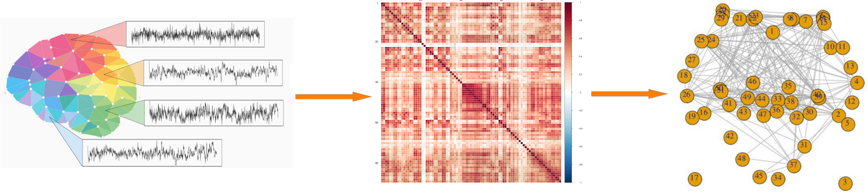
Génération de Matrices de Corrélation avec des Structures de Graphe par Optimisation Convexe

Ali Fakhar, **Kévin Polisano**, Irène Gannaz, Sophie Achard

CNRS - UGA, Laboratoire Jean Kuntzmann

29 août 2025

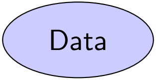
Motivation ∼ Graphical modeling



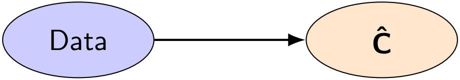
The objective is to model dependence using graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, p\}$ denotes the set of nodes and $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V}\}$ the set of edges.

- Each node $i = 1, \dots, p$ is associated to a random variable X_i.
- We consider the correlation matrix $\mathbf{C} = (\mathrm{corr}(X_i, X_j))_{i,j=1,\dots,p}$.
- There is an edge between nodes (i,j), $i \sim j$, iff $c_{i,j} \neq 0$.

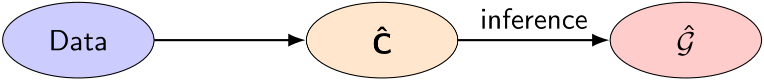
Motivation ∼ Validation of Graph Inference



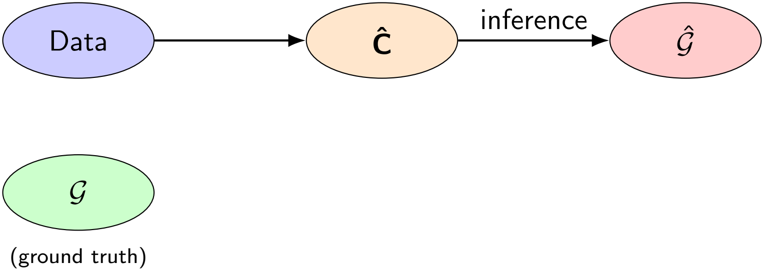
Motivation ∼ Validation of Graph Inference



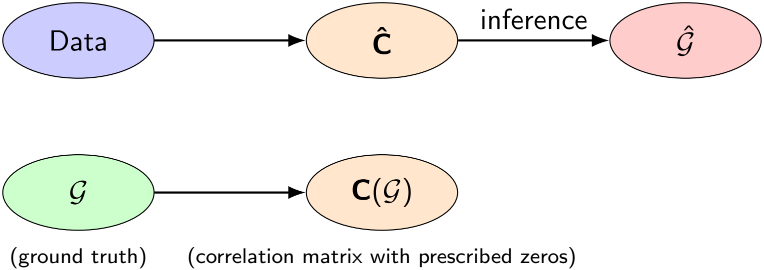
Motivation $\sim$ Validation of Graph Inference



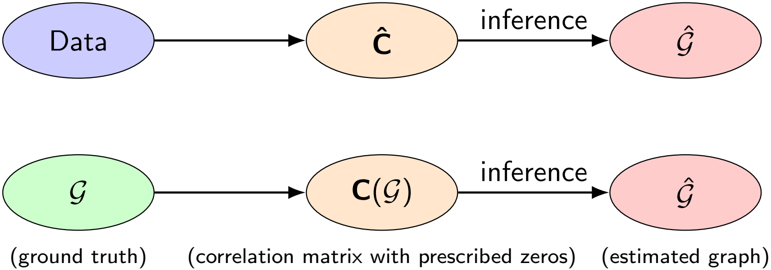
Motivation ∼ Validation of Graph Inference



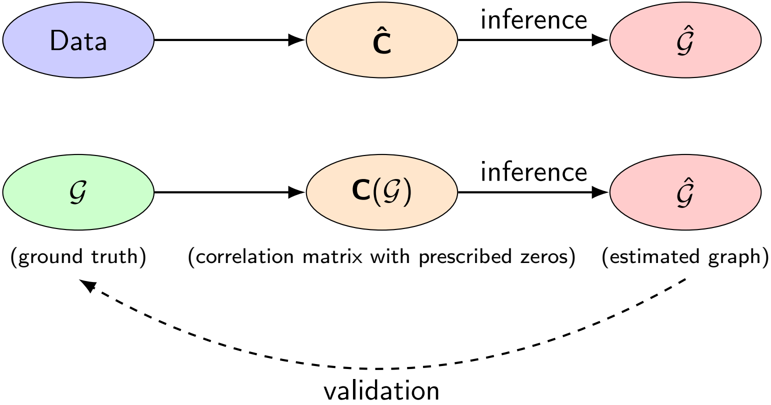
Motivation ∼ Validation of Graph Inference



Motivation $\sim$ Validation of Graph Inference



Motivation ∼ Validation of Graph Inference



Motivation $\sim$ Generation of Correlation Matrices

Challenge

Generate **theoretical** correlation matrices $\mathbf{C}$ associated with a specific graph $\mathcal{G}$ for benchmarking purposes.

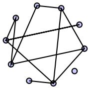
Objective

- Propose a flexible method for constructing $\mathbf{C} \in \mathcal{C}(\mathcal{G})$ that remains efficient in high dimensions.
- Control the average correlation in order to assess benchmark quality and to better mimic real data.

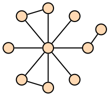
Key Definitions

- $\mathbf{C}$ in $\mathbb{R}^{p \times p}$, symmetric, PSD, $c_{ii} = 1$, $|c_{ij}| \leq 1$.
- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $c_{ij} = 0$ if $(i,j) \notin \mathcal{E}$ (or $(i,j) \in \overline{\mathcal{E}}$).

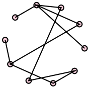
Five Key Graph Models



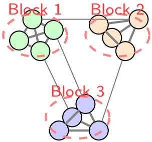
Erdős–Rényi Model



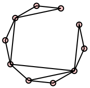
Barabási-Albert Model



Watts-Strogatz Model

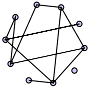


Stochastic Block Model

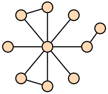


Chordal Graph

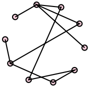
Five Key Graph Models



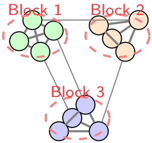
Erdős–Rényi Model



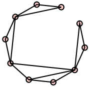
Barabási-Albert Model



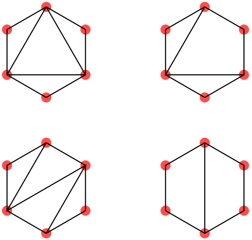
Watts-Strogatz Model



Stochastic Block Model



Chordal Graph



Chordal non-chordal

Outline

1. Motivations
2. Generation procedure
   - Related work
   - Proposed method
3. Results
   - Comparison with other methods
   - Getting close to real data
   - Effect of the graph structure
4. Conclusion

Related work

Diagonal Dominance

- Increase the diagonal entries of $\mathbf{C}$, then normalize.
- **Con:** Produces very small correlations.

Cholesky Decomposition

- $\mathbf{C} = \mathbf{U}\mathbf{U}^\top$, with $\mathbf{U}$ generated via polar parametrization (Pourahmadi and Wang, 2015) or Metropolis–Hastings (Córdoba, 2018).
- **Pro:** Enables uniform sampling in $\mathcal{C}(\mathcal{G})$.
- **Con:** Applicable only to chordal graphs.

Partial Orthogonalization

- Orthogonalize rows according to $\overline{\mathcal{E}}$ using the Gram–Schmidt algorithm (Córdoba, 2020).
- **Con:** Highly sensitive to initialization.

Proposed Method

Ali Fakhar, Kévin Polisano, Irène Gannaz, and Sophie Achard (June 2025).
"Generating Correlation Matrices with Graph Structures Using Convex Optimization".
In: IEEE Statistical Signal Processing Workshop (SSP). Edinbourg, United Kingdom

Generate $\mathbf{C} \in \mathcal{C}(\mathcal{G})$ by solving:

Matrix Completion Problem

$$\begin{aligned} \underset{\mathbf{C}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{C} - \bar{\mathbf{C}}\|_F^2 \\ \text{subject to} \quad & \mathbf{C} \succeq 0, \quad \operatorname{diag}(\mathbf{C}) = \mathbf{1}, \quad |c_{ij}| \leq 1 \\ & c_{ij} = 0, \quad \forall (i,j) \in \overline{\mathcal{E}} \end{aligned}$$

where $\bar{\mathbf{C}}$ denotes the initialization matrix.

Proposed Method

Ali Fakhar, Kévin Polisano, Irène Gannaz, and Sophie Achard (June 2025).
"Generating Correlation Matrices with Graph Structures Using Convex Optimization".
In: IEEE Statistical Signal Processing Workshop (SSP). Edinbourg, United Kingdom

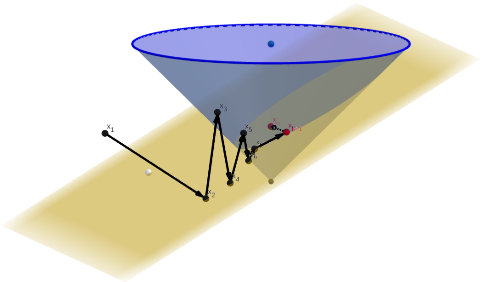
Generate $\mathbf{C} \in \mathcal{C}(\mathcal{G})$ with **additional constaints**:

Matrix Completion Problem

$$\begin{aligned} \underset{\mathbf{C}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{C} - \bar{\mathbf{C}}\|_F^2 \\ \text{subject to} \quad & \mathbf{C} \succeq 0, \quad \text{diag}(\mathbf{C}) = \mathbf{1}, \quad |c_{ij}| \leq 1 \\ & c_{ij} = 0, \quad \forall (i,j) \in \overline{\mathcal{E}} \\ & \frac{1}{2|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} c_{ij} \geq b \quad \text{(Mean constraint)} \end{aligned}$$

Solving the Convex Optimization Problem

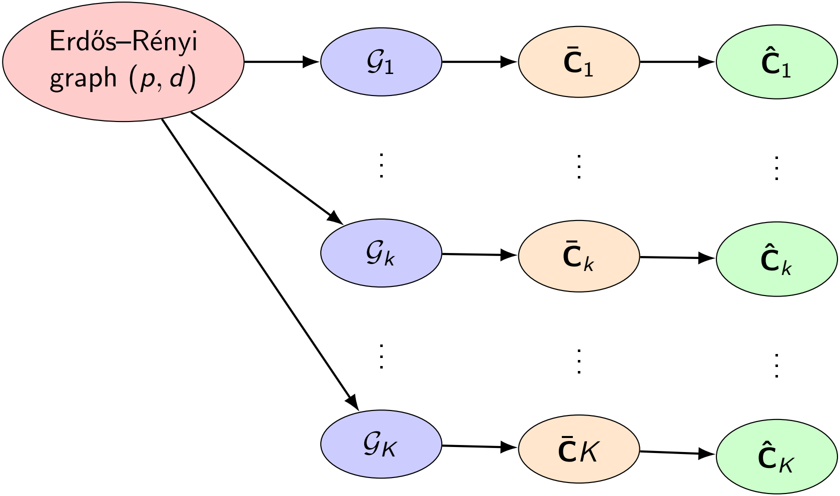
- Solved via **alternating projections method**



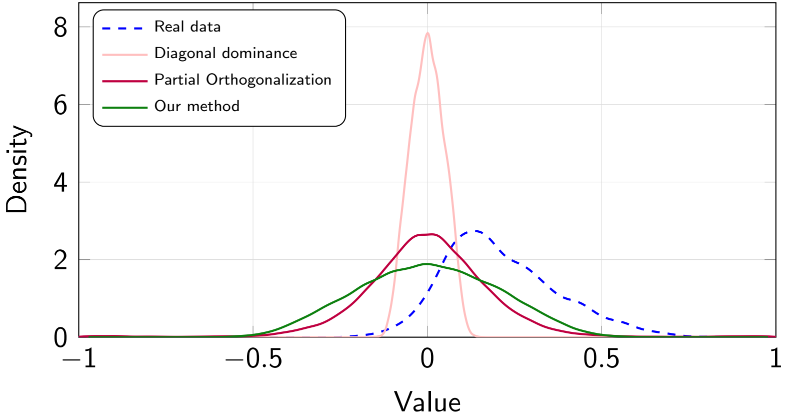
Iterative optimization procedure involving multiple projections. The sequence of black points $x_1, x_2, \ldots, x_k$ illustrates the intermediate iterates. The red point is a solution in $\mathcal{C}(\mathcal{G})$.

- Solved via **interior-point method** (CVXPY)

Sampling procedure from a given Graph Model

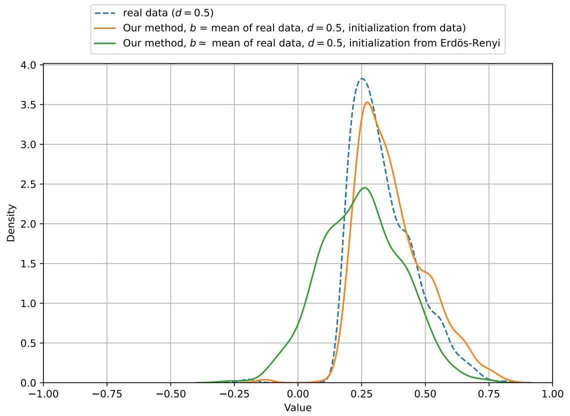


Comparison with other Methods



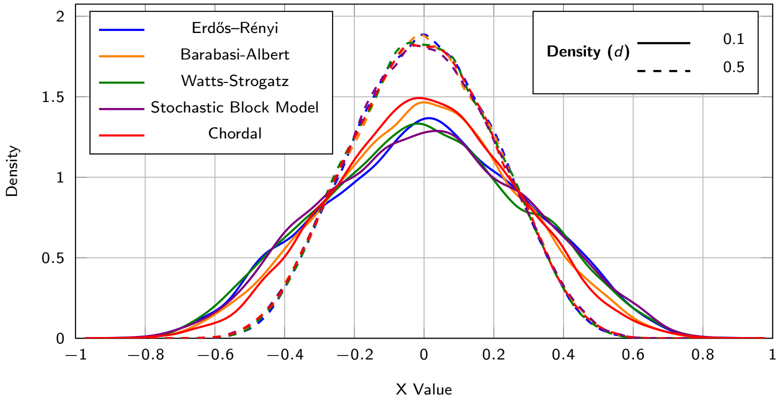
Distribution of non-zero, off-diagonal correlation values averaged over $K = 50$ Erdős–Rényi graphs ($p = 51$, $d = 0.5$). Data obtained for $\mathbf{C} \in \mathbb{R}^{51\times 51}$ with $b = -1$.

Getting Close to Real Data



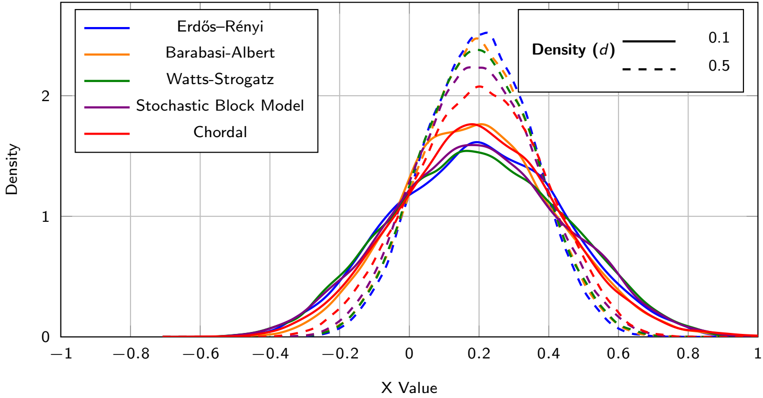
Real data: correlation of cerebral connectivity, thresholded to impose $d = 0.5$ (correlations of wavelet coefficients in the [0.06, 0.12] Hz band from fMRI data of a live rat).

Effect of Graph Structure



Distribution of correlation values averaged over $K = 50$ graphs, for various graph structures and densities (d). Data obtained for $\mathbf{C} \in \mathbb{R}^{51\times51}$ with $b = -1$.

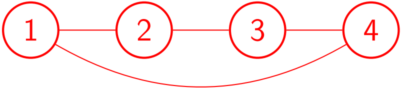
Effect of Graph Structure



Distribution of correlation values averaged over $K = 50$ graphs, for various graph structures and densities (d). Data obtained for $\mathbf{C} \in \mathbb{R}^{51\times51}$ with $b = 0.2$

SDP Matrix Completion

$$\mathbf{A}(\mathcal{G}) = \begin{pmatrix} a_{11} & a_{12} & ? & a_{14} \\ a_{21} & a_{22} & a_{23} & ? \\ ? & a_{32} & a_{33} & a_{34} \\ a_{41} & ? & a_{43} & a_{44} \end{pmatrix}$$



$\mathcal{G}$

$$X \succeq 0, \quad x_{ij} = a_{ij}, \quad (i,j) \in \mathcal{G}$$

Existence of a Solution

Definition ($\mathcal{G}$–partial positive)

A matrix $\mathbf{A}(\mathcal{G}) = [a_{ij}]_{\mathcal{G}}$ is $\mathcal{G}$–partial positive if $a_{ji} = \overline{a_{ij}}$ for all $(i, j) \in \mathcal{E}$, and for every clique C of $\mathcal{G}$, the principal submatrix $[a_{ij} : i, j \in C]$ of $\mathbf{A}(\mathcal{G})$ is positive definite.

Definition (Graph *completable*)

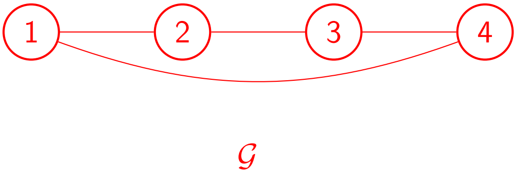
A graph $\mathcal{G}$ is *completable* if and only if every $\mathcal{G}$–partial positive matrix admits a positive completion.

Theorem (Grone, 1984)

$\mathcal{G}$ is *completable* $\Leftrightarrow$ $\mathcal{G}$ is chordal.

SDP Matrix Completion

$$\mathbf{A}(\mathcal{G}) = \begin{pmatrix} 1 & 1 & ? & 0 \\ 1 & 1 & 1 & ? \\ ? & 1 & 1 & 1 \\ 0 & ? & 1 & 1 \end{pmatrix}$$

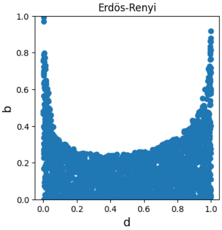


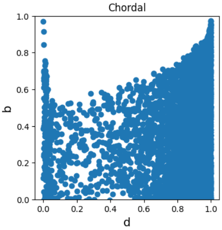
$$X \succeq 0, \quad x_{ij} = a_{ij}, \quad (i,j) \in \mathcal{G}$$

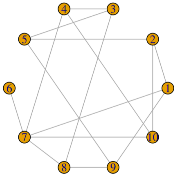
$\mathcal{G}$ is not **chordal**. There is no completion for $\mathbf{A}(\mathcal{G})$!

Feasibility Regions

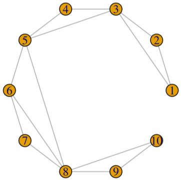
w.r.t the graph density $d = 2|\mathcal{E}|/(p(p-1))$ and the mean b:







Erdős–Rényi



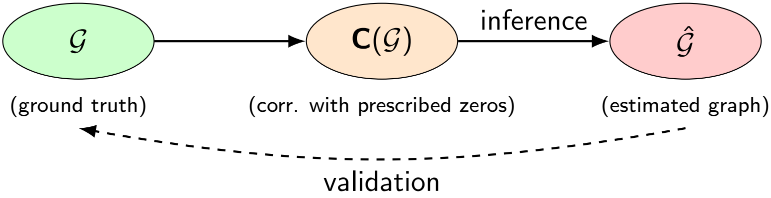
Chordal

Conclusion & Ongoing work

Conclusion

Procedure for generating correlation matrices with a given graph $\mathcal{G}$:

- Applicable to any graph structure.
- Produces larger correlation values than other algorithms.
- Allows the inclusion of additional constraints.
- Initial results on graph inference procedures.



Conclusion & Ongoing work

Conclusion

Procedure for generating correlation matrices with a given graph $\mathcal{G}$:

- Applicable to any graph structure.
- Produces larger correlation values than other algorithms.
- Allows the inclusion of additional constraints.
- Initial results on graph inference procedures.

Alice Chevaux, Ali Fahkar, Kévin Polisano, Irène Gannaz, and Sophie Achard (Sept. 2025). "Benchmarking Brain Connectivity Graph Inference: A Novel Validation Approach". In: 33rd European Signal Processing Conference (EUSIPCO 2025). Palerme, Italy. URL: https://hal.science/hal-04995510

Conclusion & Ongoing work

Conclusion

Procedure for generating correlation matrices with a given graph $\mathcal{G}$:

- Applicable to any graph structure.
- Produces larger correlation values than other algorithms.
- Allows the inclusion of additional constraints.
- Initial results on graph inference procedures.

Perspective

- Higher-dimensional settings.
- Theoretical guarantees for the existence and the sampling.
- More thorough study of inference procedures.

References I

Chevaux, Alice, Ali Fahkar, Kévin Polisano, Irène Gannaz, and Sophie Achard (Sept. 2025). "Benchmarking Brain Connectivity Graph Inference: A Novel Validation Approach". In: 33rd European Signal Processing Conference (EUSIPCO 2025). Palerme, Italy. URL: https://hal.science/hal-04995510.

Córdoba, Irene, Gherardo Varando, Concha Bielza, and Pedro Larrañaga (2018). "A fast Metropolis-Hastings method for generating random correlation matrices". In: International Conference on Intelligent Data Engineering and Automated Learning. Springer, pp. 117–124.

— (2020). "On generating random Gaussian graphical models". In: International Journal of Approximate Reasoning 125, pp. 240–250.

Fakhar, Ali, Kévin Polisano, Irène Gannaz, and Sophie Achard (June 2025). "Generating Correlation Matrices with Graph Structures Using Convex Optimization". In: IEEE Statistical Signal Processing Workshop (SSP). Edinbourg, United Kingdom.

References II

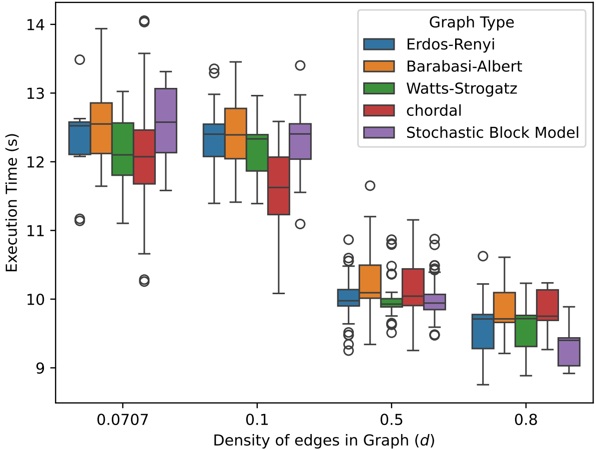
Grone, Robert, Charles R Johnson, Eduardo M Sá, and Henry Wolkowicz (1984). "Positive definite completions of partial Hermitian matrices". In: Linear algebra and its applications 58, pp. 109–124.

Pourahmadi, Mohsen and Xiao Wang (2015). "Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor". In: Statistics & Probability Letters 106, pp. 5–12.

Questions?

Thank you!

Time execution



Sampling 3 × 3 full correlation matrices

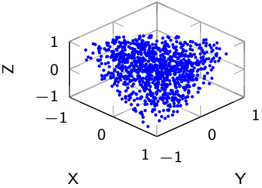
$\bar{\mathbf{C}}_k$ sampled with unif or Pourahmadi

$\downarrow$

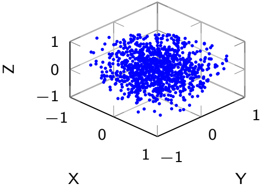
$$\hat{\mathbf{C}}_k = \begin{pmatrix} 1 & x_k & y_k \\ x_k & 1 & z_k \\ y_k & z_k & 1 \end{pmatrix}$$

$\downarrow$

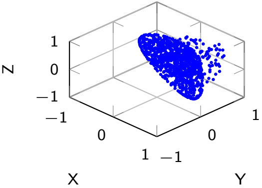
$p_k = (x_k, y_k, z_k)$



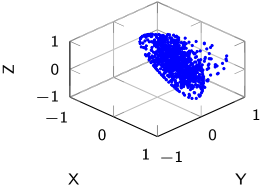
Pourahmadi ($b = -1$)



Uniform ($b = -1$)

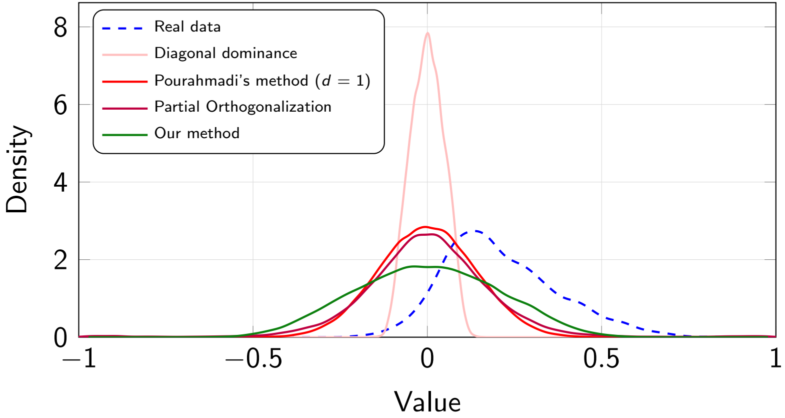


Pourahmadi ($b = 0.2$)



Uniform ($b = 0.2$)

Comparison for Chordal Graphs



Distribution of non-zero, off-diagonal correlation values averaged over $K = 50$ chordal graphs ($p = 51$, $d = 0.5$). Data obtained for $\mathbf{C} \in \mathbb{R}^{51 \times 51}$ with $b = -1$.

Benchmark for graph inference

Context

$\mathcal{G}$ graph and $\mathbf{C} \in \mathcal{C}(\mathcal{G})$

$X_i \in \mathbb{R}^p$, $i = 1, \ldots, n$ iid from $\mathcal{N}(0, \mathbf{C})$

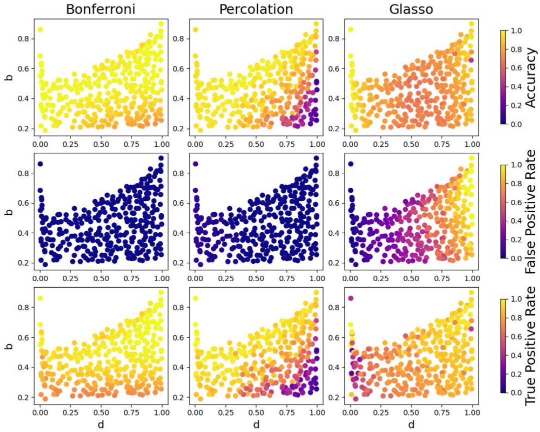
Problem

Inference of $\mathcal{G}$ via (X_i)? $\hookrightarrow$ Simulation study

Parameters

- 100 simulations for each graph
- $n = 1000$ observations
- $p = 51$ nodes
- graph structure: chordal
- d graph density - varying
- b mean constraint on non-zero correlations - varying

Benchmark for graph inference



(Chevaux, 2025)

Related works - (1) Diagonal Dominance

Method:

- Construct a symmetric matrix $\tilde{\mathbf{C}}$
- If $(i,j) \in \mathcal{E}^c$, then $\tilde{c}_{ij} = 0$
- Update rule:

$$\tilde{c}_{ii} \leftarrow \sum_{\substack{j=1,\dots,p \\ i \neq j}} |\tilde{c}_{ij}| + \text{random positive perturbation}$$

  Gershgorin theorem.

- $\mathbf{C} = \mathrm{diag}(\tilde{\mathbf{C}})^{-1/2} \tilde{\mathbf{C}} \mathrm{diag}(\tilde{\mathbf{C}})^{-1/2}$

Drawback: Yields correlation matrices with very low off-diagonal values.

Related works - (2) Cholesky decomposition

- (Pourahmadi and Wang, 2015) use Cholesky decomposition and polar transformation, using angles as random variables.
- Correlation matrix is $\mathbf{C} = \mathbf{L}\mathbf{L}^\top$, where $\mathbf{L}$ is:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \cos\theta_{21} & \sin\theta_{21} & 0 & \dots & 0 \\ \cos\theta_{31} & \cos\theta_{32} & \sin\theta_{31}\sin\theta_{32} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cos\theta_{n1} & \cos\theta_{n2} & \sin\theta_{n1}\cos\theta_{n3}\sin\theta_{n2}\sin\theta_{n1} & \dots & \prod_{k=1}^{n-1}\sin\theta_{nk} \end{bmatrix}.$$

Related works - (3) Partial Orthogonalization

- **Approach:** Starts with an initial matrix $\mathbf{C}$ (with zeros in desired pattern $(\mathcal{E}')^c$).
- **Process:** Iteratively removes additional edges using a modified Gram-Schmidt-based partial orthogonalization.
- **Mechanism:** Writes $\mathbf{C} = \mathbf{Q}\mathbf{Q}^\top$ and orthogonalizes each row $\mathbf{q_i}$ with respect to rows $\{\mathbf{q_j}$ s.t. $(i,j) \in \mathcal{E}$ and $j < i\}$.