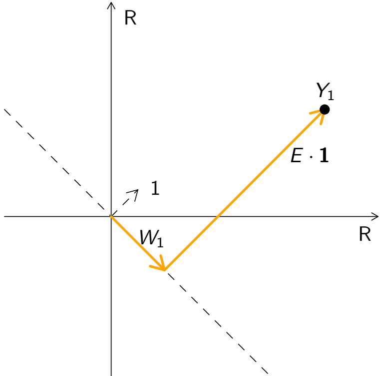
$$\lambda(y;\Gamma) = c \cdot \exp(-y^\top \mathbf{e}_d) \cdot \exp(-\tfrac{1}{2} y^\top \Theta y + y^\top r_\Theta),$$

with $\mathbf{e}_d = d^{-1}\mathbf{1}$ and $r_\Theta \perp \mathbf{1}$.

Corresponding stochastic construction:

$$Y^{(\mathbf{1})} \overset{d}{=} W + E\mathbf{1}$$
$$W \sim \mathcal{N}(\mu_\Theta, \Theta^+)$$

for $Y^{(\mathbf{1})} = Y \mid \mathbf{1}^\top Y > 0$.



Limit representation

Let $S \in \mathbb{R}^{d \times d}$ such that

$$\Gamma_{ij} = S_{ii} + S_{jj} - 2S_{ij}.$$

Then Θ can also be expressed as

$$\Theta = \lim_{t \to \infty} \left(S + t\mathbf{1}\mathbf{1}^\top\right)^{-1}.$$

Furthermore (Wan and Zhou (2023)),

$$\left(\Sigma + t\mathbf{1}\mathbf{1}^\top\right)^{-1} = \Theta + t^{-1}d^{-2}\mathbf{1}\mathbf{1}^\top.$$

Summary: Hüsler–Reiss precision matrix

	Gaussian	Hüsler–Reiss
Parameter Matrix	Σ	Γ
Precision Matrix	Σ^{-1}	$(P(-\frac{1}{2}\Gamma)P)^{+}$
Density	$\propto \exp\left(-\frac{1}{2}\|y-\mu\|_{\Theta}^{2}\right)$	$\propto \exp\left(-\frac{1}{2}\|y-\mu_{\Theta}\|_{\Theta}^{2}\right)\cdot\ldots$
Graphical structure	$\Theta_{ij}=0$	$\Theta_{ij}=0$

Inference for Gaussian graphical models (on a known graph)

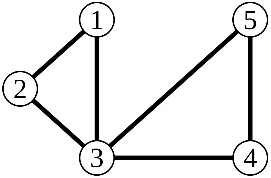
- Interpretation of entries in matrices Σ and Θ:

  Σ_{ij}: Covariance between Y_i and Y_j $\Rightarrow$ estimate using $\hat{\Sigma}_{ij}$

  Θ_{ij}: Conditional independence indicated by zeros $\Rightarrow$ known from graph structure

- Resulting matrix completion problems:

$$\Sigma = \begin{pmatrix} 5 & 4 & 4 & ? & ? \\ 4 & 8 & 4 & ? & ? \\ 4 & 4 & 8 & 6 & 6 \\ ? & ? & 6 & 10 & 7 \\ ? & ? & 6 & 7 & 13 \end{pmatrix} \quad \Theta = \begin{pmatrix} ? & ? & ? & 0 & 0 \\ ? & ? & ? & 0 & 0 \\ ? & ? & ? & ? & ? \\ 0 & 0 & ? & ? & ? \\ 0 & 0 & ? & ? & ? \end{pmatrix}$$

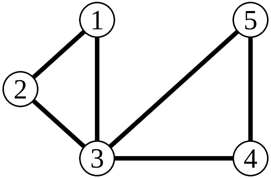


- Solution in Speed and Kiiveri (1986):
  - Explicit construction for decomposable graphs
  - Convergent algorithm for general graphs

Inference for Hüsler–Reiss graphical models (on a known graph)

- Interpretation of entries in matrices Γ and Θ:
  - Γ_{ij}: Marginal distribution of $Y_{\{ij\}}$ $\Rightarrow$ estimate how?
  - Θ_{ij}: Conditional independence indicated by zeros $\Rightarrow$ known from graph structure
- Resulting matrix completion problem:

$$\Gamma = \begin{pmatrix} 0 & 1 & 6 & ? & ? \\ 1 & 0 & 3 & ? & ? \\ 6 & 3 & 0 & 3 & 4 \\ ? & ? & 3 & 0 & 3 \\ ? & ? & 4 & 3 & 0 \end{pmatrix} \quad \Theta = \begin{pmatrix} ? & ? & ? & 0 & 0 \\ ? & ? & ? & 0 & 0 \\ ? & ? & ? & ? & ? \\ 0 & 0 & ? & ? & ? \\ 0 & 0 & ? & ? & ? \end{pmatrix}$$



- Solutions?

Empirical variogram

- Engelke and Volgushev (2020) introduce the empirical variogram $\hat{\Gamma}^{(m)}$

$$\hat{\Gamma}_{ij}^{(m)} = \widehat{\mathrm{Var}}\Big(\log\Big(1 - \tilde{F}_i(X_{ti})\Big) - \log\Big(1 - \tilde{F}_j(X_{tj})\Big) : \tilde{F}_m(X_{tm}) \geq 1 - k/n\Big),$$

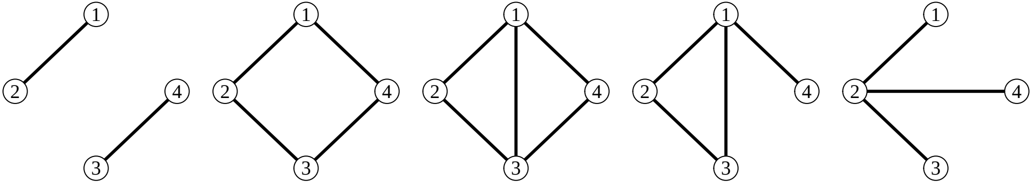
  where $\tilde{F}$ is the empirical distribution function.
- can be made independent of $m \in V$ by considering

$$\hat{\Gamma}_{ij} = \frac{1}{n} \sum_{m \in V} \hat{\Gamma}_{ij}^{(m)},$$

- which is a consistent estimator of Γ under some conditions.

Graphs

Connected graph: a graph in which there is a path between every pair of vertices.
Decomposable graph: a graph in which all cycles of length ≥ 4 have a chord.
Block graph: a decomposable graph in which all separators are of size 1.
Tree (graph): a connected graph without cycles.

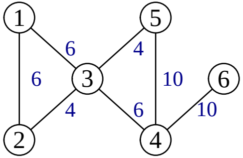


Matrix completion on block graphs and trees

- Let $G = (V, E)$ be a connected block graph (or a tree)
- Let $\mathring{\Gamma}$ be a partial variogram, specified on the edges of G
- Then we have the additivity property

$$\Gamma_{st} = \sum_{(i,j) \in \text{ph}(s,t)} \Gamma_{ij},$$

where $\text{ph}(s, t)$ denotes the shortest path between s and t and the resulting Γ has graphical structure G.
(Engelke and Hitz (2020); Asenova et al. (2021); Engelke and Volgushev (2020); Asenova and Segers (2021))

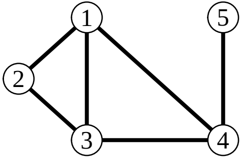


$$\mathring{\Gamma} = \begin{pmatrix} 0 & 6 & 6 & ? & ? & ? \\ 6 & 0 & 4 & ? & ? & ? \\ 6 & 4 & 0 & 6 & 4 & ? \\ ? & ? & 6 & 0 & 10 & 10 \\ ? & ? & 4 & 10 & 0 & ? \\ ? & ? & ? & 10 & ? & 0 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0 & 6 & 6 & \underline{12} & \underline{10} & \underline{22} \\ 6 & 0 & 4 & \underline{10} & \underline{8} & \underline{20} \\ 6 & 4 & 0 & 6 & 4 & \underline{16} \\ \underline{12} & \underline{10} & 6 & 0 & 10 & 10 \\ \underline{10} & \underline{8} & 4 & 10 & 0 & \underline{20} \\ \underline{22} & \underline{20} & \underline{16} & 10 & \underline{20} & 0 \end{pmatrix}$$

Matrix completion on decomposable graphs

- Let $G = (V, E)$ be a connected, decomposable graph
- Let $\mathring{\Gamma}$ be a partial variogram, specified on the edges of G



Proposition

There exists a unique completion Γ *such that*

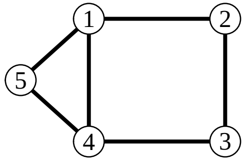
$$\Gamma_{ij} = \mathring{\Gamma}_{ij}, \forall (i,j) \in \overline{E},$$
$$\Theta_{ij} = 0, \ \forall (i,j) \notin \overline{E}.$$

$$\mathring{\Gamma} = \begin{pmatrix} 0 & 10 & 4 & 3 & ? \\ 10 & 0 & 18 & ? & ? \\ 4 & 18 & 0 & 3 & ? \\ 3 & ? & 3 & 0 & 6 \\ ? & ? & ? & 6 & 0 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0 & 10 & 4 & 3 & \underline{9} \\ 10 & 0 & 18 & \underline{15} & \underline{21} \\ 4 & 18 & 0 & 3 & \underline{9} \\ 3 & \underline{15} & 3 & 0 & 6 \\ \underline{9} & \underline{21} & \underline{9} & 6 & 0 \end{pmatrix}$$

- This completion can be computed explicitly.
- The mapping $\mathring{\Gamma} \mapsto \Gamma$ is continuous.

Matrix completion on general connected graphs



- Let $G = (V, E)$ be a connected graph
- Let $\mathring{\Gamma}$ be a partial variogram, specified on the edges of G

$$\tilde{\Gamma} = \begin{pmatrix} 0 & 0.23 & \underline{0.08} & 0.09 & 0.21 \\ 0.23 & 0 & 0.14 & \underline{0.23} & \underline{0.19} \\ \underline{0.08} & 0.14 & 0 & 0.11 & \underline{0.20} \\ 0.09 & \underline{0.23} & 0.11 & 0 & 0.16 \\ 0.21 & \underline{0.19} & \underline{0.20} & 0.16 & 0 \end{pmatrix}$$

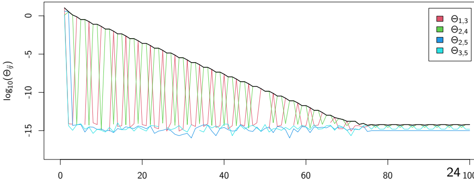
Proposition

If there exists any valid completion of $\mathring{\Gamma}$ (say $\tilde{\Gamma}$), then there exists a unique Γ such that

$$\Gamma_{ij} = \mathring{\Gamma}_{ij}, \forall (i,j) \in \overline{E},$$
$$\Theta_{ij} = 0, \ \forall (i,j) \notin \overline{E}.$$

$$\Gamma = \begin{pmatrix} 0 & 0.23 & \underline{0.18} & 0.09 & 0.21 \\ 0.23 & 0 & 0.14 & \underline{0.21} & \underline{0.35} \\ \underline{0.18} & 0.14 & 0 & 0.11 & \underline{0.26} \\ 0.09 & \underline{0.21} & 0.11 & 0 & 0.16 \\ 0.21 & \underline{0.35} & \underline{0.26} & 0.16 & 0 \end{pmatrix}$$

- This completion can be computed as the limit of a convergent sequence of matrices, starting with $\tilde{\Gamma}$.
- The mapping $\mathring{\Gamma} \mapsto \Gamma$ is continuous.



Statistical inference through matrix completions

- Hüsler–Reiss graphical model on known graph $G = (V, E)$
- Parametrized by unknown variogram Γ
- $\widehat{\mathring{\Gamma}}_n$ sequence of consistent estimators for Γ_{ij}, $(ij) \in E$, satisfying $\operatorname{diag} \widehat{\mathring{\Gamma}}_n \equiv \mathbf{0}$

Proposition

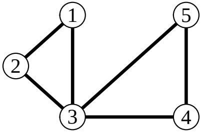
For $\widehat{\mathring{\Gamma}}$ as above,

- *with probability going to 1, there exists a graphical completion $\widehat{\Gamma}_n^G \in \mathcal{D}$ of $\widehat{\mathring{\Gamma}}_n$.*
- *This completion is consistent for all $(i, j) \in V \times V$, i.e.,*

$$\mathbb{P}\left(\max_{(i,j) \in V \times V} \left| \widehat{\Gamma}_{ij}^G - \Gamma_{ij} \right| < \varepsilon \right) \to 1, \quad n \to \infty.$$

Estimation strategies on sparse graphs

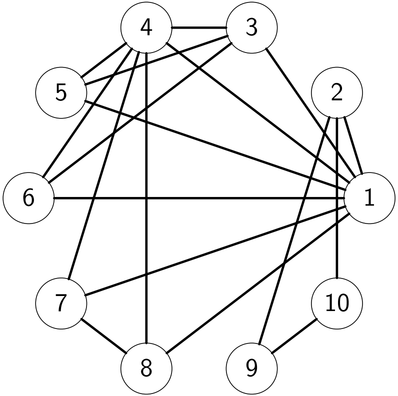
- Setting:
  - Known, connected graph $G = (V, E)$ with clique set $\mathcal{C}$
  - Unknown variogram matrix Γ
  - Samples from a Hüsler–Reiss graphical model parametrized by G, Γ
  - Consistent estimator $\hat{\Gamma}_{S \times S}$ for $S \subseteq V$ available
- To be estimated:
  - Unknown variogram matrix Γ



Estimation strategies:

$$\text{“Full”:} \begin{pmatrix} 0 & \color{red}{1} & \color{red}{6} & \color{red}{9} & \color{red}{10} \\ 1 & 0 & \color{red}{3} & \color{red}{6} & \color{red}{7} \\ 6 & 3 & 0 & \color{red}{3} & \color{red}{4} \\ 9 & 6 & 3 & 0 & \color{red}{3} \\ 10 & 7 & 4 & 3 & 0 \end{pmatrix} \quad \text{“Graphical”:} \begin{pmatrix} 0 & \color{red}{1} & \color{red}{6} & ? & ? \\ 1 & 0 & \color{red}{3} & ? & ? \\ 6 & 3 & 0 & \color{red}{3} & \color{red}{4} \\ ? & ? & 3 & 0 & \color{red}{3} \\ ? & ? & 4 & 3 & 0 \end{pmatrix} \quad \text{“Clique-wise”:} \begin{pmatrix} 0 & \color{red}{1} & \color{red}{6} & ? & ? \\ 1 & 0 & \color{red}{3} & ? & ? \\ 6 & 3 & 0 & \color{green}{3} & \color{green}{4} \\ ? & ? & 3 & 0 & \color{green}{3} \\ ? & ? & 4 & 3 & 0 \end{pmatrix}$$

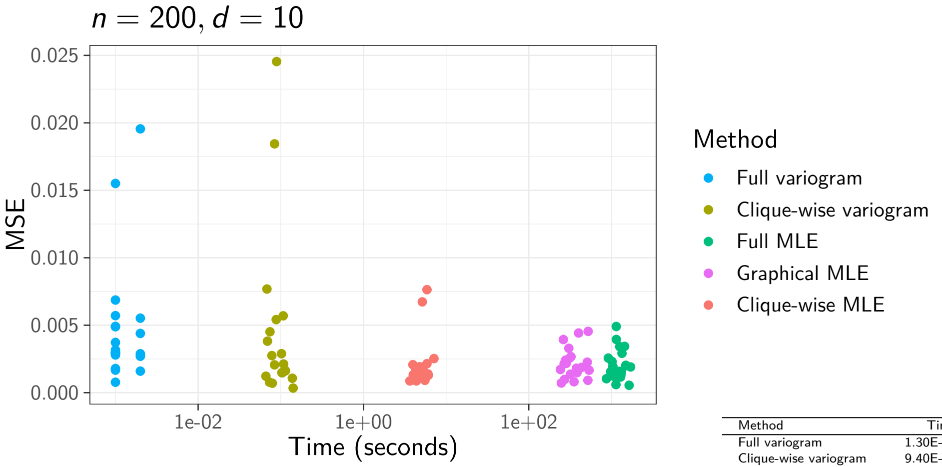
Simulation study – Setting



- Graph $G = (V, E)$ with $|V| = 10$, $|E| = 18$.
- Sample size $n = 200$
- Parameter matrix Θ:

$$\begin{pmatrix}
24.15 & -3.95 & 2.45 & -10.61 & -1.42 & -4.85 & -4.51 & -1.25 & 0 & 0 \\
-3.95 & 77.85 & 0 & 0 & 0 & 0 & 0 & 0 & 7.13 & -81.03 \\
2.45 & 0 & 22.58 & -15.78 & -8.16 & -1.09 & 0 & 0 & 0 & 0 \\
-10.61 & 0 & -15.78 & 27.10 & -2.18 & 1.69 & 4.32 & -4.53 & 0 & 0 \\
-1.42 & 0 & -8.16 & -2.18 & 11.76 & 0 & 0 & 0 & 0 & 0 \\
-4.85 & 0 & -1.09 & 1.69 & 0 & 4.25 & 0 & 0 & 0 & 0 \\
-4.51 & 0 & 0 & 4.32 & 0 & 0 & 3.97 & -3.77 & 0 & 0 \\
-1.25 & 0 & 0 & -4.53 & 0 & 0 & -3.77 & 9.55 & 0 & 0 \\
0 & 7.13 & 0 & 0 & 0 & 0 & 0 & 0 & 40.63 & -47.76 \\
0 & -81.03 & 0 & 0 & 0 & 0 & 0 & 0 & -47.76 & 128.79
\end{pmatrix}$$

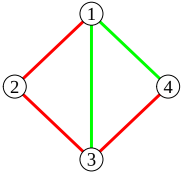
Simulation Study – Results



Method	Time	MSE
Full variogram	1.30E-03	4.87E-03
Clique-wise variogram	9.40E-02	4.51E-03
Clique-wise MLE	4.97E+00	2.07E-03
Graphical MLE	3.57E+02	2.19E-03
Full MLE	1.24E+03	2.09E-03

Colored models

- Number of edges in the graph/cliques might still be very large
- Desirable to further reduce the dimensionality of the problem
- Röttger et al. (2023) suggest colored graphical models:
  - Edge coloring: $\lambda : E \to \{1, \dots, r\}$
  - RCON: $\Theta_{ij} = \Theta_{kl}$ if $\lambda((ij)) = \lambda((kl))$
  - RVAR: $\Gamma_{ij} = \Gamma_{kl}$ if $\lambda((ij)) = \lambda((kl))$
- Inference:
  - Determine λ through a clustering step
  - Determine r through cross-validation



$$\Theta = \begin{pmatrix} 5 & -1 & -2 & -2 \\ -1 & 2 & -1 & 0 \\ -2 & -1 & 4 & -1 \\ -2 & 0 & -1 & 3 \end{pmatrix}$$

$$\neq$$

$$\Gamma = \begin{pmatrix} 0 & 2 & 4 & 4 \\ 2 & 0 & 2 & 3 \\ 4 & 2 & 0 & 2 \\ 4 & 3 & 2 & 0 \end{pmatrix}$$

Tree graphs

- Recall the additivity property (for trees/block graphs):

$$\Gamma_{st} = \sum_{(i,j) \in \mathrm{ph}(s,t)} \Gamma_{ij},$$

- Similar result for extremal correlation χ

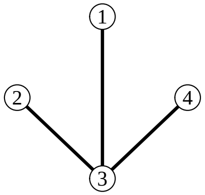
$$\chi_{ij} \leq \chi_{st} \qquad \forall\, (ij) \in \mathrm{ph}(s,t)$$

- Minimum spanning tree (EMST):

$$T_{mst} = \underset{(V,E) \in \mathcal{T}}{\mathrm{argmin}} \sum_{(ij) \in E} \rho_{ij},$$

with $\rho_{ij} = \Gamma_{ij}$ or $\rho_{ij} = -\log \chi_{ij}$

- Also works for non-Hüsler–Reiss tree models (Engelke and Volgushev (2020))



$$\Gamma = \begin{pmatrix} 0 & 0.75 & 0.25 & 1.25 \\ 0.75 & 0 & 0.5 & 1.5 \\ 0.25 & 0.5 & 0 & 1 \\ 1.25 & 1.5 & 1 & 0 \end{pmatrix}$$

$$\chi = \begin{pmatrix} 1 & 0.67 & 0.80 & 0.58 \\ 0.67 & 1 & 0.72 & 0.54 \\ 0.80 & 0.72 & 1 & 0.62 \\ 0.58 & 0.54 & 0.62 & 1 \end{pmatrix}$$

“Ideal extremal graphical lasso”

- Normal distribution:

$$f(x) \propto \sqrt{|\Theta|} \exp((x-\mu)^\top \Theta (x-\mu))$$
$$(ij) \notin E \Rightarrow \Theta_{ij} = 0$$

- Gaussian graphical lasso:

$$\Theta_{\text{glasso}} = \underset{\Theta \in \mathcal{S}^+}{\operatorname{argmin}} -\ell(\Theta, \hat{\Sigma}) + \lambda \|\Theta\|_1$$

- Efficiently estimates structure and parameters at once!

- Hüsler–Reiss distribution:

$$\lambda(y; \Theta) \propto c_\Theta \sqrt{|\Theta|_+} \exp((y-\mu_\Theta)^\top \Theta (y-\mu_\Theta))$$
$$(ij) \notin E \Rightarrow \Theta_{ij} = 0$$
$$Y^{(\mathbf{1})} \overset{d}{=} W + E\mathbf{1}$$

- “Extremal graphical lasso”?

$$\Theta_{\text{eglasso}} = \underset{\Theta \in \mathcal{S}_{\mathbf{1}}^+}{\operatorname{argmin}} -\ell(\Theta, \hat{\Gamma}) + \lambda \|\Theta\|_{1,\text{off}}$$

- Does not work! Ying et al. (2020, 2021)

Majority voting

- Engelke et al. (2022b) consider the matrices

$$\Theta^{(1)}, \ldots, \Theta^{(d)} \in \mathbb{R}^{(d-1)\times(d-1)}$$

  - $\Theta^{(k)}_{ij} = 0 \quad \Leftrightarrow \quad \Theta_{ij} = 0$
  - $\Theta^{(k)}$ is a full rank precision matrix

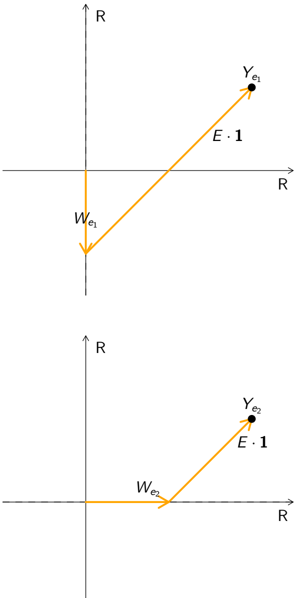
- Apply graphical lasso for each k, yielding sparse

$$\widehat{\Theta}^{(1)}, \ldots, \widehat{\Theta}^{(d)}$$

- Obtain $\widehat{G} = (V, \widehat{E})$ by a majority vote:

$$\#\{k \neq i,j \mid \widehat{\Theta}^{(k)}_{ij} \neq 0\} > \frac{d-2}{2} \implies (ij) \in \widehat{E}$$

- Some sparsistency results, but no parameter estimate!



Score matching

- Alternative to MLE that minimizes the Fisher information distance

$$J(f) = \int_{\mathbb{R}^d} f_0(x) \|\nabla_x \log f(x) - \nabla_x \log f_0(x)\|_2^2 dx$$

- Computationally efficient without computing the normalization constant
- Can be ℓ^1-penalized to obtain sparse graphs
- Lederer and Oesting (2023) apply this approach to Hüsler–Reiss models
  - Optimize over a superset of Hüsler–Reiss parametrizations
    $\Rightarrow$ might yield invalid Θ matrices
  - Provide concentration guarantees

Parameter shift

- Wan and Zhou (2023) consider for some $c \in \mathbb{R}$ the full rank matrices

$$\begin{aligned}\Theta^* &= \Theta + c\mathbf{1}\mathbf{1}^\top \\ \Sigma^* &= (\Theta^*)^{-1} \\ &= \Sigma + c^{-1}d^{-2}\mathbf{1}\mathbf{1}^\top\end{aligned}$$

- Estimate $\hat{\Sigma}$ from data and set $\widehat{\Sigma}^* = \widehat{\Sigma} + c^{-1}d^{-2}\mathbf{1}\mathbf{1}^\top$
- Apply graphical lasso with modified penalty term

$$\hat{\Theta}^* = \underset{\Theta^* \in \mathcal{S}^+}{\operatorname{argmin}} -\ell\big(\Theta^*, \widehat{\Sigma}^*\big) + \lambda \|\Theta^* - c\mathbf{1}\mathbf{1}^\top\|_{1,\text{off}}$$

- Obtain $\hat{\Theta} = \hat{\Theta}^* - c\mathbf{1}\mathbf{1}^\top$
- No guarantee that $\hat{\Theta}$ is a valid precision matrix
- But still consistent

EMTP_2

- Röttger et al. (2021) introduce extremal multivariate total positivity (EMTP_2)
- Requires $Y^{(k)}$ to be MTP_2, i.e. their densities satisfy

$$f(x \vee y) f(x \wedge y) \geq f(x) f(y)$$

- For Hüsler–Reiss models encoded as

$$\Theta_{ij} \leq 0 \quad \forall i \neq j$$

- Constraint typically yields sparse estimates $\hat{\Theta}_+$
- Some asymptotic structural consistency results

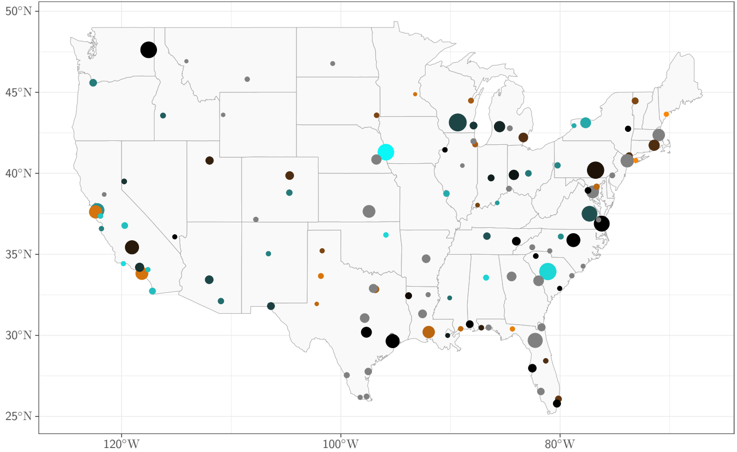
$$\Theta = \begin{pmatrix} 4 & 0 & -4 & 0 \\ 0 & 2 & -2 & 0 \\ -4 & -2 & 7 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

The data set

- Domestic flight data from the U.S. Bureau of Transportation Statistics[1]
- Airports and airlines with $\geq 1\%$ marketshare
- Years 2005-2020
- Filtered:
    - Airports with at least 1000 flights per year
    - in the contiguous United States
- Total positive delays aggregated per day
- Resulting in $n = 5601$ observations of $d = 118$ airports

[1] https://www.bts.dot.gov/browse-statistical-products-and-data/bts-publications/airline-service-quality-performance-234-time

The data set – Illustration

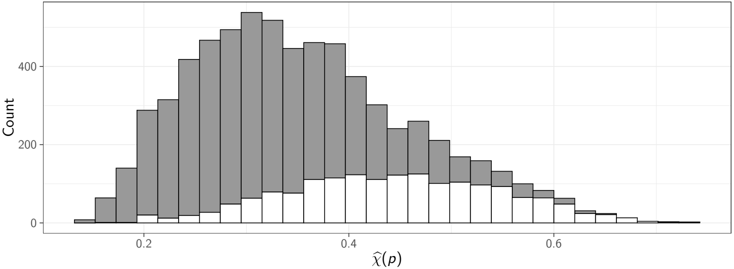


- Size proportional to daily flights
- Color indicates shape of univariate generalized Pareto MLE:
    - Cyan < 0
    - Black $= 0$
    - Orange > 0

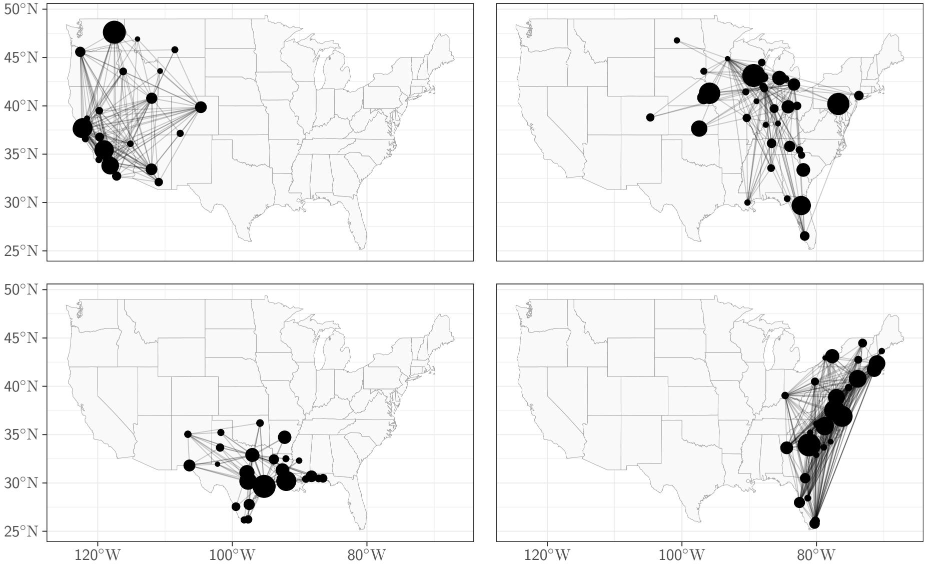
$\rightarrow$ Normalized marginals from here

Clustering

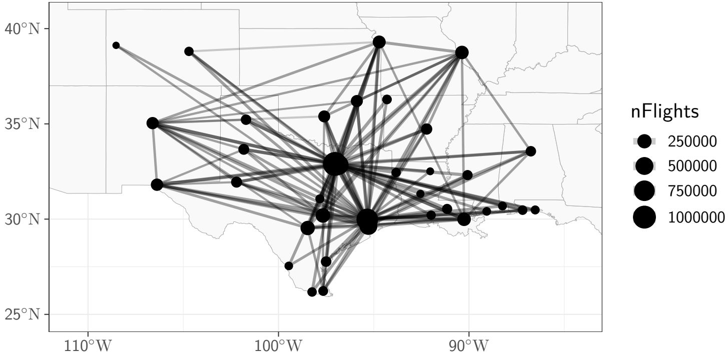
- Ensuring (strong) extremal dependence within the analyzed set of airports
- Using k-medoids clustering
  - Empirical extremal correlation $\hat{\chi}(p = 0.85)$ as (dis)similarity measure
  - $k = 4, \ldots, 7$ yields "reasonable" results
- Clusters show strong geographic proximity (even though no explicit geographic information is used!)



Clustering – Results



Clustering – The “Texas cluster”



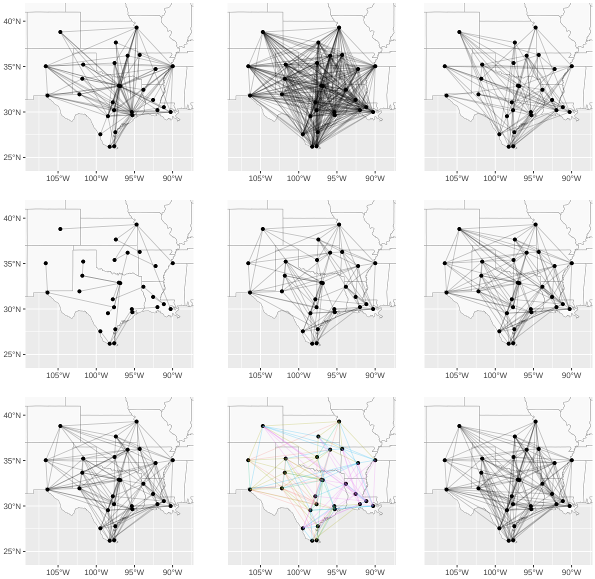
- Centered around Houston, Dallas
- 29 Airports
- 83 Pairs with ($\geq$ monthly) connections
- Size proportional to daily flights

Estimation

- Train/test split: before/after 2010
- Structure learning:
    - "Flight graph"
    - Full graph
    - Random graph
    - Minimum spanning tree ("EMST")
    - Majority voting ("EGLearn")
    - Parameter shift

  $\rightarrow$ Use matrix completion with empirical variogram at threshold $p = 0.95$ for parameters
- Joint structure and parameter learning:
    - $EMTP_2$
    - Score matching
- Colored graphical model on the $EMTP_2$ graph

Results – Graph structures



(Flight graph, full variogram, random graph)

(EMST, EGLearn, $EMTP_2$)

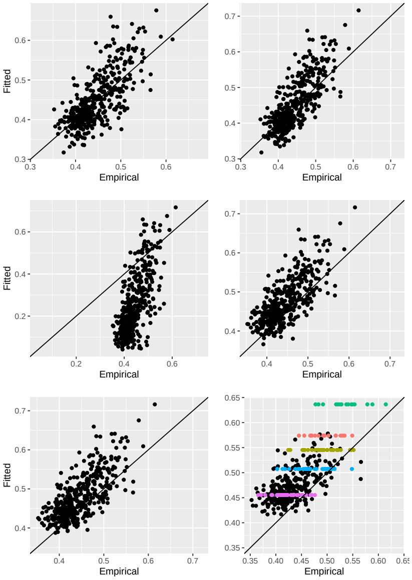
(Parameter shift, colored, score matching)

Results – Edge counts and validation likelihoods

	Edges	Log-likelihood
Flight Graph	129	-9.23
Full Variogram	406	-293.68
Random Graph	100	-713.79
EMST	28	-6435.20
EGLearn	101	1123.45
$EMTP_2$	126	1212.97
Parameter Shift	142	1192.02
Colored Graph	126	1341.17
Score Matching	173	1144.97

- Hüsler–Reiss likelihoods with standardized margins, computed on the validation set
- Random graph was set to have 100 edges
- EMST must have $d - 1 = 28$ edges

Results – Implied χ



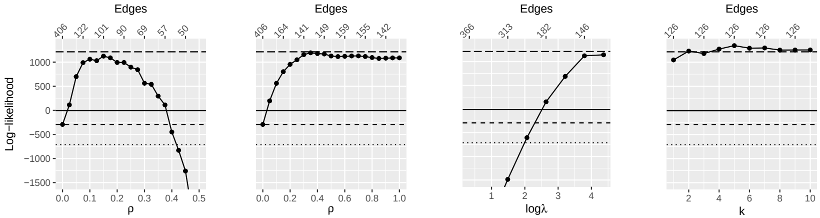
(Flight graph, full variogram, random graph)

(EMST, EGLearn, EMTP$_2$)

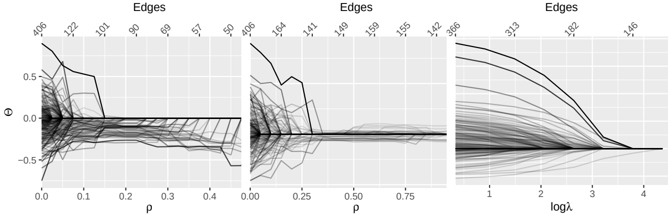
(Parameter shift, colored, score matching)

Results – Hyperparameters

Validation log-likelihoods for different hyperparameters:



Convergence of entries in Θ to zero for penalized methods:



(EGLearn, parameter shift, score matching, [colored])

Discussion points/open questions

- Stochastic interpretation of extremal conditional independence?
- Relation to pre-asymptotic graphical models, domain of attraction?
  - Do pre-asymptotic graphical models converge to extremal graphical models?
- Focus on Hüsler–Reiss distribution
  - Comparable to focus on Gaussian graphical models?
  - Corresponding limit theorem?
  - Do results translate to general covariance/variogram statements?
  - "Good enough approximation" of other distributions?
- Disconnected graphical models
  - "Asymptotic independence" very different from normal independence!
  - Do methods for connected models transfer to disconnected ones?

References

Asenova, S., Mazo, G., and Segers, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes*, 24:461–500.

Asenova, S. and Segers, J. (2021). Extremes of Markov random fields on block graphs. Available from https://arxiv.org/abs/2112.04847.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Engelke, S., Ivanovs, J., and Strokorb, K. (2022a). Graphical models for infinite measures with applications to extremes and lévy processes.

Engelke, S., Lalancette, M., and Volgushev, S. (2022b). Learning extremal graphical structures in high dimensions. Available from https://arxiv.org/abs/2111.00840.

Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. Available from https://arxiv.org/abs/2012.06179.

Hentschel, M., Engelke, S., and Segers, J. (2022). Statistical inference for hüsler-reiss graphical models through matrix completions. Available from https://arxiv.org/abs/2210.14292.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Prob. Letters*, 7(4):283–286.

Lederer, J. and Oesting, M. (2023). Extremes in high dimensions: Methods and scalable algorithms. Available from https://arxiv.org/abs/2303.04258.

Röttger, F., Engelke, S., and Zwiernik, P. (2021). Total positivity in multivariate extremes. Available from https://arxiv.org/abs/2112.14727.

Röttger, F., Coons, J. I., and Grosdos, A. (2023). Parametric and nonparametric symmetries in graphical models for extremes. Available from https://arxiv.org/abs/2306.00703.

Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.*, 14(1):138–150.

Wan, P. and Zhou, C. (2023). Graphical lasso for extremes. Available from https://arxiv.org/abs/2307.15054.

Ying, J., Cardoso, J. M., and Palomar, D. (2021). Minimax estimation of Laplacian constrained precision matrices. In *Int. Conf. Artif. Intell. Stat.*, pages 3736–3744. PMLR.

Ying, J., Cardoso, J. V. d. M., and Palomar, D. P. (2020). Does the ℓ_1-norm learn a sparse graph under Laplacian constrained graphical models? Available from https://arxiv.org/abs/2006.14925.

Thank you!