

# Stochastic Block Model Prior with Ordering Constraints for Gaussian Graphical Models

**Teo Bucci, Filippo Cipriani,  
Filippo Pagella, Flavia Petruso,  
Andrea Puricelli, Giulio Venturini**  
Supervisor: Dr. Alessandro Colombi

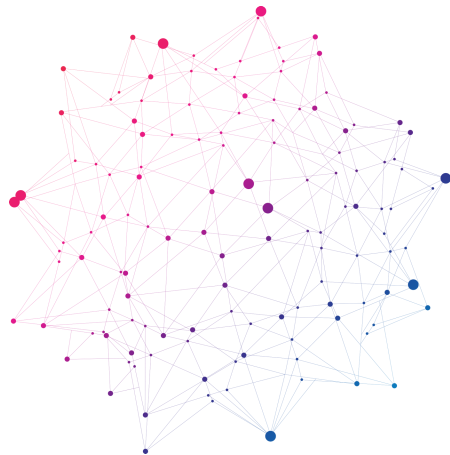
Bayesian Statistics  
MSc. Mathematical Engineering  
Politecnico di Milano

November 11, 2022



# CONTENTS

1. Background
  - 1.1 Gaussian Graphical Models
  - 1.2 Stochastic Block Models
  - 1.3 Changepoint Models
2. Project goals and next steps



BACKGROUND

## GAUSSIAN GRAPHICAL MODELS

Models for the **conditional dependence structure** among variables, represented through an undirected graph  $G = (V, E)$

$$\mathbf{y}_1, \dots, \mathbf{y}_n \mid \mathbf{K} \stackrel{\text{iid}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{K}^{-1})$$

Conditional independence described through a **map** between a **graph** and a family of multivariate **probability models**

$$Y_i \perp\!\!\!\perp Y_j \mid Y_{-(ij)} \iff (i, j) \notin E \iff k_{ij} = 0$$

Usual prior for  $\mathbf{K}$  conditionally to the graph is a G-Wishart

$$\mathbf{K} \mid G \sim \text{G-Wishart}(b, D)$$

$G$  is a r.v. in the space of undirected graphs with  $p$  nodes.

If we assume a possible **grouping** of the variables, we need a prior  $\pi(G)$  that induces a **block structure on its adjacency matrix**.

## STOCHASTIC BLOCK MODELS FOR THE PRIOR ON $G$

Given a random network, SBM **infer a node partition** based on similarity of connectivity patterns. Let

- $H$  be the fixed number of clusters
- $\mathbf{z} \in \mathbb{R}^p$  be the vector of group memberships,  $z_i \in \{1, \dots, H\}$

Then the model of the graph conditionally on  $\mathbf{z}$  is the following

$$P((i, j) \in E \mid \mathbf{z}, Q) = Q_{z_i z_j}, \text{ independent}$$

$$Q_{rs} \stackrel{\text{ind}}{\sim} \text{Beta}(a, b), \quad 1 \leq r \leq s \leq H$$

$$\mathbf{z} \sim f_{\mathbf{z}}(\mathbf{z})$$

Problems:

## STOCHASTIC BLOCK MODELS FOR THE PRIOR ON $G$

Given a random network, SBM **infer a node partition** based on similarity of connectivity patterns. Let

- $H$  be the fixed number of clusters
- $\mathbf{z} \in \mathbb{R}^p$  be the vector of group memberships,  $z_i \in \{1, \dots, H\}$

Then the model of the graph conditionally on  $\mathbf{z}$  is the following

$$P((i, j) \in E \mid \mathbf{z}, Q) = Q_{z_i z_j}, \text{ independent}$$

$$Q_{rs} \stackrel{\text{ind}}{\sim} \text{Beta}(a, b), \quad 1 \leq r \leq s \leq H$$

$$\mathbf{z} \sim f_z(\mathbf{z})$$

Problems:

- How to **identify the number of clusters**  $H$ ?

# STOCHASTIC BLOCK MODELS FOR THE PRIOR ON $G$

Given a random network, SBM **infer a node partition** based on similarity of connectivity patterns. Let

- $H$  be the fixed number of clusters
- $\mathbf{z} \in \mathbb{R}^p$  be the vector of group memberships,  $z_i \in \{1, \dots, H\}$

Then the model of the graph conditionally on  $\mathbf{z}$  is the following

$$P((i, j) \in E \mid \mathbf{z}, Q) = Q_{z_i z_j}, \text{ independent}$$

$$Q_{rs} \stackrel{\text{ind}}{\sim} \text{Beta}(a, b), \quad 1 \leq r \leq s \leq H$$

$$\mathbf{z} \sim f_{\mathbf{z}}(\mathbf{z})$$

Problems:

- How to **identify the number of clusters**  $H$ ?
- How to model **constraints on the ordering of the nodes** when generating the partition?

## CHANGEPOINT MODELS

Let  $\mathbf{y} = (y_1, \dots, y_n)$  be **ordered** observations, each depending on a parameter  $\vartheta_i$ . Consider a process where the underlying generating mechanism changes at time  $\tau_j$ ,  $j = 1, \dots, k + 1$ , where  $\tau_j, k$  are unknown. The likelihood can be modeled as

$$f(y_1, \dots, y_n \mid \vartheta_1, \vartheta_2, \dots, \vartheta_{k+1}, \mathbf{z}) = \prod_{j=1}^{k+1} \prod_{i=\tau_{j-1}+1}^{\tau_j} f(y_i \mid \vartheta_j), \quad k = \sum_{i=1}^{n-1} z_i$$

where  $\tau_j$  are called **changepoints** and  $z_i = 1$  if  $i$  is a changepoint, 0 otherwise. We are interested in **generating a random partition with ordering constraints**

$$\pi(\mathbf{z}, \boldsymbol{\vartheta} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\vartheta}, \mathbf{z}) \pi(\boldsymbol{\vartheta} \mid \mathbf{z}) \pi(\mathbf{z})$$



To define the distribution of  $\mathbf{z}$  we assign a probability law over the space of all **admissible** partitions.

To define the distribution of  $\mathbf{z}$  we assign a probability law over the space of all **admissible** partitions.

Let  $\rho_p = \{C_1, \dots, C_M\}$  be a partition of the  $p$  nodes.

$$\rho_p \text{ is admissible} \iff \forall x \in C_i, y \in C_j, i < j \text{ implies } x < y$$

To define the distribution of  $\mathbf{z}$  we assign a probability law over the space of all **admissible** partitions.

Let  $\rho_p = \{C_1, \dots, C_M\}$  be a partition of the  $p$  nodes.

$$\rho_p \text{ is admissible} \iff \forall x \in C_i, y \in C_j, i < j \text{ implies } x < y$$

Martínez and Mena 2014 assign

$$P(\rho_p = \{C_1, \dots, C_M\}) = p'(n_1, \dots, n_k) = \begin{cases} \binom{n}{n_1, \dots, n_k} \frac{1}{k!} p(n_1, \dots, n_k), & \rho_p \text{ admissible} \\ 0, & \rho_p \text{ not admissible} \end{cases}$$

where  $n_j = |C_j|$ ,  $p(n_1, \dots, n_k)$  any Exchangeable Partition Probability Function (EPPF).

To define the distribution of  $\mathbf{z}$  we assign a probability law over the space of all **admissible** partitions.

Let  $\rho_p = \{C_1, \dots, C_M\}$  be a partition of the  $p$  nodes.

$$\rho_p \text{ is admissible} \iff \forall x \in C_i, y \in C_j, i < j \text{ implies } x < y$$

Martínez and Mena 2014 assign

$$P(\rho_p = \{C_1, \dots, C_M\}) = p'(n_1, \dots, n_k) = \begin{cases} \binom{n}{n_1, \dots, n_k} \frac{1}{k!} p(n_1, \dots, n_k), & \rho_p \text{ admissible} \\ 0, & \rho_p \text{ not admissible} \end{cases}$$

where  $n_j = |C_j|$ ,  $p(n_1, \dots, n_k)$  any Exchangeable Partition Probability Function (EPPF).

The vector  $\mathbf{z}$  is derived by setting

$$z_i = m \iff i \in C_m, \quad i = 1, \dots, p, \quad m = 1, \dots, M.$$

# PROJECT GOALS AND NEXT STEPS

## GOAL OF THE PROJECT AND NEXT STEPS

**Goal:** propose a **new prior** that accounts for

- Ordering Constraint: taking advantage of the study of changepoint models
- Block Structure: using a stochastic block model prior

$$\mathbf{y}_1, \dots, \mathbf{y}_n \mid \mathbf{K} \stackrel{\text{iid}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{K}^{-1})$$

$$\mathbf{K} \mid G \sim \text{G-Wishart}(b, D)$$

$$P((i, j) \in E \mid \mathbf{z}, Q) = Q_{z_i z_j}, \text{ independent}$$

$$Q_{rs} \mid \mathbf{z} \stackrel{\text{ind}}{\sim} \text{Beta}(a, b), 1 \leq r \leq s \leq H$$

$$\rho_p \sim \mathcal{L}(\rho_p)$$

## GOAL OF THE PROJECT AND NEXT STEPS

**Goal:** propose a **new prior** that accounts for

- Ordering Constraint: taking advantage of the study of changepoint models
- Block Structure: using a stochastic block model prior

$$\mathbf{y}_1, \dots, \mathbf{y}_n \mid \mathbf{K} \stackrel{\text{iid}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{K}^{-1})$$

$$\mathbf{K} \mid G \sim \text{G-Wishart}(b, D)$$

$$P((i, j) \in E \mid \mathbf{z}, Q) = Q_{z_i z_j}, \text{ independent}$$






$$Q_{rs} \mid \mathbf{z} \stackrel{\text{ind}}{\sim} \text{Beta}(a, b), 1 \leq r \leq s \leq H$$

$$\rho_p \sim \mathcal{L}(\rho_p)$$

### Next steps:

- understand the nonparametric prior on ordered partitions
- implement the sampling strategy

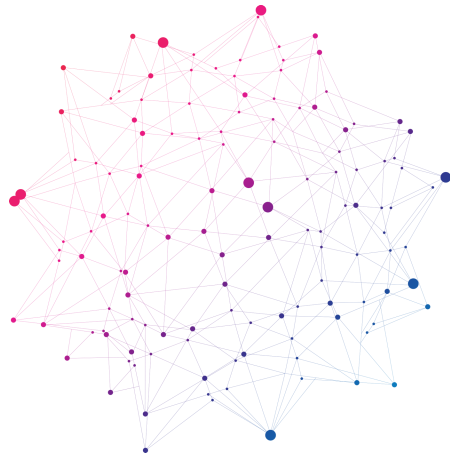
## MAIN REFERENCES

-  Benson, A. and N. Friel (2018). "Adaptive MCMC for Multiple Changepoint Analysis with Applications to Large Datasets". In: Electronic Journal of Statistics 12.2.
-  Colombi, A., R. Argiento, L. Paci and A. Pini (2022). "Learning block structured graphs in Gaussian graphical models". In: arXiv preprint arXiv:2206.14274.
-  Legramanti, S., T. Rigon, D. Durante and D. B. Dunson (2022). "Extended stochastic block models with application to criminal networks". In: The Annals of Applied Statistics 16.4, pp. 2369–2395.
-  Martínez, A. F. and R. H. Mena (2014). "On a Nonparametric Change Point Detection Model in Markovian Regimes". In: Bayesian Analysis 9.4.
-  Mohammadi, A. and E. C. Wit (2015). "Bayesian Structure Learning in Sparse Gaussian Graphical Models". In: Bayesian Analysis 10.1.



Thank you!

**Any questions?**



# STOCHASTIC BLOCK MODELS FOR THE PRIOR ON $G$

$$G_{ij} \mid \Theta, \mathbf{z} \stackrel{\text{iid}}{\sim} \text{Be}(\vartheta_{ij}), G \in \mathbb{R}^{p \times p}, \quad \vartheta_{ij} = \Theta_{z_i z_j}, \quad 1 \leq i < j \leq p$$

$$\Theta_{z_i z_j} \stackrel{\text{iid}}{\sim} \text{Beta}(\alpha, \beta)$$

$$\mathbf{z} \stackrel{\text{iid}}{\sim} \pi_{\mathbf{z}}(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^p$$

$\Theta$  can be marginalized out via beta-binomial conjugacy.



## CHOOSING THE NUMBER OF CLUSTERS $H$

$H$  is usually chosen using data driven methods **before** estimating the model. However

- uncertainty quantification for  $H$  at this stage is ignored
- the model does not account for new clusters (**prediction**)

Among others, Legramanti et al. 2022 proposed the **Extended Stochastic Block Models (ESBM)**, which use priors for  $\mathbf{z}$  that naturally allow to adaptively modify the number of groups  $H$ , e.g.,

- Finite Dirichlet
- Infinite Dirichlet
- Finite mixtures with a random number of components

**None of the ESBM models pose constraints on the ordering of the nodes when generating the partition, that may be of relevance in a real-life context (e.g., gene expression).**