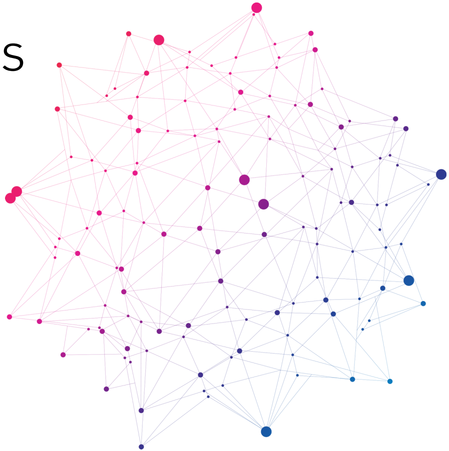
Stochastic Block Model Prior with Ordering Constraints for Gaussian Graphical Models

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

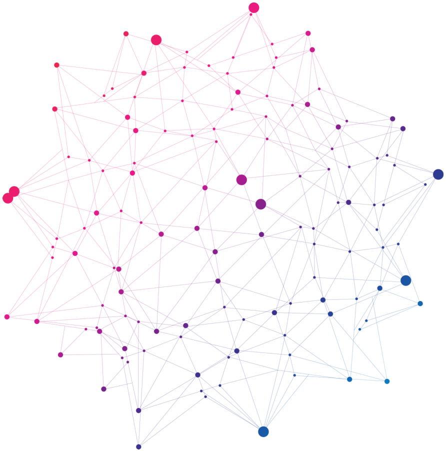
Bayesian Statistics
MSc. Mathematical Engineering
Politecnico di Milano

February 14, 2023



CONTENTS

1. The model
2. Quick review of the sampling strategy
3. Simulations and Posterior analysis
4. Conclusions
5. Extra



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.

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

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

The prior for the **partition** $P(\boldsymbol{z})$ is (5) from Martínez and Mena (2014).

To improve the mixing of the chain, the two parameters of $P(\boldsymbol{z})$, namely ϑ and σ, are updated as in the original paper.

Q is marginalized out.

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

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

where

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

QUICK REVIEW OF THE SAMPLING STRATEGY

BLOCK GIBBS SAMPLER

Conditional distributions for our model:

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

We implement a block Gibbs sampling strategy:

1. **Sampling Graph and Precision Matrix**
   $\boldsymbol{G}$ 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. The modified Birth and Death rates take into account the dependency on the random partition.
2. **Sampling the Random Partition**
   Conditionally on $\boldsymbol{G}$, we can sample $\boldsymbol{z}$ through an adaptive split and merge sampler.

SIMULATIONS AND POSTERIOR ANALYSIS

SIMULATIONS STRUCTURE

The Beta was reparametrized with mean and variance, with mean set to the graph density.

We ran a total of 41 simulations, varying different hyperparameters. Simulations divided into groups, depending on the parameter tuned

n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL
500	25	BD	22111996	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187
500	25	BD	31051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187
500	25	BD	27051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187

POSTERIOR ANALYSIS

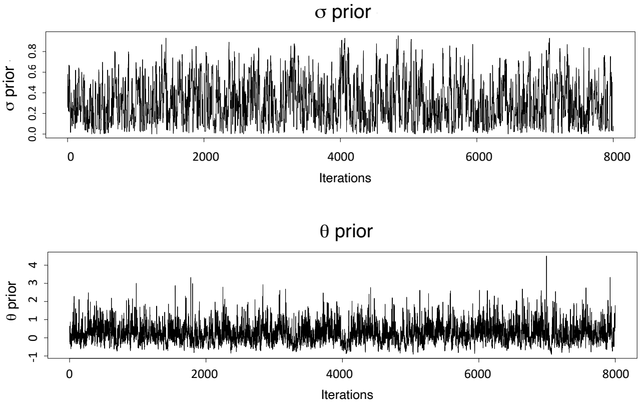
ρ_{est}	`accept`	VI	RI	KL

1. $\boldsymbol{\rho}$ estimated by solving a minimization problem on the space of visited partitions using **Variation of Information** (**VI**) loss function
2. Mean **acceptance rate**
3. **Rand index** (**RI**) used to measure the similarity between partitions
4. **Kullback-Leibler** (**KL**) distance to compare generating and estimated precision matrices

The adjacency matrix estimated by taking the **p-links** matrix (probability that a link is included in the graph) and choosing a threshold value s using BFDR.

THETA AND SIGMA TRACEPLOTS

The traceplots of the prior parameters ϑ and σ are similar throughout all the simulations.



GROUP 1: VARYING DATA GENERATING SEED

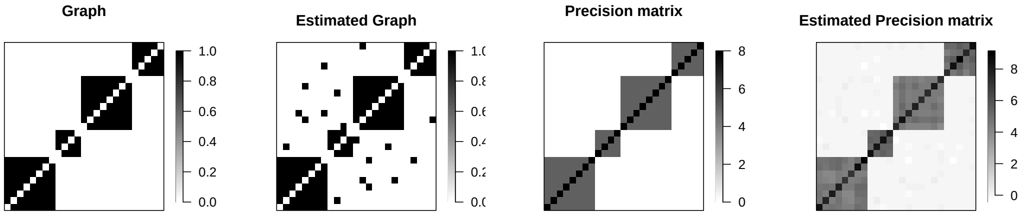
	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
01	500	25	BD	22111996	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21400 mins
02	500	25	BD	31051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21800 mins
03	500	25	BD	27051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21700 mins
04	500	25	BD	29061999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21700 mins
05	500	25	BD	12091997	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21600 mins
06	500	25	BD	27091999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21500 mins
07	500	25	BD	27121996	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21700 mins

Results are consistent for every seed:

- Estimated partition coincides with the generating partition
- KL distance converges to 0.187
- Estimated adjacency matrix is close to the generating adjacency matrix

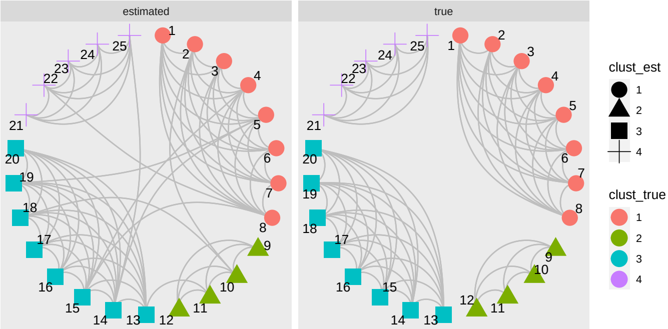
GROUP 1: VARYING DATA GENERATING SEED

sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
02	500	25	BD	31051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21800 mins



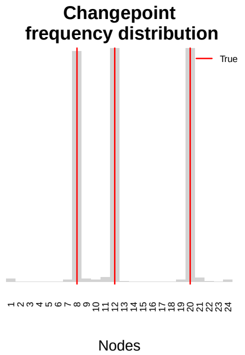
GROUP 1: VARYING DATA GENERATING SEED

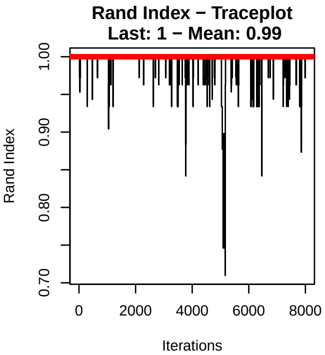
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
02	500	25	BD	31051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21800 mins

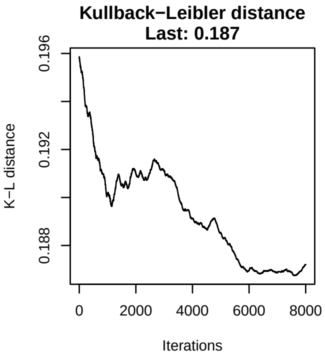


GROUP 1: VARYING DATA GENERATING SEED

sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
02	500	25	BD	31051999	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.21800 mins





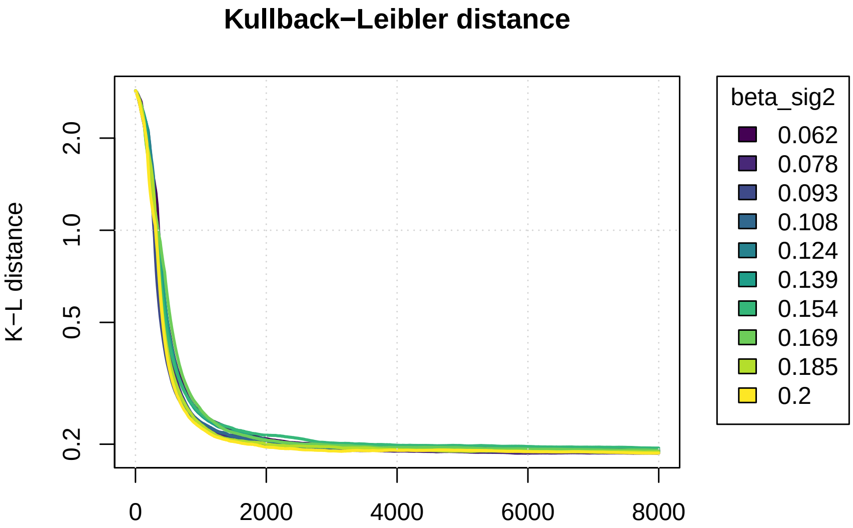


GROUP 2: CHANGING THE VARIANCE OF THE BETA PRIOR

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
08	500	25	BD	27121996	25	0.062	8,4,8,5	8,4,8,5	0.066	0.034	1.000	0.191	1.22400 mins
09	500	25	BD	27121996	25	0.078	8,4,8,5	8,4,8,5	0.080	0.043	1.000	0.186	1.20700 mins
10	500	25	BD	27121996	25	0.093	8,4,8,5	8,4,8,5	0.075	0.039	1.000	0.188	1.20600 mins
11	500	25	BD	27121996	25	0.108	8,4,8,5	8,4,8,5	0.079	0.049	1.000	0.188	1.20800 mins
12	500	25	BD	27121996	25	0.124	8,4,8,5	8,4,8,5	0.067	0.040	1.000	0.188	1.20100 mins
13	500	25	BD	27121996	25	0.139	8,4,8,5	8,4,8,5	0.060	0.035	1.000	0.188	1.20200 mins
14	500	25	BD	27121996	25	0.154	8,4,8,5	8,4,8,5	0.056	0.030	1.000	0.193	1.20000 mins
15	500	25	BD	27121996	25	0.169	8,4,8,5	8,4,8,5	0.047	0.024	1.000	0.191	1.20100 mins
16	500	25	BD	27121996	25	0.185	8,4,8,5	8,4,8,5	0.034	0.022	1.000	0.189	1.20000 mins
17	500	25	BD	27121996	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.16200 mins

Changing the **variance of the Beta prior** does not yield significant changes in posterior analysis.

GROUP 2: CHANGING THE VARIANCE OF THE BETA PRIOR



GROUP 3: VARYING INITIAL PARTITION

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
18	500	25	BD	27121996	25	0.2	8,4,8,5	8,4,8,5	0.019	0.017	1.000	0.187	1.16300 mins
19	500	25	BD	27121996	singletons	0.2	8,4,8,5	8,4,8,5	0.020	0.015	1.000	0.189	1.24800 mins

Starting from different initial partitions (i.e. one single partition or all singletons) does not affect the final result.

GROUP 4: VARYING GENERATING PARTITION'S CLUSTER NUMEROSITIES

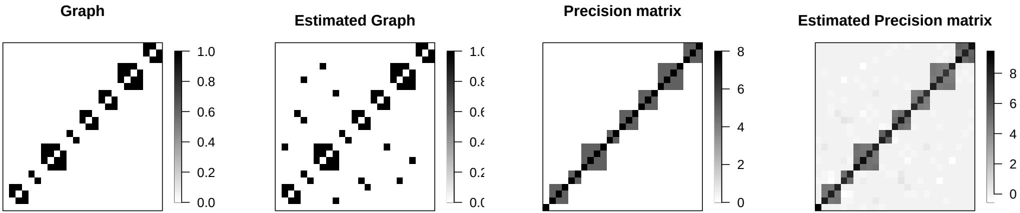
	Data								Analysis					
`sim_id`	n	p	`data_gen`	`seed`	$\boldsymbol{\rho}_0$	`beta_sig2`	$\boldsymbol{\rho}_{\text{true}}$		$\boldsymbol{\rho}_{\text{est}}$	`accept`	VI	RI	KL	`time`
20	500	25	BD	27121996	25	0.2	8,4,8,5		8,4,8,5	0.019	0.017	1.000	0.187	1.16000 mins
21	500	25	BD	27121996	25	0.2	1,10,2,9,3		1,10,2,9,3	0.072	0.109	1.000	0.181	1.25500 mins
22	500	25	BD	27121996	25	0.02	1,3,2,4,2,3,3,4,3		18,7	0.397	1.020	0.129	0.126	1.10700 mins
23	500	25	BD	27121996	25	0.2	12,13		12,13	0.129	0.073	1.000	0.255	1.36700 mins

Changing the generating partition's cluster numerosities does not generally influence the result. The only exception comes with small and numerous clusters, as in simulation 22.

Let's have a closer look.

GROUP 4: VARYING GENERATING PARTITION'S CLUSTER NUMEROSITIES

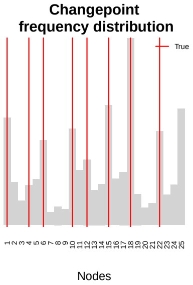
	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
22	500	25	BD	27121996	25	0.02	1,3,2,4,2,3,3,4,3	18,7	0.397	1.020	0.129	0.126	1.10700 mins

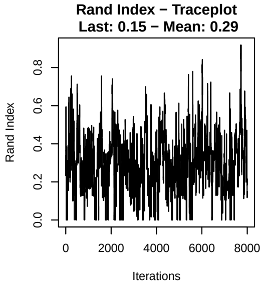


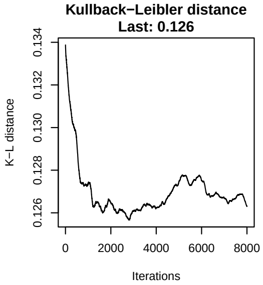
GROUP 4: VARYING GENERATING PARTITION'S CLUSTER NUMEROSITIES

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
22	500	25	BD	27121996	25	0.02	1,3,2,4,2,3,3,4,3	18,7	0.397	1.020	0.129	0.126	1.10700 mins

The problem partially lies in the criteria for selecting the partition.


Changepoint frequency distribution


Rand Index – Traceplot
Last: 0.15 – Mean: 0.29

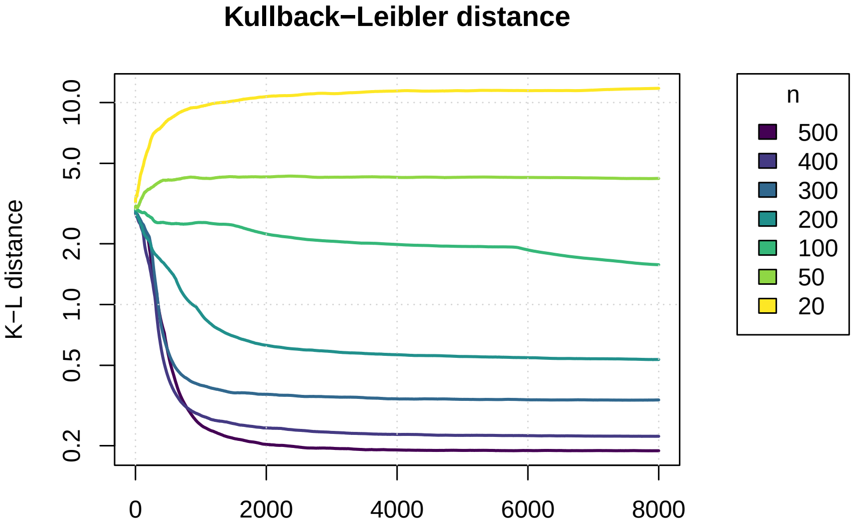

Kullback–Leibler distance
Last: 0.126

GROUP 5: VARYING NUMBER OF OBSERVATIONS

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
24	500	25	BD	27121996	25	0.1	8,4,8,5	8,4,8,5	0.082	0.046	1.000	0.187	1.17300 mins
25	400	25	BD	27121996	25	0.1	8,4,8,5	8,4,8,5	0.072	0.036	1.000	0.221	1.15900 mins
26	300	25	BD	27121996	25	0.1	8,4,8,5	8,4,8,5	0.070	0.041	1.000	0.336	1.16800 mins
27	200	25	BD	27121996	25	0.1	8,4,8,5	8,4,8,5	0.074	0.045	1.000	0.532	1.15400 mins
28	100	25	BD	27121996	25	0.1	8,4,8,5	8,4,8,5	0.111	0.241	1.000	1.487	1.11800 mins
29	50	25	BD	27121996	25	0.1	8,4,8,5	20,5	0.236	0.539	0.273	4.212	1.09000 mins
30	20	25	BD	27121996	25	0.1	8,4,8,5	25	0.280	0.241	0.000	11.854	1.07600 mins

As n diminishes and becomes comparable to p, we notice that the KL distance plateaus on progressively higher values, and VI and RI behave accordingly. The acceptance rates rises in more difficult situations.

GROUP 5: VARYING NUMBER OF OBSERVATIONS

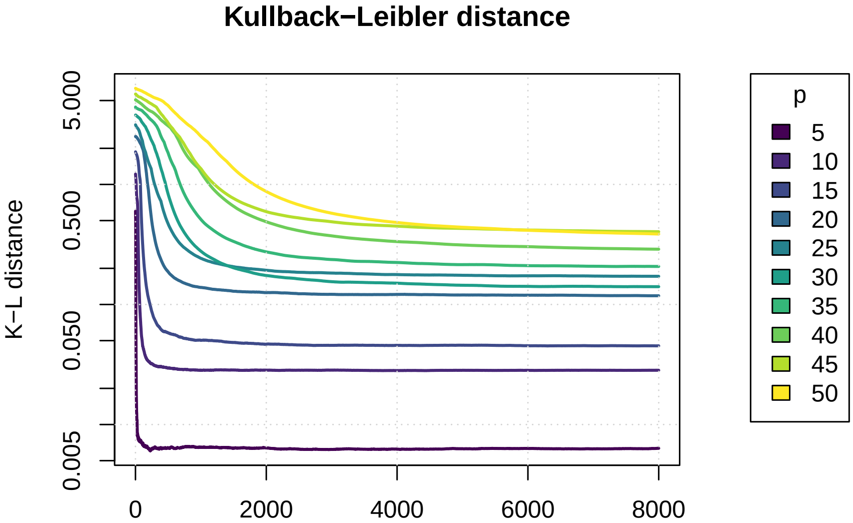


GROUP 6: VARYING NUMBER OF NODES

	Data							Analysis					
`sim_id`	n	p	`data_gen`	`seed`	ρ_0	`beta_sig2`	ρ_{true}	ρ_{est}	`accept`	VI	RI	KL	`time`
31	500	5	BD	27121996	5	0.0625	3,2	3,2	0.543	0.752	1.000	0.006	0.79960 mins
32	500	10	BD	27121996	10	0.0625	5,5	5,5	0.188	0.115	1.000	0.028	0.85055 mins
33	500	15	BD	27121996	15	0.0625	5,5,5	5,5,5	0.129	0.087	1.000	0.045	1.09400 mins
34	500	20	BD	27121996	20	0.0625	5,5,5,5	5,5,5,5	0.103	0.070	1.000	0.118	1.20100 mins
35	500	25	BD	27121996	25	0.0625	5,5,5,5,5	5,5,5,5,5	0.104	0.074	1.000	0.170	1.33600 mins
36	500	30	BD	27121996	30	0.0625	5,5,5,5,5,5	5,5,5,5,5,5	0.080	0.061	1.000	0.140	1.49600 mins
37	500	35	BD	27121996	35	0.0625	5,5,5,5,5,5,5	5,5,5,5,5,5,5	0.066	0.149	1.000	0.205	1.71700 mins
38	500	40	BD	27121996	40	0.0625	5,5,5,5,5,5,5,5	5,5,5,5,5,5,5,5	0.061	0.384	1.000	0.282	1.99900 mins
39	500	45	BD	27121996	45	0.0625	5,5,5,5,5,5,5,5,5	5,5,5,5,5,10,10	0.071	0.761	0.756	0.396	2.45600 mins
40	500	50	BD	27121996	50	0.0625	5,5,5,5,5,5,5,5,5,5	25,5,5,5,5,5	0.059	0.853	0.364	0.367	3.05700 mins

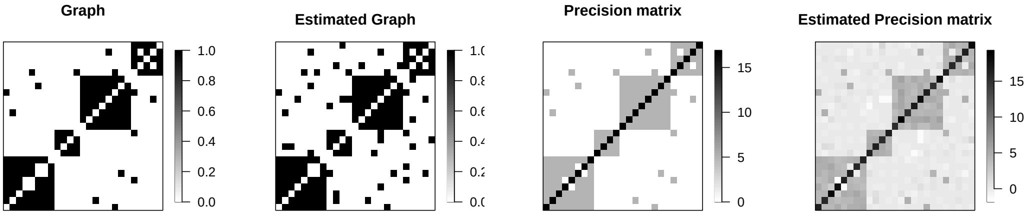
Increasing p yields consequences comparable to diminishing n.

GROUP 6: VARYING NUMBER OF NODES



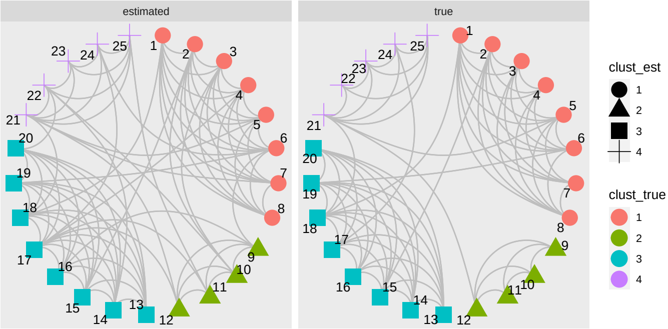
GROUP 7: USING NOISED BLOCK STRUCTURE AS DATA GENERATOR

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
41	500	25	B	27121996	25	0.2	8,4,8,5	8,4,8,5	0.013	0.104	1.000	0.176	1.03100 mins



GROUP 7: USING NOISED BLOCK STRUCTURE AS DATA GENERATOR

	Data							Analysis					
sim_id	n	p	data_gen	seed	ρ_0	beta_sig2	ρ_{true}	ρ_{est}	accept	VI	RI	KL	time
41	500	25	B	27121996	25	0.2	8,4,8,5	8,4,8,5	0.013	0.104	1.000	0.176	1.03100 mins



CONCLUSIONS

CONCLUSIONS AND FUTURE DIRECTIONS

Final considerations

- Effective choice of the prior distribution for the partition
- Good performance (Rand index, VI, KL), across a wide range of parameter perturbations

Limitations and future directions

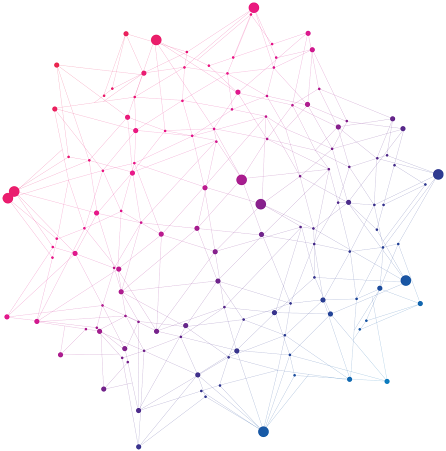
- Refine hyperparameter tuning
- Further analysis of convergence speed in the first iterations and the computational complexity
- Apply the model to real data

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?



EXTRA

PRIOR FOR THE PARTITION

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 with $j = 1, \dots, M$ the cardinalities of the groups, ϑ and σ are parameters.

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

σ is assigned a prior Beta(a, b) while ϑ is assigned a prior ShiftedGamma$(c, d, -\sigma)$.

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.

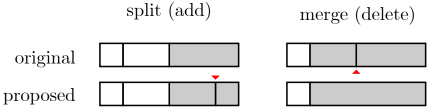
	Target distribution	B/D rates
Before	$P(\boldsymbol{G}, \boldsymbol{K} \mid \boldsymbol{Y}) \propto P(\boldsymbol{Y} \mid \boldsymbol{K}) P(\boldsymbol{K} \mid \boldsymbol{G}) P(\boldsymbol{G})$	$\dfrac{P(\boldsymbol{G}')}{P(\boldsymbol{G})}$
After	$P(\boldsymbol{G}, \boldsymbol{K} \mid \boldsymbol{Y},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

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

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

GENERAL STEPS FOR UPDATING THE PARTITION

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

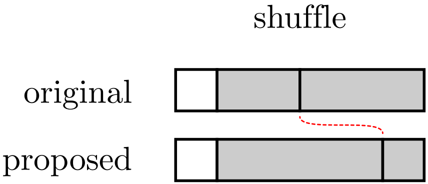


1. With probability α_{split}, usually 0.5, choose an **split move**, otherwise a **merge move**. Unless we are forced by extreme cases.
   1.1 Propose a new partition by splitting one group into two or merging two adjacent, using two vectors of **weights**, $\boldsymbol{a}^{(t)}$ and $\boldsymbol{d}^{(t)}$, to choose where to perform the move.
   1.2 Accept or reject using Metropolis Hastings. The target is: $f(\boldsymbol{z} \mid \boldsymbol{G}) \approx P(\boldsymbol{G} \mid \boldsymbol{z})P(\boldsymbol{z})$

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

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.



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

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 $\alpha_{\text{merge}} = 1 - \alpha_{\text{split}}$.

PERFORMANCE INDEXES: KULLBACK-LEIBLER

Suppose that we have two multivariate normal distributions, with means μ_0, μ_1 and with (non-singular) covariance matrices Σ_0, Σ_1. If the two distributions have the same dimension, k, then the relative entropy between the distributions is as follows:

$$D_{\mathrm{KL}}\left(\mathcal{N}_0 \| \mathcal{N}_1\right) = \frac{1}{2}\left(\operatorname{tr}\left(\Sigma_1^{-1}\Sigma_0\right) - k + \left(\mu_1 - \mu_0\right)^{\top} \Sigma_1^{-1}\left(\mu_1 - \mu_0\right) + \ln\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right)\right)$$

PERFORMANCE INDEXES: RAND INDEX

Given a set of n elements $S = \{o_1, \ldots, o_n\}$ and two partitions of S to compare, $X = \{X_1, \ldots, X_r\}$, a partition of S into r subsets, and $Y = \{Y_1, \ldots, Y_s\}$, a partition of S into s subsets, define the following:

- a, the number of pairs of elements in S that are in the same subset in X and in the same subset in Y
- b, the number of pairs of elements in S that are in different subsets in X and in different subsets in Y
- c, the number of pairs of elements in S that are in the same subset in X and in different subsets in Y
- d, the number of pairs of elements in S that are in different subsets in X and in the same subset in Y

The Rand index, R, is:

$$R = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}}$$

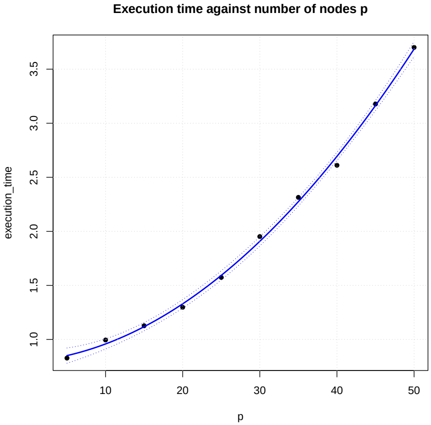
PERFORMANCE INDEXES: BFDR

Given the estimated p-links matrix $\hat{p}_{jk}$ the Bayesian False Discovery rate is

$$\mathrm{BFDR}(s) = \frac{\sum_{j<k} \left(1 - \hat{p}_{jk}\right) \mathbb{1}_{(\hat{p}_{jk} \geq s)}}{\sum_{j<k} \mathbb{1}_{(\hat{p}_{jk} \geq s)}}$$

and the threshold s is selected so that BFDR is below 0.05.

PRELIMINARY RESULTS ON COMPUTATIONAL COMPLEXITY



The execution time appears to scale with p^2.