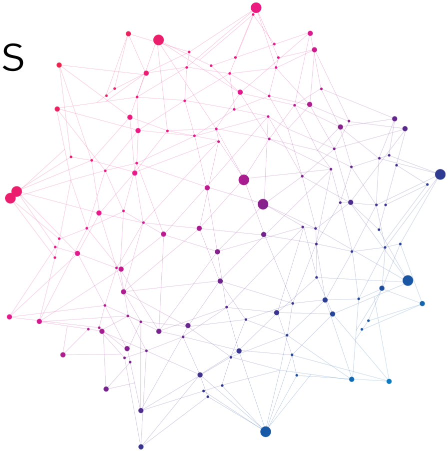
Stochastic Block Model Prior with Ordering Constraints for Gaussian Graphical Models

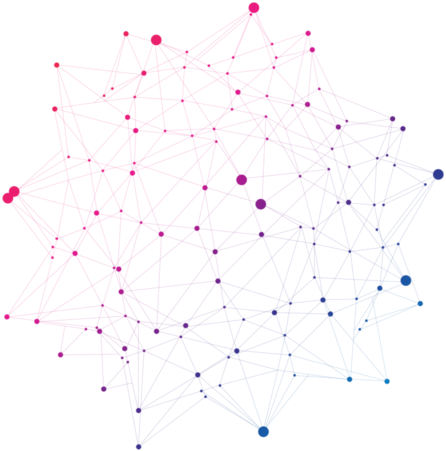
Teo Bucci, Filippo Cipriani,
Filippo Pagella, Flavia Petruso,
Andrea Puricelli, Giulio Venturini
Supervisor: Dr. Alessandro Colombi

Bayesian Statistics
MSc. Mathematical Engineering
Politecnico di Milano

January 09, 2023



CONTENTS



THE MODEL

THE MODEL

Goal: given a set of n data with p variables, simultaneously infer the conditional dependence structure of such variables and their clustering.

THE MODEL

Goal: given a set of n data with p variables, simultaneously infer the conditional dependence structure of such variables and their clustering.

Constraint: the partition must respect the original order of the variables.

THE MODEL

Goal: given a set of n data with p variables, simultaneously infer the conditional dependence structure of such variables and their clustering.

Constraint: the partition must respect the original order of the variables.

$$\begin{aligned} \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid \boldsymbol{K} &\overset{\text{iid}}{\sim} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{K}^{-1}) \\ \boldsymbol{K} \mid G &\sim \text{G-Wishart}(b, D) \\ P((i,j) \in E \mid \boldsymbol{z}, Q) &= Q_{z_i z_j}, \text{ independent} \\ Q_{rs} \mid \boldsymbol{z} &\overset{\text{ind}}{\sim} \text{Beta}(\alpha, \beta), 1 \leq r \leq s \leq M \\ \rho &\sim f_\rho(\rho) \end{aligned}$$

Notation for the partition: ρ vector of cardinalities, $\boldsymbol{z}$ vector of groups memberships.

The prior for the **partition** $f_\rho(\rho)$ is (5) from Martínez and Mena 2014.

Q is marginalized out.

Q is marginalized out.

The prior for the **graph** given the vector of group memberships is

$$P(G \mid \boldsymbol{z}) = \prod_{u=1}^{M} \prod_{v=u}^{M} \frac{B(\alpha + S_{uv}, \beta + S_{uv}^{\star})}{B(\alpha, \beta)}$$

where

- S_{uv} is the sum of the edges between group u and v.
- $S_{uv}^{\star}$ is the sum of the "non-edges", namely $S_{uv}^{\star} = T_{uv} - S_{uv}$ and T_{uv} is the total number of possible edges.

THE SAMPLING STRATEGY

BLOCK GIBBS SAMPLER

Conditional distributions for our model:

Graph and Precision	$P(\boldsymbol{K}, G \mid \boldsymbol{Y}, \boldsymbol{z}) \propto P(\boldsymbol{Y} \mid \boldsymbol{K})P(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})$
Random Partition	$P(\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{K}, G) \propto P(\boldsymbol{Y} \mid \boldsymbol{K})P(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})P(\boldsymbol{z}) \propto P(G \mid \boldsymbol{z})P(\boldsymbol{z})$

BLOCK GIBBS SAMPLER

Conditional distributions for our model:

Graph and Precision	$P(\boldsymbol{K}, G \mid \boldsymbol{Y}, \boldsymbol{z}) \propto P(\boldsymbol{Y} \mid \boldsymbol{K})P(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})$
Random Partition	$P(\boldsymbol{z} \mid \boldsymbol{Y}, \boldsymbol{K}, G) \propto P(\boldsymbol{Y} \mid \boldsymbol{K})P(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})P(\boldsymbol{z}) \propto P(G \mid \boldsymbol{z})P(\boldsymbol{z})$

We implement a Block Gibbs-Sampler strategy:

1. **Sampling Graph and Precision Matrix**
   G and $\boldsymbol{K}$ - given $\boldsymbol{z}$ - are sampled using a modified version of a Birth-and-Death chain (Mohammadi and Wit 2015), changing one link at a time.
2. **Sampling the Random Partition**
   Conditionally on G, we can sample $\boldsymbol{z}$ through an adaptive split and merge sampler.

BIRTH AND DEATH ALGORITHM FOR UPDATING THE GRAPH

`BDGraph` is an algorithm that follows a Birth-and-Death approach to decide whether to **add** a new edge to the graph or **delete** an already existing one.

BIRTH AND DEATH ALGORITHM FOR UPDATING THE GRAPH

`BDGraph` is an algorithm that follows a Birth-and-Death approach to decide whether to **add** a new edge to the graph or **delete** an already existing one.

	Target distribution	B/D rates
Before	$P(G, \boldsymbol{K} \mid \boldsymbol{Y}) \propto \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{K})\mathcal{L}(\boldsymbol{K} \mid G)P(G)$	$\frac{P(G')}{P(G)}$
After	$P(G, \boldsymbol{K} \mid \boldsymbol{Y}, \boldsymbol{z}) \propto \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{K})\mathcal{L}(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})$	$\frac{P(G' \mid \boldsymbol{z})}{P(G \mid \boldsymbol{z})}$

where $G' = G^{\pm e}$ and e is an edge.

BIRTH AND DEATH ALGORITHM FOR UPDATING THE GRAPH

`BDGraph` is an algorithm that follows a Birth-and-Death approach to decide whether to **add** a new edge to the graph or **delete** an already existing one.

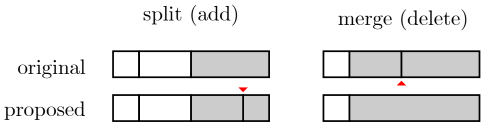
	Target distribution	B/D rates
Before	$P(G, \boldsymbol{K} \mid \boldsymbol{Y}) \propto \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{K})\mathcal{L}(\boldsymbol{K} \mid G)P(G)$	$\frac{P(G')}{P(G)}$
After	$P(G, \boldsymbol{K} \mid \boldsymbol{Y}, \boldsymbol{z}) \propto \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{K})\mathcal{L}(\boldsymbol{K} \mid G)P(G \mid \boldsymbol{z})$	$\frac{P(G' \mid \boldsymbol{z})}{P(G \mid \boldsymbol{z})}$

where $G' = G^{\pm e}$ and e is an edge.

$$\text{Birth rate} \propto \frac{P(G^{+e} \mid \boldsymbol{z})}{P(G \mid \boldsymbol{z})} = \frac{S_{uv} + \alpha}{S^{\star}_{uv} + \beta} \qquad \text{Death rate} \propto \frac{P(G^{-e} \mid \boldsymbol{z})}{P(G \mid \boldsymbol{z})} = \frac{S^{\star}_{uv} + \beta}{S_{uv} + \alpha}$$

GENERAL STEPS FOR UPDATING THE PARTITION

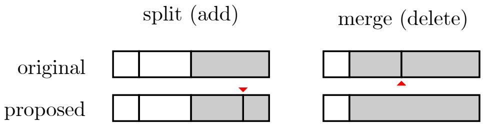
We perform an adaptive **split and merge**.



1. With probability α_{split}, usually 0.5, choose an **split move**, otherwise a **merge move**. Unless we are forced by extreme cases.
   1.1 Propose a new partition by splitting one group into two or merging two adjacent.
   1.2 Accept or reject using Metropolis Hastings. The target is: $f(\rho \mid G) \approx f_G(G \mid \rho) f_\rho(\rho)$

GENERAL STEPS FOR UPDATING THE PARTITION

We perform an adaptive **split and merge**.

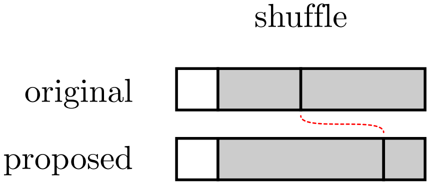


1. With probability α_{split}, usually 0.5, choose an **split move**, otherwise a **merge move**. Unless we are forced by extreme cases.
   1.1 Propose a new partition by splitting one group into two or merging two adjacent.
   1.2 Accept or reject using Metropolis Hastings. The target is: $f(\rho \mid G) \approx f_G(G \mid \rho) f_\rho(\rho)$

$$\alpha_{\text{accept}} = \min\left\{1, \overbrace{\underbrace{\frac{f_G(G \mid \rho')}{f_G(G \mid \rho)}}_{\substack{\text{graph}\\ \text{ratio}}} \underbrace{\frac{f_\rho(\rho')}{f_\rho(\rho)}}_{\substack{\text{partition}\\ \text{ratio}}}}^{\text{target ratio}} \underbrace{\frac{Q(\rho', \rho)}{Q(\rho, \rho')}}_{\substack{\text{proposal}\\ \text{ratio}}}\right\}$$

SHUFFLE MOVE

2. To improve the mixing of the chain we also perform a **shuffle move**.
   2.1 Propose a new partition by moving some nodes from a group to an adjacent one.
   2.2 Accept or reject using Metropolis Hastings.



PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

Suppose a split move.

PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

Suppose a split move.

At each iteration, **draw** node i with probability proportional to $a_i^{(t)}$.

PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

Suppose a split move.

At each iteration, **draw** node i with probability proportional to $a_i^{(t)}$.

Denoting by $a^{\star} = \sum_{j \in \{\text{admissible split nodes}\}} a_j^{(t)}$ and $d^{\star} = \sum_{j \in \{\text{admissible merge nodes}\}} d_j^{(t)}$

PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

Suppose a split move.

At each iteration, **draw** node i with probability proportional to $a_i^{(t)}$.

Denoting by $a^\star = \sum_{j\in\{\text{admissible split nodes}\}} a_j^{(t)}$ and $d^\star = \sum_{j\in\{\text{admissible merge nodes}\}} d_j^{(t)}$

$$\frac{Q(\rho',\rho)}{Q(\rho,\rho')} = \frac{P(\text{choose merge})}{P(\text{choose split})} \cdot \frac{P(\text{merge at node } i)}{P(\text{split at node } i)} = \frac{1-\alpha_{\text{split}}}{\alpha_{\text{split}}} \cdot \frac{\frac{d_i^{(t)}}{d^\star + d_i^{(t)}}}{\frac{a_i^{(t)}}{a^\star}}$$

PROPOSAL RATIO

We introduce $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, two $(p-1)$-dimensional vectors of **weights** to choose the node where to perform the split or the merge. They are unnormalized discrete densities.

Suppose a split move.

At each iteration, **draw** node i with probability proportional to $a_i^{(t)}$.

Denoting by $a^\star = \sum_{j \in \{\text{admissible split nodes}\}} a_j^{(t)}$ and $d^\star = \sum_{j \in \{\text{admissible merge nodes}\}} d_j^{(t)}$

$$\frac{Q(\rho', \rho)}{Q(\rho, \rho')} = \frac{P(\text{choose merge})}{P(\text{choose split})} \cdot \frac{P(\text{merge at node } i)}{P(\text{split at node } i)} = \frac{1-\alpha_{\text{split}}}{\alpha_{\text{split}}} \cdot \frac{\frac{d_i^{(t)}}{d^\star + d_i^{(t)}}}{\frac{a_i^{(t)}}{a^\star}}$$

The extreme cases "every node belonging to the same group" and "every node has its own group" are dealt with separately.

ADAPTIVE STEP

The two weights vectors $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$ are updated at each iteration t as in Benson and Friel 2018 using the following adaptation scheme.

ADAPTIVE STEP

The two weights vectors $\mathbf{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$ are updated at each iteration t as in Benson and Friel 2018 using the following adaptation scheme.

- If a **split** move at node i has been accepted, then update:

$$\log(a_i^{(t+1)}) = \log(a_i^{(t)}) + \frac{h}{t/p}(\alpha_{\text{split}} - \alpha_{\text{target}}).$$

- If a **merge** move at node i has been accepted, then update:

$$\log(d_i^{(t+1)}) = \log(d_i^{(t)}) + \frac{h}{t/p}(\alpha_{\text{merge}} - \alpha_{\text{target}}).$$

Where $h > 0$ is the initial adaptation, t/p are the iterations (t) per number of nodes (p), α_{target} is the target MH acceptance rate and $\alpha_{\text{merge}} = 1 - \alpha_{\text{split}}$.

NEXT STEPS

NEXT STEPS

The first draft of the code is completed.

Next things to do:

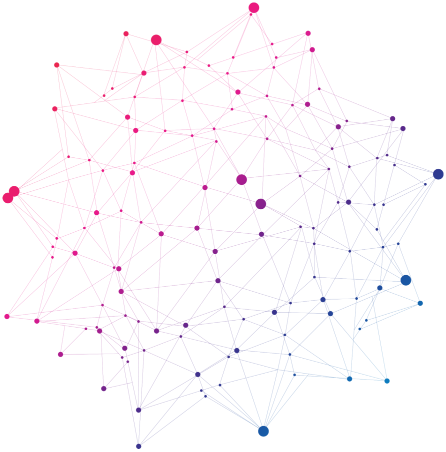
- Debugging the code.
- Run simulations.
- Perform posterior analysis.

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?



PARTITION RATIO

As a prior, we use an EPPF induced by the **two-parameter Poisson-Dirichlet** process (Pitman-Yor process) from Martínez and Mena 2014.

Let M and p be the number of groups and nodes, respectively, and $n_j, j = 1, \dots, M$ the cardinalities of the groups, ϑ and σ are parameters.

$$P(\rho = \{n_1, \dots, n_M\}) = \begin{cases} \frac{p!}{M!} \frac{\prod_{i=1}^{M-1} (\vartheta + i\sigma)}{(\vartheta+1)_{(p-1)\uparrow}} \prod_{j=1}^{M} \frac{(1-\sigma)_{(n_j-1)\uparrow}}{n_{j\uparrow}}, & \rho \text{ admissible} \\ 0, & \rho \text{ not admissible.} \end{cases}$$

Hence, in the split case, after simplifying common factors, the **partition ratio** is:

$$\frac{f_\rho(\rho')}{f_\rho(\rho)} = \frac{1}{M}(\vartheta + M\sigma)\frac{(1-\sigma)_{(n'_s-1)\uparrow}(1-\sigma)_{(n'_s+1)\uparrow}}{(1-\sigma)_{(n_s-1)\uparrow}} \frac{n_s!}{n'_s!n'_{s+1}!}$$

GRAPH RATIO

Suppose a split move.

The **graph ratio**, after simplifying common factors, is:

$$\frac{f_G(G \mid \rho')}{f_G(G \mid \rho)} = \left(\frac{1}{B(\alpha, \beta)}\right)^{M+1} \times$$

$$\frac{\prod_{l=1}^{S-1} f_B(C'_l, C'_S) f_B(C'_l, C'_{S+1}) \cdot \prod_{m=S+2}^{M+1} f_B(C'_S, C'_m) f_B(C'_{S+1}, C'_m) \cdot f_B(C'_S, C'_{S+1}) f_B(C'_S, C'_S) f_B(C'_{S+1}, C'_{S+1})}{\prod_{l=1}^{S-1} f_B(C_l, C_S) \prod_{m=S+1}^{M} f_B(C_S, C_m) \cdot f_B(C_S, C_S)}$$

where

$$f_B(C_u, C_v) = B(\alpha + S_{uv}, \beta + S^{\star}_{uv})$$