Prior sensitivity in Bayesian structure learning of Gaussian Graphical Models

Marwin Carmo & Philippe Rast

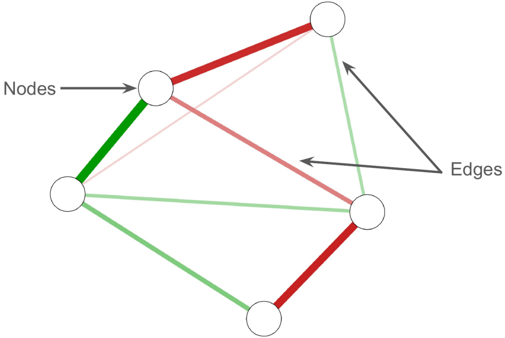
University of California Davis

January 9, 2026

Introduction

Introduction

- Gaussian Graphical Models (GGMs), or partial correlation networks, help us visualize the unique relationships between variables while controlling for all other variables in the model.
- A key goal of GGMs estimation is **structure learning**: determining which edges exist based on the data.



Introduction

- The GGM for a given graph $G = (V, E)$ can be defined as the family of distributions
$$\mathcal{N}_G = \{N(0, \Sigma) : K = \Sigma^{-1} \in P_G\}$$
- P_G denotes the cone of positive definite matrices whose zero entries corresponds to the absence of edges in G.
- The partial correlation $\rho_{ij \cdot \mathbf{V} \setminus \{i,j\}}$ between variables i and j is calculated by standardizing the off-diagonal elements of the precision matrix;

$$\rho_{ij \cdot \mathbf{V} \setminus \{i,j\}} = -\frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}}$$

The Bayesian Approach

- **Why Bayesian Methods?** They allow us to quantify our uncertainty and incorporate prior knowledge into the model.
- New Bayesian methods are as efficient and accurate as frequentist methods in structure learning, even for large-scale applications.
- **Prior sensitivity**: The final estimated network can be highly sensitive to the chosen prior distributions.

Motivation

- For researchers, it's often unclear how to choose these priors or how much an incorrect choice will affect the results.
- There are several distinct approaches to Bayesian Structure Learning (BSL), but little is known about their comparative performance under circumstances of **prior misspecification**.
- The goal of this project is to evaluate the performance of popular and accessible Bayesian algorithms under a range of prior settings and to investigate how these choices influence the inferred graph structure and precision matrix.

Algorithms Under Investigation

- **BGGM:** Tests individual partial correlations using BF and Matrix-F as the prior for $\mathbf{K}$.
- **SSSL:** Stochastic Search Structure Learning using continuous spike-and-slab priors with latent indicators to identify the graph by shrinking unimportant precision matrix elements towards zero.
- **BDMCMC:** Uses a Birth-Death MCMC search over the graph space, accelerated by a closed-form approximation for the G-Wishart normalizing constant ratio.
- **RJ-WWA:** Reversible jump MCMC is used for moving between models of different dimensions. Prioritize proposals that are more likely to increase the posterior probability.

Summary of Priors for Considered GGM Algorithms

Priors in Bayesian Structure Learning Algorithms for GGMs:

Algorithm	**Graph Prior**	**Parameter Prior**
BGGM (Williams & Mulder, 2020)	Not modeled directly: $p(\rho_{ij} = 0, \boldsymbol{\phi}_{ij} \mid \rho_{ij} \in (-1, 1))$	Matrix-F: $\mathbf{K} \sim F(\nu, \delta, \mathbf{B})$
SSSL (Wang, 2015)	Mixture on a_{ij}: $(1-\pi)N(0, v_0^2) + \pi N(0, v_1^2)$	Spike & Slab on elements: $N(a_{ij} \vert 0, v_{z_{ij}}^2)$.
BDMCMC (Mohammadi et al., 2023)	Bernoulli on edges: $P(G) \propto \pi^{\vert E\vert}(1-\pi)^{m-\vert E\vert}$.	G-Wishart on $\mathbf{K}$: $\mathbf{K}\vert G \sim \mathcal{W}_G(\delta, \mathbf{D})$.
RJ-WWA (van den Boom et al., 2022)	Bernoulli on edges: $p(G) \propto \pi^{\vert E\vert}(1-\pi)^{m-\vert E\vert}$.	G-Wishart on $\mathbf{K}$: $\mathbf{K}\vert G \sim \mathcal{W}_G(\delta, \mathbf{D})$.

Methods

Experimental Conditions

Data Characteristics:

- **Dimensionality (p):**
  10, 25, 100 nodes
- **Sample Size (n):**
  Small ($n = p$)
  Medium ($n = 5p$)
  Large ($n = 20p$)
- **True Graph Structure:**
  Random, Small-World
- **True Graph Density:**
  Sparse (20%)
  Average (50%)
  Dense (80%)
- **Replications:** 25

Prior Misspecification:

- **G-Wishart** (648 conditions):
  - $\delta = \{3, 10\}$
  - $\pi = \{0.2, 0.5, 0.8\}$
- **Matrix-F** (216 conditions):
  - $\delta = \{0.2, 0.5\}$
- **Spike and Slab** (648 conditions):
  - $v_0 = \{0.02, 0.1\}$
  - $v_1 = h \cdot v_0, \quad h = \{10, 50\}$
  - $\pi = \{0.2, 0.5, 0.8\}$

Selected Performance Metrics

- **F1-Score:** The harmonic mean of Precision and Sensitivity.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}, \quad \text{where}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{Precision} = \frac{TP}{TP + FP}$$

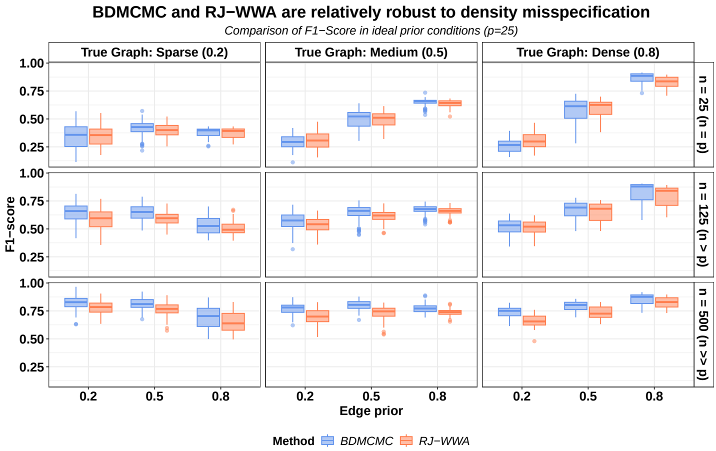
- **Relative RMSE:** The root mean squared error of the partial correlations (ρ_{ij}), normalized by the average magnitude of the true partial correlations.

$$\text{Relative RMSE} = \frac{\sqrt{\frac{1}{M}\sum_{i<j}(\hat{\rho}_{ij} - \rho_{ij}^{\text{true}})^2}}{\frac{1}{M}\sum_{i<j}|\rho_{ij}^{\text{true}}|}$$

Results

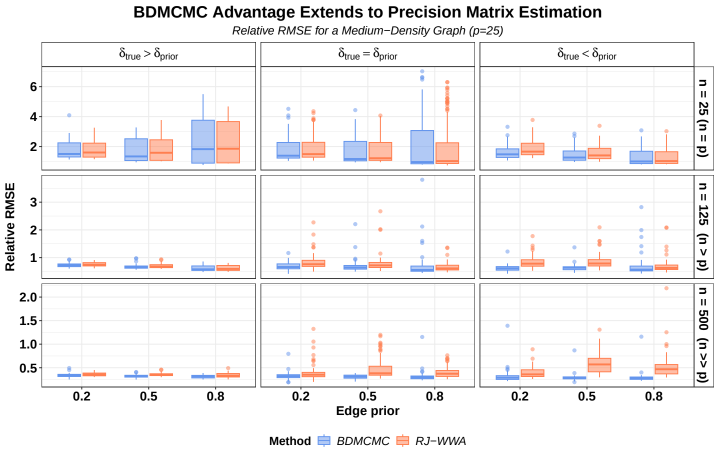
Results

The G-Wishart Approach: BDMCMC & RJ-WWA



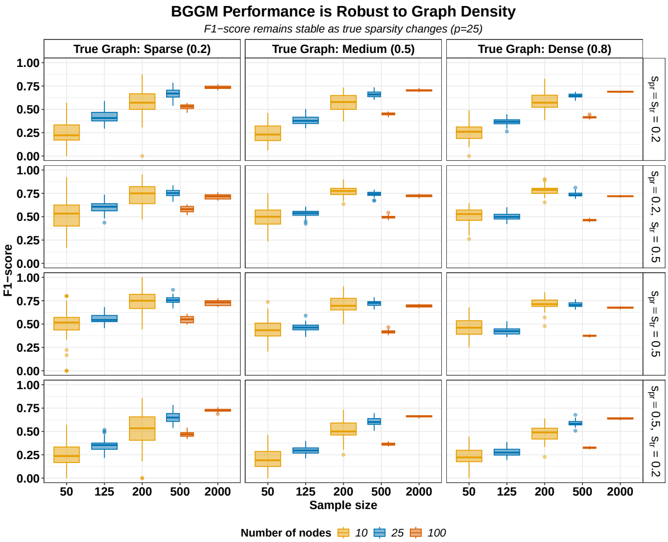
Results

The G-Wishart Approach: BDMCMC & RJ-WWA



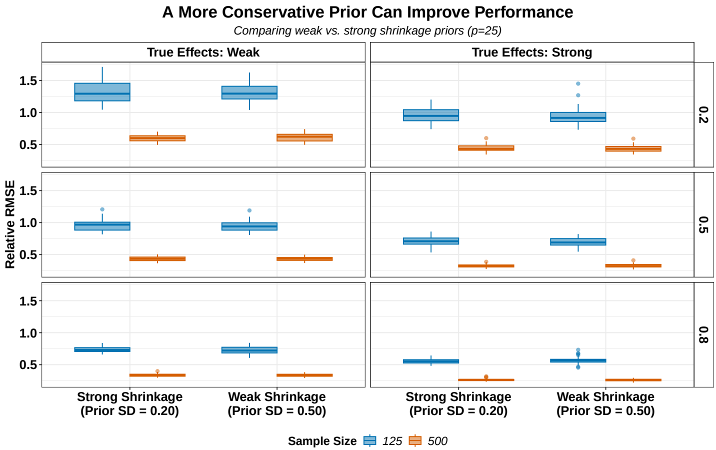
Results

The F-Matrix Approach: BGGM



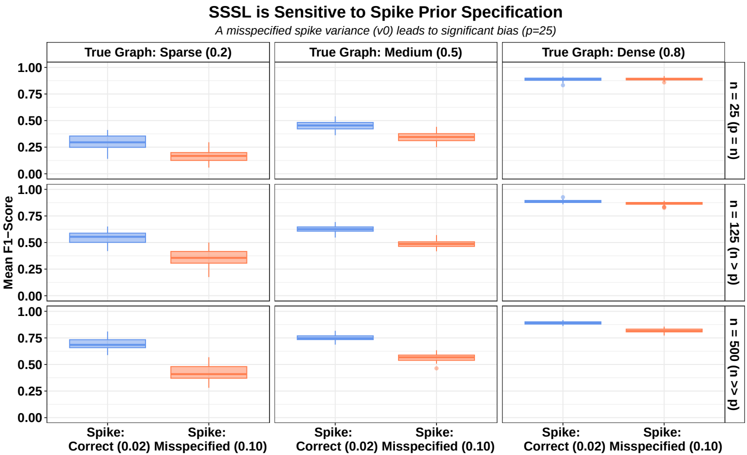
Results

The F-Matrix Approach: BGGM



Results

Stochastic Search Structure Learning



Conclusion

Conclusion

General results

- Data speaks: In high-dimensional regimes, all algorithms perform poorly. Given enough data ($n \geq 5p$), all of the tested methods are relatively robust to prior misspecification.
- There is no strong, systematic evidence significantly better or worse performance on random vs. small-world graphs of similar density.

Conclusion

Comparative results

- The G-Wishart methods, BDA and RJ-WWA, provided the best overall balance of finding true edges and avoiding false ones when priors were reasonably well-specified.
- BGGM is the most conservative and robust choice if you want to minimize false positives and are uncertain about graph density.
- SSSL is the most sensitive of the group, requiring fine tuning to perform well.

Limitations

- Conclusions are restricted to continuous, gaussian data.
- Only complete datasets were generated.
- Some priors of different algorithms are not directly comparable.

Thank you

- Email: `mmcarmo@ucdavis.edu`
- Website: `marwincarmo.github.io`
- Special thanks to my advisor Philippe Rast, unofficial mentors Mijke Rhemtulla and Jonathan Park, and labmates Abraham Iniguez and Teddy Chen.