Generating Correlation Matrices with Graph Structures via Convex Optimization

Ali Fakhar[1], Kévin Polisano[1], Irène Gannaz[2], Sophie Achard[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, Inria, LJK, F-38000 Grenoble, France
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France

Introduction & Motivation

Graphical models represent dependencies (e.g., genetics, neuroscience). Inference often uses correlation ($\mathbf{C}$) or precision matrices.
Challenge: Generate *theoretical* $\mathbf{C}$ with specific sparsity (graph $\mathcal{G}$) for benchmarking. Existing methods often yield mean-zero correlations.
Our Goal: Flexible method for $\mathbf{C} \in \mathcal{C}(\mathcal{G})$ with controlled mean, mimicking real data (e.g., brain connectivity [1]).

Key Definitions:
- $\mathbf{C}$: symmetric, PSD, $c_{ii} = 1$, $|c_{ij}| \le 1$.
- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: sparsity means $c_{ij} = 0$ if $(i,j) \notin \mathcal{E}$.
- Graph density (d): $2|\mathcal{E}|/(p(p-1))$.

Related Work

Cholesky/Polar Parametrizations [2, 6]:
- $\mathbf{C} = \mathbf{U}\mathbf{U}^\top$.
- **Pro:** Uniform sampling in $\mathcal{C}(\mathcal{G})$.
- **Con:** Only for chordal graphs.

Diagonal Dominance [3]:
- Diagonally dominant $\tilde{\mathbf{C}}$, then normalize.
- **Pro:** Works for any $\mathcal{G}$.
- **Con:** Very small correlations.

Partial Orthogonalization [2]:
- Orthogonalize rows based on $\mathcal{E}$.
- **Pro:** Non-chordal graphs.
- **Con:** Sensitive to initialization, mean-zero matrices.

Proposed Approach: Convex Optimization

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

$$\begin{aligned} \underset{\mathbf{C}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{C} - \bar{\mathbf{C}}\|_F^2 \\ \text{subject to} \quad & \mathbf{C} \succeq 0 \quad \text{(PSD)} \\ & \operatorname{diag}(\mathbf{C}) = \mathbf{1} \\ & |c_{ij}| \le 1, \quad \forall (i,j) \\ & c_{ij} = 0, \quad \forall (i,j) \notin \mathcal{E} \\ & \underbrace{\frac{1}{2|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} c_{ij} \ge b}_{\text{Novelty: Mean constraint}} \quad (*) \end{aligned}$$

$\bar{\mathbf{C}}$ = target matrix, b = mean of edge correlations.

Key Aspects:
- Solved via convex optimization (CVXPY [4]).
- Objective: feasible $\mathbf{C}$ close to $\bar{\mathbf{C}}$.
- Constraint $(*)$ controls mean via b.
- Unique solution if feasible ($b \le 0$ guarantees feasibility).
- **Post-processing:** If needed, $\tilde{\mathbf{C}}_\epsilon = \tilde{\mathbf{C}} + \epsilon\mathbf{I}$, $\mathbf{C} = \tilde{\mathbf{C}}_\epsilon/(1+\epsilon)$.

Results: Comparison & Mean Control

Simulations: $p = 51$, 50 runs. Target $\bar{\mathbf{C}} \sim \mathrm{U}(-1, 1)$ unless stated.

Comparison (ER, $d = 0.5, b = -1$):

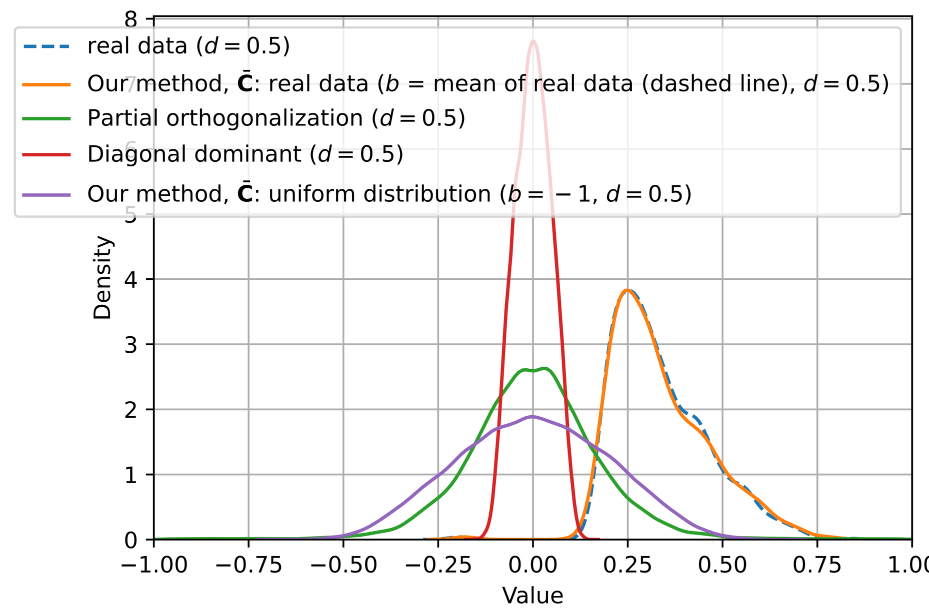

Figure 1: Density plots. Our method (purple) vs. others (red, green). Approximates real fMRI (blue) when tuned (orange).

Effect of Mean Constraint (b):
- Fig. 1 (orange) fits fMRI [5] using empirical b and $\bar{\mathbf{C}}$.
- Feasibility depends on b and graph density d.

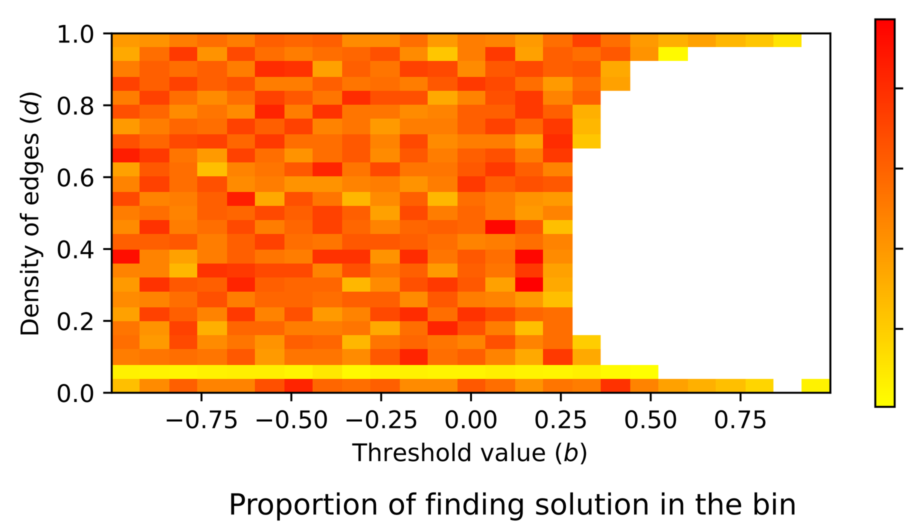

Figure 2: Feasibility map for ER graphs ($p = 51$). White = infeasible (d, b).

Results: Influence of Graph Structure

Tested: ER, BA, WS, SBM, Chordal.

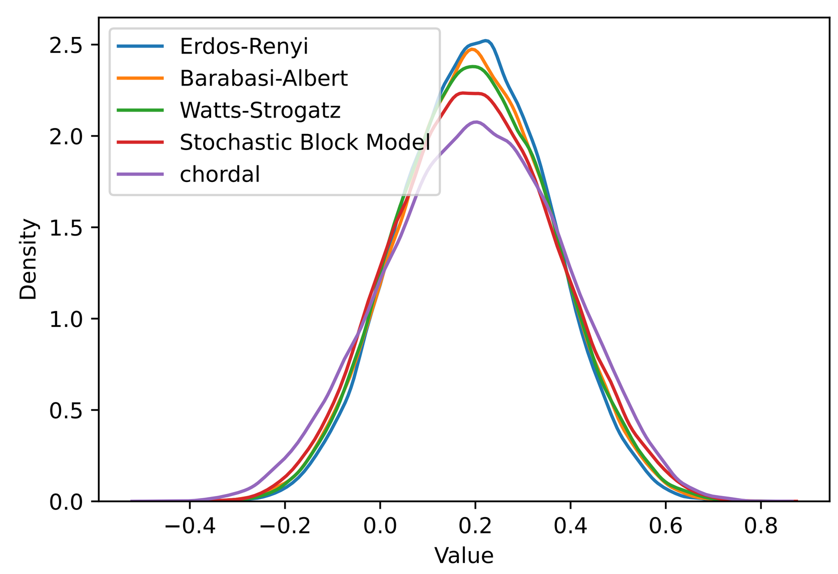

Figure 3: Density plots ($p = 51$, $d = 0.5$, $b = 0.2$). Minimal impact of graph structure at this scale.

Observations:
- For $p = 51$, graph type has limited impact on correlation distribution given fixed d and b.
- Influence may increase for larger p.
- More computationally intensive but better control over statistical properties.

Conclusion

- Convex optimization framework to generate $\mathbf{C}$ with graph structure $\mathcal{G}$.
- **Key advantage:** Explicit control over mean correlation (b) for realistic data generation.
- **Improvement:** Supports any graph, avoids near-zero correlations, introduces mean control.
- Application: benchmarking inference methods on controlled graph structures.
- Future: scaling to large p, deeper study of graph structure role.

Code: [7]

Acknowledgements & References

Funding: ANR France 2030 (ANR-23-IACL-0006), GRICAD infrastructure.

Selected References:

[1] S. Achard et al. (2022). NeuroImage.
[2] I. Córdoba et al. (2020). J Comput Graph Stat.
[3] I. Córdoba et al. (2020). TEST.
[4] S. Diamond & S. Boyd (2016). JMLR.
[5] G. Becq et al. (2020). NeuroImage, 219, 116945.
[6] M. Pourahmadi (2015). Stat & Prob Let.
[7] A. Fakhar et al. Code. https://github.com/alifakhar7/corr_mat_gen

LaTeX TikZposter