

**Piotr Graczyk (Angers, France)**

**ERASMUS LECTURE**

Politechnika Wroclawska, February 2021

**Part 1. Graphical and Colored  
Graphical Models in statistics**

**Part 2. Model Selection of  
colored Gaussian models**

*Piotr Graczyk(Angers), Hideyuki Ishi(Osaka),  
Bartosz Kołodziejek(Warsaw),  
Hélène Massam(Toronto)*

# **GRAPHICAL MODELS IN STATISTICS:**

**coding conditional independence by a graph**

(mathematical and statistical theory started  
by S. Lauritzen and collaborators, 1980's)

## Simpson paradox

A university has 48 000 students

Half boys(24 000), half girls(24 000)

At the final exams: 10 000 boys and 14 000 girls fail

Feminist organizations threaten to close the university,  
girl students want to lynch the president!

However, the president of the university proves that the **results**  $R$  of the exams are **conditionally independent** of the **sex**  $S$  of a student, **knowing the department**  $D$  (notation  $R \perp\!\!\!\perp S|D$ )

3 departments

**A**(literature, history, languages),

**B**(law),

**C**(sciences)

16 000 students each

	<b>A</b>	Succ.	Fail	<b>B</b>	Succ.	Fail	<b>C</b>	Succ.	Fail
Girls	3	3	9	4	4	4	3	3	1
Boys	1	1	3	4	4	4	9	9	3

Actually  $R \perp\!\!\!\perp S \mid D = d$  for  $d = \mathbf{A, B, C}$

**The notion of the conditional independence is necessary to understand the Simpson paradox.**

## Graphical coding of conditional independence

Let  $\mathcal{G}$  be a graph with vertices  $v_i$ . If 2 vertices  $v_i, v_j$  are connected by an (undirected) edge, we write  $\boxed{v_i \sim v_j}$ .

$R \perp\!\!\!\perp S \mid D$  will be coded by  $R \not\sim S$     **no edge between  $R$  and  $S$**

Results depend on Department (knowing Sex):  $R \sim D$   
 $S$  depends on  $D$  (knowing Results):  $S \sim D$



**Remark.**  $D$  separates  $R$  from  $S$

(any path from  $R$  to  $S$  goes through  $D$ . Direct route from  $R$  to  $S$  impossible.)

We consider undirected graphs  $\mathcal{G} = (V, E)$  where:

- the vertex (node) set  $V = \{1, \dots, n\}$
- the edge set  $E \subset \{F \subset V \mid \text{card}(F) = 2\}$

Any 2-element set  $\{i, j\} \in E$  will be called an edge.  
Then we write  $i \sim j$ .

Consider a system of random variables  $X_1, \dots, X_p$  on the same probability space  $(\Omega, \mathcal{T}, P)$ .

The information on **conditional independence** between the  $X_i$ 's is schematized by an undirected graph  $\mathcal{G} = (V, E)$  such that

$$X_l \perp\!\!\!\perp X_m \mid X_{V \setminus \{l, m\}} \iff l \not\sim m.$$

The graph  $\mathcal{G}$  is called the *dependence graph* of the system of random variables  $X_1, \dots, X_p$ .

## Conditional Independence in a Gaussian vector

Let  $X = (X_1, \dots, X_p)$  be a Gaussian vector on  $\mathbb{R}^p$ , with law  $N_p(\xi, \Sigma)$  and invertible  $\Sigma$ .

The matrix  $K = \Sigma^{-1}$  is called the precision matrix of the Gaussian vector  $X$ .

( $K$  is also called concentration matrix)



This is the precision matrix  $K = (\kappa_{ij})_{i,j \leq p}$  that appears in the Gaussian density

$$f(x) = (2\pi)^{-p/2} (\det K)^{1/2} e^{-(x - \xi)^T K (x - \xi)/2}$$

**Proposition.** Let  $X$  be a Gaussian vector in  $\mathbb{R}^p$ . Denote  $V = \{1, \dots, p\}$  the index set. Let  $l, m \in V$  and  $l \neq m$ .

The **marginals**  $X_l, X_m$  **are conditionally independent** w.r. to all the other variables  $X_{V \setminus \{l, m\}}$

$$X_l \perp\!\!\!\perp X_m \mid X_{V \setminus \{l, m\}}$$

**if and only if**

$$\kappa_{lm} = 0$$

i.e. the  $lm$ -term of the precision matrix  $K$  is equal 0.

### Proof in dim 3.

$$X \sim N(0, \Sigma), \quad \boxed{X_1 \perp\!\!\!\perp X_2 | X_3} \quad K = (\kappa_{ij})_{i,j \leq 3} = \Sigma^{-1}$$

Factorization Property of the joint density:

$$X_1 \perp\!\!\!\perp X_2 | X_3 \Leftrightarrow f(x_1, x_2, x_3) = g(x_1, x_3)h(x_2, x_3)$$

(analogous to  $X_1 \perp\!\!\!\perp X_2 \Leftrightarrow f(x_1, x_2) = g(x_1)h(x_2)$ )

$$f(x_1, x_2, x_3) = \frac{(\det K)^{1/2}}{(2\pi)^{3/2}} \times$$
$$e^{-(\kappa_{11}x_1^2 + \kappa_{22}x_2^2 + \kappa_{33}x_3^2 + 2\kappa_{12}x_1x_2 + 2\kappa_{13}x_1x_3 + 2\kappa_{23}x_2x_3)/2}$$

Suppose  $\boxed{X_1 \perp\!\!\!\perp X_2 | X_3} \Leftrightarrow f(x_1, x_2, x_3) = g(x_1, x_3)h(x_2, x_3) = C \times$

$$e^{-(\kappa_{11}x_1^2 + \kappa_{22}x_2^2 + \kappa_{33}x_3^2 + 2\kappa_{12}x_1x_2 + 2\kappa_{13}x_1x_3 + 2\kappa_{23}x_2x_3)/2}$$

Obligatorily  $2\kappa_{12}x_1x_2 = 0 \Leftrightarrow \boxed{\kappa_{12} = 0}$ .

## GAUSSIAN GRAPHICAL MODELS

Let  $V = \{1, \dots, p\}$  and let  $\mathcal{G} = (V, E)$  be an undirected graph. Let

$$\mathcal{S}(\mathcal{G}) = \{Z \in \text{Sym}(n \times n) \mid i \not\sim j \Rightarrow Z_{ij} = 0\},$$

the space of symmetric matrices with **obligatory zero terms**  $Z_{ij} = 0$  for  $i \not\sim j$ .

**Definition.** The **GAUSSIAN GRAPHICAL MODEL governed by the graph  $\mathcal{G}$**  is the set of all random Gaussian vectors  $X = (X_v)_{v \in V} \sim N(\xi, \Sigma)$ , with precision matrix  $K = \Sigma^{-1} \in \mathcal{S}(\mathcal{G})$ .

**Remark.** The **complete graph**  $\mathcal{G}$  (i.e.  $\mathcal{G}$  containing all possible edges) defines Gaussian graphical model containing all Gaussian vectors supported by  $\mathbb{R}^p$ , with no constraint. Such model is called **saturated**.

**Precision matrix  $K = \Sigma^{-1}$  of a Gaussian vector contains:**

- **information on conditional independence:**

$$X_l \perp\!\!\!\perp X_m \mid X_{V \setminus \{l,m\}} \iff \kappa_{lm} = 0$$

- **conditional precision matrices ( $K_{X_A|X_B} = K_{AA}$ )**  
where  $A \cup B = V, A \cap B = \emptyset$

- in practice, we use conditional correlation

$$\rho_{lm|V\setminus\{l,m\}} = \frac{\text{Cov}(X_l, X_m | X_{V\setminus\{l,m\}})}{\text{Var}(X_l | X_{V\setminus\{l,m\}})^{\frac{1}{2}} \text{Var}(X_m | X_{V\setminus\{l,m\}})^{\frac{1}{2}}} = -\tilde{\kappa}_{lm}$$

where  $\tilde{\kappa}_{lm} \stackrel{\text{df}}{=} \frac{\kappa_{lm}}{\sqrt{\kappa_{ll}}\sqrt{\kappa_{mm}}}$  is an element of the so-called **scaled precision matrix  $\tilde{K}$** .

Intuitively,

$$\rho_{lm|V\setminus\{l,m\}} = -\tilde{\kappa}_{lm} \approx 0 \Rightarrow X_l \perp\!\!\!\perp X_m | X_{V\setminus\{l,m\}}.$$

The matrix  $R$  with terms  $\rho_{lm|V\setminus\{l,m\}}$  is called **partial correlation matrix**.

For off-diagonal terms,  $R_{lm} = -\tilde{\kappa}_{lm}$ .

**Example.** Marks of 88 students in 5 exams:  $N_5(\xi, \Sigma)$

Table 1.1.1: Marks in five mathematics exams for 88 students. From Mardia, Kent and Bibby (1979). [WHITTAKER]

me	ve	al	an	st	me	ve	al	an	st	me	ve	al	an	st
77	82	67	67	81	30	69	50	52	45	62	44	36	22	42
63	78	80	70	81	46	49	53	59	37	48	38	41	44	33
75	73	71	66	81	40	27	54	61	61	34	42	50	47	29
55	72	63	70	68	31	42	48	54	68	18	51	40	56	30
63	63	65	70	63	36	59	51	45	51	35	36	46	48	29
53	61	72	64	73	56	40	56	54	35	59	53	37	22	19
51	67	65	65	68	46	56	57	49	32	41	41	43	30	33
59	70	68	62	56	45	42	55	56	40	31	52	37	27	40
62	60	58	62	70	42	60	54	49	33	17	51	52	35	31
64	72	60	62	45	40	63	53	54	25	34	30	50	47	36
52	64	60	63	54	23	55	59	53	44	46	40	47	29	17
55	67	59	62	44	48	48	49	51	37	10	46	36	47	39
50	50	64	55	63	41	63	49	46	34	46	37	45	15	30
65	63	58	56	37	46	52	53	41	40	30	34	43	46	18
31	55	60	57	73	46	61	46	38	41	13	51	50	25	31
60	64	56	54	40	40	57	51	52	31	49	50	38	23	9
44	69	53	53	53	49	49	45	48	39	18	32	31	45	40
42	69	61	55	45	22	58	53	56	41	8	42	48	26	40
62	46	61	57	45	35	60	47	54	33	23	38	36	48	15
31	49	62	63	62	48	56	49	42	32	30	24	43	33	25
44	61	52	62	46	31	57	50	54	34	3	9	51	47	40
49	41	61	49	64	17	53	57	43	51	7	51	43	17	22
12	58	61	63	67	49	57	47	39	26	15	40	43	23	18
49	53	49	62	47	59	50	47	15	46	15	38	39	28	17
54	49	56	47	53	37	56	49	28	45	5	30	44	36	18
54	53	46	59	44	40	43	48	21	61	12	30	32	35	21
44	56	55	61	36	35	35	41	51	50	5	26	15	20	20
18	44	50	57	81	38	44	54	47	24	0	40	21	9	14
46	52	65	50	35	43	43	38	34	49					
32	45	49	57	64	39	46	46	32	43					

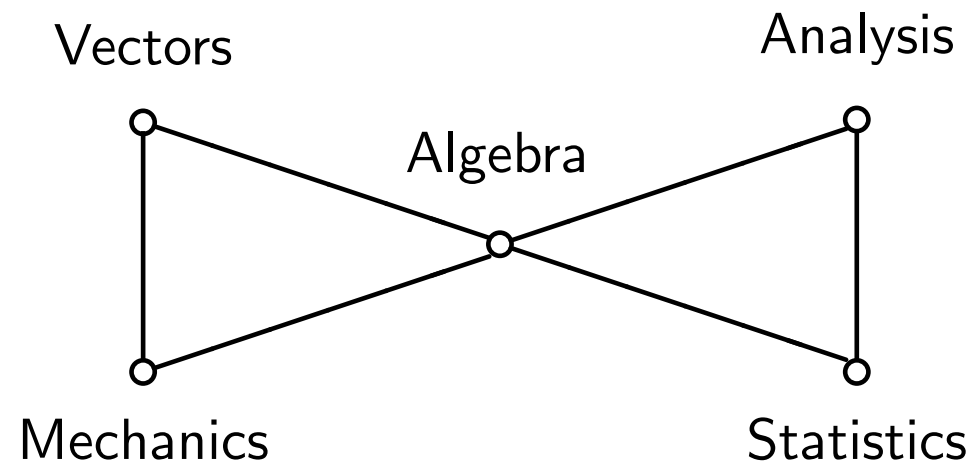


# Mathematics marks

Examination marks of 88 students in 5 different mathematical subjects. The empirical concentrations (on or above diagonal) and partial correlations (below diagonal) are

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	5.24	-2.44	-2.74	0.01	-0.14
Vectors	0.33	10.43	-4.71	-0.79	-0.17
Algebra	0.23	0.28	26.95	-7.05	-4.70
Analysis	-0.00	0.08	0.43	9.88	-2.02
Statistics	0.02	0.02	0.36	0.25	6.45

# Graphical model for mathmarks



This analysis is from Whittaker (1990).  
We have  $An, Stats \perp\!\!\!\perp Mech, Vec \mid Alg$ .

A well-known reference on Graphical Models in statistics is the book:

S.L. Lauritzen, *Graphical Models*, Oxford 1996.

A very recent reference on Graphical Models in statistics is the book:

Maathuis, M., Drton, M., Lauritzen, S. and Wainwright, M. editors ,  
*Handbook of Graphical Models*,  
Chapman and Hall - CRC Handbooks of Modern Statistical Methods, Chapman and Hall, **2018**, 536 p.

## COLORED GRAPHICAL GAUSSIAN MODELS

In order to make Graphical Gaussian Models a viable modeling tool in the modern Big Data Science, i.e. when the number of variables outgrows the number of observations,  $p \gg n$ ,

Højsgaard and Lauritzen introduced in 2008 models which impose **equality restrictions on certain entries of precision matrix  $K$  or partial correlation matrix  $R$ .**

Such models can be represented by **colored** graphs:

**the same color edges of the graph encode the equal entries of the matrix**

**Exemple.**  $p = 3$

The complete graph (saturated model).  $\dim =$

A colored complete graph.  $\dim =$

**Exemple.**  $p = 3$

The linear graph (Simpson paradox model).  $\dim =$

A colored linear graph.  $\dim =$

Three types of restriction on graphical Gaussian models are introduced by Højsgaard and Lauritzen :

- **RCON models:**  
equality among specified elements of the **concentration matrix**,  
(*R for Restriction, CON for concentration = RCON*)
- **RCOR models:**  
equality among specified partial(conditional) variances  $\kappa_{ll}$  and **partial correlations**  $R_{lm}$  ,
- **RCOP models:**  
restrictions on concentrations and partial correlations generated by a subgroup of the group of **permutations**

## **RCOP models** $(\mathcal{G}, \Gamma)$

*(Permutation invariant coloured models)*

Consider a Gaussian graphical model on  $\mathbb{R}^p$  governed by graph  $\mathcal{G}$ .

Permutation-generated colourings are done according to a subgroup

$$\Gamma \subset \text{Aut}(\mathcal{G}) \subset \mathfrak{S}_p$$

of the permutation (symmetric) group  $\mathfrak{S}_p$ .

$\text{Aut}(\mathcal{G})$  denotes permutations which are automorphisms of the graph  $\mathcal{G}$ , **mapping edges to edges**

$$\sigma(\alpha) \sim \sigma(\beta) \Leftrightarrow \alpha \sim \beta \quad \text{if } \sigma \in \Gamma.$$



**Example.**  $p = 3$

The linear graph (Simpson paradox model).

$\text{Aut}(\mathcal{G}) =$

By definition, permutation invariant colouring  $(\mathcal{V}, \mathcal{E})$  of  $\mathcal{G}$  according to  $\Gamma$  is given by the orbits of  $\Gamma$  in  $V$  and  $E$  respectively.

Two vertices  $\alpha, \beta \in V$  have the same colour whenever there exists  $\sigma \in \Gamma$  mapping  $\alpha$  to  $\beta$ , and similarly for the edges.

**Example.**  $p = 3$

The linear graph (Simpson paradox model).

$$\text{Aut}(\mathcal{G}) = \{\text{id}, (1, 3)\}$$

2 possible subgroups  $\Gamma \subset \text{Aut}(\mathcal{G})$ :

$\Gamma = \{\text{id}\}$ : no coloring

$\Gamma = \text{Aut}(\mathcal{G})$

For a subgroup  $\Gamma \subset \mathfrak{S}_p$ , we define the space of symmetric matrices invariant under  $\Gamma$ , or the **colored matrix space**,

$$\mathcal{Z}_\Gamma := \left\{ x \in \text{Sym}(p; \mathbb{R}) \mid x_{ij} = x_{\sigma(i)\sigma(j)} \text{ for all } \sigma \in \Gamma \right\},$$

and the cone of positive definite matrices in  $\mathcal{Z}_\Gamma$ ,

$$\mathcal{P}_\Gamma := \mathcal{Z}_\Gamma \cap \text{Sym}^+(p; \mathbb{R}).$$

Equivalently,

$$\mathcal{Z}_\Gamma = \left\{ x \in \text{Sym}(p; \mathbb{R}) \mid R(\sigma) \cdot x = x \cdot R(\sigma) \text{ for all } \sigma \in \Gamma \right\},$$

where  $R(\sigma)$  denotes the (permutation) matrix of  $\sigma$ .

**Example.**  $p = 3, \sigma = (1, 3)$

$$R(\sigma) = ? \quad R(\sigma) \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_3 \\ v_2 \\ v_1 \end{pmatrix}$$

In **RCOP Gaussian models**  $N(0, \Sigma)$ , we have, **equivalently,**  
**equalities in the covariance matrix  $\Sigma$ :**

$$\text{For any } \sigma \in \Gamma, \Sigma_{ij} = \Sigma_{\sigma(i)\sigma(j)}$$

*Proof.*

RCOP invariance of the concentration matrix of a Gaussian model is equivalent to

$$R(\sigma) K R(\sigma)^t = K$$

for all  $\sigma \in \Gamma$ .

Take the inverse matrix.

We get  $R(\sigma) \Sigma R(\sigma)^t = \Sigma$  for all  $\sigma \in \Gamma$ .

**Summary.** RCOP model  $(\mathcal{G}, \Gamma)$  means two properties:

(i)  $K \in \mathcal{S}(\mathcal{G})$

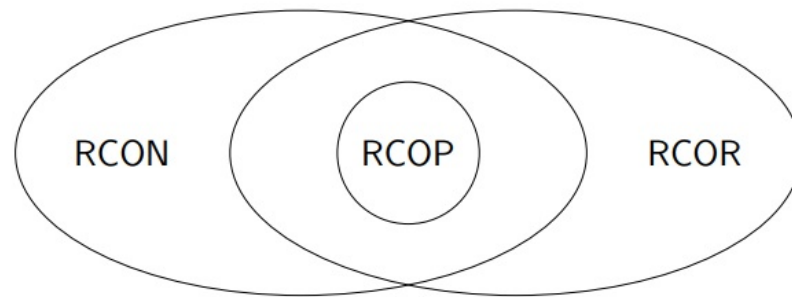
**Zeros** are disposed in  $K$  according to missing edges in  $\mathcal{G}$

(ii) For any  $\sigma \in \Gamma$ ,  $K_{ij} = K_{\sigma(i)\sigma(j)}$

"Clusters" (colors) of equal terms are disposed in  $K$  according to the orbits of  $\Gamma$

We denote by  $\mathcal{S}^\Gamma(\mathcal{G})$  the symmetric matrices satisfying the constraints (i) and (ii) imposed by the RCOP model  $(\mathcal{G}, \Gamma)$ .

## Relations between coloured graphical model types



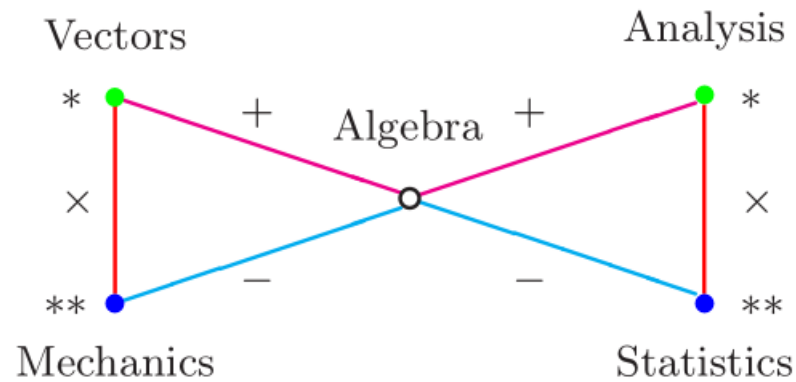
**Fig. 10.** Relations between symmetry models. Models given by permutation symmetry (RCOP) have similar symmetries for concentrations (RCON) and partial correlations (RCOR). RCON models are not necessarily RCOR and *vice versa* but a model can be simultaneously RCON and RCOR without being RCOP.



**Example** of an RCOP model from [Højsgaard and Lauritzen]:

**Table 6.** Fitted concentrations  $\times 1000$  (on and above diagonal) and fitted partial correlations (below diagonal) for the examination marks in five mathematical subjects assuming the RCOP model with coloured graph as in Figure 8.

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	5.75	-2.28	-3.70	0	0
Vectors	0.30	9.96	-6.44	0	0
Algebra	0.29	0.39	27.4	-6.44	-3.70
Analysis	0	0	0.39	9.79	-2.28
Statistics	0	0	0.29	0.30	5.75



This colored graph is invariant by the group containing two permutations:

$$\Gamma = \{id, (An, Ve) \circ (Me, St)\}$$

This group  $\Gamma$  is generated by one permutation

$$\sigma = (An, Ve) \circ (Me, St)$$

(note that  $\sigma^{-1} = \sigma$ )

Groups with one generator are called **cyclic groups**.

## Advantages of RCOP models

- We believe that there are permutation symmetries in the nature. RCOP models are colored according to groups of permutations
- A deep mathematical tool:  
theory of groups of permutations and of finite groups
- RCOP models have good statistical properties:  
Højsgaard-Lauritzen, Gehrman-Lauritzen: developed MLE theory of  $K$  and  $\Sigma$  for RCOP Gaussian models
- RCOP models are less numerous than RCON and RCOR models

## PART 2

P. Graczyk, H. Ishi, B. Kołodziejek, H. Massam

**Model selection in the space of Gaussian models invariant by symmetry**, submitted 2020

*<https://arxiv.org/abs/2004.03503>*

The open problems we address are motivated by

**Bayesian Model Selection on RCOP models  $(\mathcal{G}, \Gamma)$**

We consider an  $n$ -sample of  $X \sim N_p(\xi, \Sigma = K^{-1})$ , a Gaussian character in  $\mathbb{R}^p$ . The density of  $X$  is

$$f(x, K) = (2\pi)^{-p/2} (\det K)^{1/2} e^{-(x - \xi)^T K (x - \xi) / 2} = (2\pi)^{-p/2} (\det K)^{1/2} e^{-\frac{1}{2} \text{tr}(Ku)}$$

where  $u = (x - \xi)(x - \xi)^T$ .

In Bayesian statistics, the parameter  $K$  is **random**.

The formula for  $f(x, K)$  suggests using as an *a priori* law of  $K$  (prior distribution of  $K$ ) the law with density

$$K \rightarrow \frac{1}{I_{\mathcal{G}}^{\Gamma}(s, \theta)} (\text{Det} K)^{\frac{s}{2}} e^{-\frac{1}{2} \text{tr}(K\theta)}, \quad K \in \mathcal{S}^{\Gamma}(\mathcal{G})^+ \stackrel{\text{df}}{=} \mathcal{P}_{\mathcal{G}}^{\Gamma},$$

where  $\theta \in \mathcal{S}^{\Gamma}(\mathcal{G})^+$ . This is a **Diaconis-Ylvisaker conjugate prior** for the precision matrix  $K$ .

The constant  $\frac{1}{I_{\mathcal{G}}^{\Gamma}(s, \theta)}$  normalizes the above function to the probability density, with

$$I_{\mathcal{G}}^{\Gamma}(s, \theta) = \int_{\mathcal{P}_{\mathcal{G}}^{\Gamma}} (\det K)^{\frac{s}{2}} e^{-\frac{1}{2} \text{tr}(K\theta)} dK.$$

Computation of this Laplace-type integral  $I_{\mathcal{G}}^{\Gamma}(s, \theta)$  may be, in general, very difficult (or hopeless...), because of the complicated structure of the matrix cone  $\mathcal{P}_{\mathcal{G}}^{\Gamma}$ .

The a priori law on  $K$

$$K \rightarrow \frac{1}{I_{\mathcal{G}}^{\Gamma}(s, \theta)} (\text{Det}K)^{\frac{s}{2}} e^{-\frac{1}{2}\text{tr}(K\theta)}, \quad K \in \mathcal{S}^{\Gamma}(\mathcal{G})^{+} \stackrel{\text{df}}{=} \mathcal{P}_{\mathcal{G}}^{\Gamma},$$

is called a  $(\mathcal{G}, \Gamma)$ -Wishart law on the matrix cone  $\mathcal{P}_{\mathcal{G}}^{\Gamma}$ , with parameters  $s, \theta$ , denoted  $W_{\mathcal{G}}^{\Gamma}(s, \theta)$ .

This is a matrix analog of KHI<sup>2</sup> law  $\chi_s^2$  sur  $\mathbb{R}^+$ .

## Bayesian Model Selection on RCOP models $(\mathcal{G}, \Gamma)$

Without loss of generality, suppose the centered case  $\xi = 0$ . We assume a uniform prior distribution on the space of RCOP models  $(\mathcal{G}, \Gamma)$  of cardinal  $g$ .

The likelihood function of the sample  $X^{(1)}, \dots, X^{(n)} \in \mathbb{R}^p$  equals:

$$\begin{aligned} f(x^{(1)}, \dots, x^{(n)}; K; \mathcal{G}, \Gamma) &= \\ &= \frac{1}{g} \prod_{k=1}^n \left\{ (2\pi)^{-p/2} (\det K)^{1/2} \exp(-x^{(k)T} K x^{(k)} / 2) \right\} \\ &= \frac{1}{g} (2\pi)^{-np/2} (\det K)^{n/2} \exp(-\frac{1}{2} \text{tr}(K u(x))) \end{aligned}$$

where  $u(x) = \sum_{k=1}^n x^{(k)} x^{(k)T}$  equals the empirical sample covariance  $\times n$ .



In order to get the joint distribution of  $(X, \mathcal{G}, \Gamma)$ , we integrate with respect to the prior Wishart law of  $K$ :

$$\begin{aligned}
 f(x^{(1)}, \dots, x^{(n)}; \mathcal{G}, \Gamma) &= \\
 \frac{1}{g} (2\pi)^{-np/2} \int_{\mathcal{P}_{\mathcal{G}}^{\Gamma}} (\det K)^{n/2} \exp\left(-\frac{1}{2} \text{tr}(Ku)\right) dW_{\mathcal{G}}^{\Gamma}(s, \theta)(K) \\
 &= \frac{1}{g} (2\pi)^{-np/2} \frac{I_{\mathcal{G}}^{\Gamma}(s + n, \theta + u(x))}{I_{\mathcal{G}}^{\Gamma}(s, \theta)}
 \end{aligned}$$

This is proportional (with factor  $\frac{1}{f(x^{(1)}, \dots, x^{(n)})}$ ) to the posterior distribution  $f(\mathcal{G}, \Gamma | X = x)$  of  $(\mathcal{G}, \Gamma)$  given the observation  $x = (x^{(1)}, \dots, x^{(n)})$ .

We apply the [Bayesian paradigm](#), as described for non-colored graphical models in [Handbook of Graphical Models, 2018]:

*In a Bayesian framework, the classical approach for choosing between two models  $\mathcal{M}_{(\mathcal{G}, \Gamma)}$  and  $\mathcal{M}_{(\mathcal{G}', \Gamma')}$  is to compute their posterior probability density and choose the model with the highest posterior probability.*

The **Bayes factor** equals the ratio of the posterior distributions of two compared models

$$B_{(\mathcal{G},\Gamma),(\mathcal{G}',\Gamma')} = \frac{I_{\mathcal{G}}^{\Gamma}(s+n, \theta+u(x))I_{\mathcal{G}'}^{\Gamma'}(s, \theta)}{I_{\mathcal{G}}^{\Gamma}(s, \theta)I_{\mathcal{G}'}^{\Gamma'}(s+n, \theta+u(x))}$$

Knowing the Bayes factors  $B_{(\mathcal{G},\Gamma),(\mathcal{G}',\Gamma')}$ , we travel through the space of RCOP models  $(\mathcal{G}, \Gamma)$  using the **Metropolis-Hastings algorithm**.

## Metropolis-Hastings algorithm over cyclic permutation groups

Starting from a permutation  $\sigma_0 \in \mathfrak{S}_p$ , repeat the following two steps for  $t = 1, 2, \dots$ :

1. Sample  $x_t$  uniformly from the set  $\mathcal{T}$  of all transpositions and set  $\sigma' = \sigma_{t-1} \circ x_t$ ;
2. Accept the move  $\sigma_t = \sigma'$  with probability

$$\min \left\{ 1, \frac{I_{\langle \sigma' \rangle}(s+n, \theta+u(x)) I_{\langle \sigma_{t-1} \rangle}(s, \theta)}{I_{\langle \sigma' \rangle}(s, \theta) I_{\langle \sigma_{t-1} \rangle}(s+n, \theta+u(x))} \right\}.$$

If the move is rejected, set  $\sigma_t = \sigma_{t-1}$ .

## Simulations

Let the covariance matrix  $\Sigma$  be the symmetric circulant matrix

$$\Sigma = \begin{pmatrix} c_0 & c_1 & \dots & c_2 & c_1 \\ c_1 & c_0 & c_1 & \dots & c_2 \\ \vdots & c_1 & c_0 & \dots & \vdots \\ c_2 & & \dots & \dots & c_1 \\ c_1 & c_2 & \dots & c_1 & c_0 \end{pmatrix},$$

with  $c_0 = 1 + 1/p$  and  $c_k = 1 - k/p$  for  $k = 1, \dots, \lfloor p/2 \rfloor$ . It is easily seen that this matrix belongs to  $\mathcal{P}_{\langle \sigma^* \rangle}$  with  $\sigma^* = (1, 2, \dots, p-1, p)$ .

We performed on Python:

$T = 100\,000$  steps of M-H Algorithm with  $\sigma_0 = \text{id}$ ,  
 $p = 100$ ,  $n = 200$ ,  $s = 3$  and  $\theta = I_{100}$ .

Let us note that for  $p = 100$ , there are about  $4 \cdot 10^{155}$  cyclic subgroups and this is the number of models we consider in our model search.

The highest estimated posterior probability was obtained for  $c^* := \langle \sigma^* \rangle$ , where

$\sigma^* = (1, 2, 3, 4)(6, 8, 15)(7, 10, 9)(11, 16, 12)(13, 17, 14)(18, 19, 20, 22, 21)(23, 26)$   
 $(24, 42, 28, 44)(25, 31, 30, 32)(27, 34)(29, 37)(33, 45)(35, 39, 36, 40)$   
 $(38, 47, 41, 48)(43, 51, 46, 49)(50, 52, 53, 54)(56, 58, 57)(59, 66, 67)$   
 $(60, 65, 63)(61, 62, 64)(68, 71, 72, 70, 69)(73, 93)(74, 77)(75, 98, 81, 100)$   
 $(76, 84, 78, 83)(79, 85)(80, 94, 82, 91)(86, 92, 87, 90)(88, 96, 89, 97)(95, 99).$

The estimate of the posterior probability  $\pi_{c^*}$  is equal to 37%.

The order of  $c^*$  is  $|c^*| = 60$ .

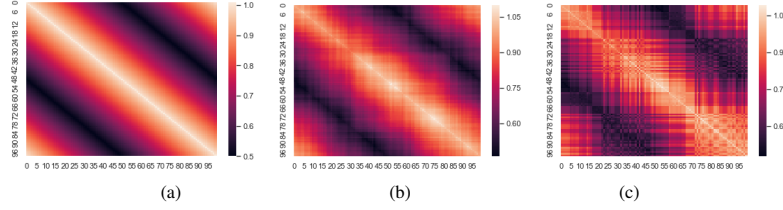


FIG 3. Heat map of matrix  $\Sigma$  (a) and matrix  $U/n$  (b) and projection of  $U/n$  onto  $\tilde{\mathcal{Z}}_{c^*}$ .

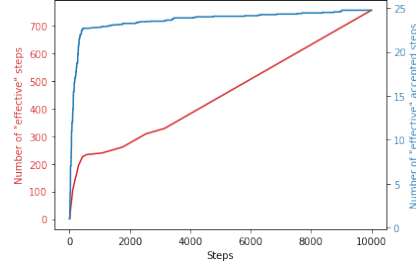


FIG 4. Number of “effective” steps (red) and number of “effective” accepted steps (blue).

$$(76, 84, 78, 83)(79, 85)(80, 94, 82, 91)(86, 92, 87, 90)(88, 96, 89, 97)(95, 99).$$

The order of  $c^*$  is  $|c^*| = 60$  and  $\Phi(c^*) = 16$ . The estimate of the posterior probability  $\pi_{c^*}$  is equal to (recall (47))

$$\frac{\frac{1}{\Phi(c^*)} \sum_{t=1}^T \mathbf{1}_{c^* = \langle \sigma_t \rangle}}{\sum_{t=1}^T \frac{1}{\Phi(\langle \sigma_t \rangle)}} \approx \frac{2361.5}{6381.5} \approx 37\%.$$

The true covariance matrix  $\Sigma$ , the data matrix  $U/n$  and the projection  $\Pi_{c^*}(U/n)$  are illustrated in Fig. 3.

We visualize the performance of the algorithm on Fig 4. In red color, a sequence  $\left(\sum_{t=1}^k \frac{1}{\Phi(\langle \sigma_t \rangle)}\right)_k$  is depicted, which can be thought of as an “effective” number of steps of the algorithm (for an explanation, see the paragraph at the end of Subsection 4.1.2). In blue, we present a sequence  $\left(\sum_{t=1}^k \frac{1}{\Phi(\langle \sigma_t \rangle)} \mathbf{1}_{\langle \sigma_t \rangle \neq \langle \sigma_{t-1} \rangle}\right)_k$ , which represents the number of weighted accepted steps, where the weight of the  $k$ th step equals  $\frac{1}{\Phi(\langle \sigma_k \rangle)}$ . We restricted the plot to steps  $k = 1, \dots, 10000$ , because after 10000 steps, the Markov chain  $(\sigma_t)_{10000 \leq t \leq 100000}$  changed its state only 9 times. For  $k = 10000$ , the value of the blue curve is 25.75, while the value of red one is 6381.5.

The model suffers from poor acceptance rate, which could be improved by an appropriate choice of the hyper-parameter  $D$  or by allowing the Markov chain to do bigger steps.

**Acknowledgements.** The authors would like to thank Steffen Lauritzen for his interest and encouragements.



The computation of the **normalizing constants**

$$I_{\mathcal{G}}^{\Gamma}(s, \theta) = \int_{\mathcal{P}_{\mathcal{G}}^{\Gamma}} (\text{Det}K)^{\frac{s}{2}} e^{-\frac{1}{2}\text{tr}(K\theta)} dK$$

is essential for the computation of the **Bayes factor** and for the M-H Algorithm.

We need to compute the integral

$$I_{\mathcal{G}}^{\Gamma}(s, \theta) = \int_{\mathcal{P}_{\mathcal{G}}^{\Gamma}} (\det K)^{\frac{s}{2}} e^{-\frac{1}{2}\text{tr}(K\theta)} dK.$$

This is a generalized Gamma Integral.

Recall [Gamma integral](#). For  $s > 0$

$$\int_0^{\infty} e^{-xy} x^{s-1} dx = \Gamma(s) y^{-s}$$

## Gamma Integrals on the cone of positive definite symmetric matrices $S_n^+ = \text{Sym}^+(n, \mathbb{R})$

Crucial in multivariate statistics, first computed by Wishart and Ingham 1928-1935.

Gamma function of  $S_n^+$ : for  $s > \frac{n-1}{2}$  and  $c_n = (2\pi)^{\frac{n(n-1)}{4}}$

$$\Gamma_{S_n^+}(s) = \int_{S_n^+} e^{-\text{tr}(x)} (\det x)^{s - \frac{n+1}{2}} dx = c_n \prod_{j=1}^n \Gamma\left(s - \frac{j-1}{2}\right)$$

## Gamma-Siegel integral

for  $s > \frac{n-1}{2}$

$$\int_{S_n^+} e^{-\text{tr}(xy)} (\det x)^{s - \frac{n+1}{2}} dx = \Gamma_{S_n^+}(s) (\det y)^{-s}$$

**Gamma integrals over the irreducible symmetric cones  $\Omega_i = \text{Herm}(r_i, \mathbb{K}_i)^+$ ,  $\mathbb{K}_i \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ .**

Let  $d_i = \dim_{\mathbb{R}} \mathbb{K}_i \in \{1, 2, 4\}$ .

Let  $N_i = r_i + \frac{r_i(r_i-1)}{2}d_i$  denote the dimension of  $\text{Herm}(r_i, \mathbb{K}_i)$

$$\int_{\Omega_i} e^{-\text{tr } xy} (\det x)^{\lambda - N_i/r_i} dx = \Gamma_{\Omega_i}(\lambda) (\det y)^{-\lambda}$$
$$\left( \lambda > \frac{(r_i - 1)d_i}{2}, y \in \Omega_i \right)$$

where  $\Gamma_{\Omega_i}$  denotes the Gamma function associated to the symmetric cone  $\Omega_i$  given by (see [Faraut-Koranyi])

$$\Gamma_{\Omega_i}(\lambda) = (2\pi)^{(N_i - r_i)/2} \Gamma(\lambda) \Gamma(\lambda - d_i/2) \dots \Gamma(\lambda - (r_i - 1)d_i/2).$$

**Our main result:**

**Computing normalizing constant of the Wishart distribution in RCOP model for the full(saturated) graph  $\mathcal{G}$ .**

Such explicit formulas were not known before.

**Method:** the representation theory (over  $\mathbb{R}$ ) of a finite group.

**Application.** **Bayesian Model selection:** selecting the best coloring subgroup  $\Gamma$  of  $\mathfrak{S}_p$ , for a Gaussian vector  $X \in \mathbb{R}^p$  with no information on conditional independence of  $X_i$ 's (saturated model)

**In progress.** The case when  $\mathcal{G}$  is a decomposable graph.

From now on,  $\mathcal{G}$  is a complete graph with  $p$  vertices and all edges.

For a subgroup  $\Gamma \subset \mathfrak{S}_p$ , we define the colored matrix vector space

$$\mathcal{Z}^\Gamma := \left\{ x \in \text{Sym}(p, \mathbb{R}) \mid x_{ij} = x_{\sigma(i)\sigma(j)} \text{ for all } \sigma \in \Gamma \right\},$$

We need to compute Gamma integrals on the cone

$$\mathcal{P}^\Gamma = \mathcal{Z}^\Gamma \cap \text{Sym}^+(p, \mathbb{R})$$

**Theorem 1. Crucial property: reduction of colored saturated models to **bloc diagonal models****

Any matrix  $X \in \mathcal{Z}^\Gamma$  (so also when  $X \in \mathcal{P}_\Gamma$ ) can be written as

$$X = U_\Gamma \left( \begin{array}{cccc} M_{\mathbb{K}_1}(x_1) \otimes I_{\frac{k_1}{d_1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M_{\mathbb{K}_2}(x_2) \otimes I_{\frac{k_2}{d_2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & M_{\mathbb{K}_L}(x_L) \otimes I_{\frac{k_L}{d_L}} \end{array} \right) U_\Gamma^T$$

where: •  $U_\Gamma$  is an orthogonal (change of basis) matrix **independent on  $X \in \mathcal{Z}^\Gamma$** ,

•  $M_{\mathbb{K}_i}(x_i)$  is a real matrix representation of a Hermitian  $r_i \times r_i$  matrix  $x_i$  with entries in  $\mathbb{K}_i = \mathbb{R}, \mathbb{C}$  or  $\mathbb{H}$ .

The maps  $X \rightarrow x_i$  are linear  $\mathcal{Z}^\Gamma \rightarrow \text{Herm}(r_i, \mathbb{K}_i)$

•  $k_i, d_i, r_i, k_i/d_i, i = 1, \dots, L$  are integer constants

Theorem 1 says that in an adequate basis of  $\mathbb{R}^p$ , any  $\Gamma$ -colored matrix  $X \in \mathcal{Z}^\Gamma$  can be written as a **bloc diagonal matrix**.

Recall that both  $\text{Herm}(r; \mathbb{C})$  and  $\text{Herm}(r; \mathbb{H})$  can be realized as real symmetric matrices, but of bigger dimension ( $2r$  and  $4r$ , respectively).

For  $z = a + bi \in \mathbb{C}$  define  $M_{\mathbb{C}}(z) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ . The function  $M_{\mathbb{C}}$  is a matrix representation of  $\mathbb{C}$ .

Similarly, any  $r \times r$  complex matrix can be realized as a  $(2r) \times (2r)$  real matrix. Hermitian complex matrix becomes a real symmetric matrix. Positive definiteness is preserved.



Analogous real  $4 \times 4$  matrix realization is available for quaternions  $\mathbb{H}$ .

**Example.** For  $p = 3$ ,  $\Gamma = \mathfrak{S}_3$ , the cone  $\mathcal{P}_\Gamma$  of positive definite matrices  $X$  invariant under  $\Gamma$ , that is such that  $X_{ij} = X_{\sigma(i)\sigma(j)}$  for all  $\sigma \in \Gamma$ , is

$$\mathcal{P}_\Gamma = \left\{ \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix} \mid a > 0 \text{ and } b \in (-a/2, a) \right\}.$$

The bloc diagonal realization of  $\mathcal{P}_\Gamma$  and

$$\begin{pmatrix} a & b & b \\ b & a & b \\ a & a & b \end{pmatrix} = U_\Gamma \begin{pmatrix} a + 2b & 0 & 0 \\ 0 & a - b & 0 \\ 0 & 0 & a - b \end{pmatrix} {}^t U_\Gamma,$$

Here  $L = 2$ ,  $\mathbb{K}_i = \mathbb{R}$ ,  $r_i = 1$ ,  $i = 1, 2$  and the constants  $k_1 = d_1 = 1$ ,  $k_2 = 2$ ,  $d_2 = 1$ ,  $k_1/d_1 = 1$ ,  $k_2/d_2 = 2$ .

This bloc diagonal realization is obtained choosing  $U_\Gamma := \left( v_1^{(1)} \quad v_2^{(1)} \quad v_3^{(1)} \right) \in O(3)$  with

$$v_1^{(1)} := \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad v_2^{(1)} := \begin{pmatrix} \sqrt{2/3} \\ -1/\sqrt{6} \\ -1/\sqrt{6} \end{pmatrix}, \quad v_3^{(1)} := \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}.$$

## *Proof of Theorem 1.*

Let  $R(\sigma)$  be the matrix of a permutation  $\sigma$  seen as a linear map on  $\mathbb{R}^p$ .

We apply a decomposition

$$R = \bigoplus_{i=1}^L \pi_i^{\oplus r_i}$$

of the natural representation

$$\Gamma \ni \sigma \mapsto R(\sigma) \in GL(p, \mathbb{R})$$

of the group  $\Gamma$  on  $\mathbb{R}^p$  into irreducible real representations  $\pi_i$  over  $V_i = \mathbb{R}^{k_i}$ , where:  $k_i = \dim_{\mathbb{R}} \pi_i$  and  $r_i$  is the multiplicity of  $\pi_i$  in the decomposition of  $R(\sigma)$ . □

## Theorem 2.

If  $Y \in \mathcal{P}_\Gamma$  and  $s > \max_{i=1, \dots, L} \left\{ -\frac{1}{k_i} \right\}$ , then

$$\begin{aligned} I(s, Y) &= \int_{\mathcal{P}_\Gamma} \text{Det}^s X e^{-\text{Tr} Y \cdot X} dX = \\ &= e^{-A_\Gamma s - B_\Gamma} \prod_{i=1}^L \Gamma_{\Omega_i} \left( k_i s + \frac{\dim \Omega_i}{r_i} \right) \text{Det}^{-s} Y \varphi_\Gamma(Y) \end{aligned}$$

where  $A_\Gamma := \sum_{i=1}^L r_i k_i \log k_i$ ,  
 $B_\Gamma := \frac{1}{2} \sum_{i=1}^L (\dim \Omega_i) (\log k_i)$ ,  
 $\varphi_\Gamma(Y) = \prod_{i=1}^L (\det \phi_i(Y))^{-\dim \Omega_i / r_i}$ ,  
with natural projections  $\phi_i : \mathcal{Z}_\Gamma \ni X \mapsto x_i \in \text{Herm}(r_i; \mathbb{K}_i)$ .

## STATISTICAL AND NUMERICAL APPLICATIONS OF Theorem 2

In order to compute the normalizing constants  $I(s, \theta)$  on  $\mathcal{P}^\Gamma$ , and the Bayes factors, we have to:

- find the constants  $k_i, d_i, r_i, i = 1, \dots, L$
- find the polynomials  $\det \phi_i(X)$

Doing so for every  $\Gamma$  visited during the model selection process is computationally heavy.

For important classes of subgroups  $\Gamma$  we can find  $U_\Gamma$  and the constants  $k_i, d_i, r_i, i = 1, \dots, L$ :

- **cyclic** (generated by one permutation) subgroups  $\Gamma$   
(model selection done with R for  $p = 100$ )
- Abelian subgroups  $\Gamma$

**Without restrictions on  $\Gamma$  and for small to moderate dimensions  $p \leq 8$ ,** we can obtain the constants  $k_i, d_i, r_i, i = 1, \dots, L$  as well as the expression of  $\text{Det}(X)$  and  $\varphi_\Gamma(X)$  without having to compute  $U_\Gamma$ , in the following way.

We use the packages of factorization of multivariate polynomials available on either Mathematica or Python.

**Fact.**  $\text{Det}(X)$  admits a unique irreducible factorization of the form

$$\text{Det}(X) = \prod_{i=1}^L \text{Det}\left(M_{\mathbb{K}_i}(x_i)\right)^{k_i/d_i} \quad (X \in \mathcal{Z}_\Gamma). \quad (1)$$



On the other hand, Mathematica or Python provide a unique irreducible factorization of the form

$$\text{Det}(X) = \prod_{j=1}^{L'} f_j(X)^{a_j} \quad (X \in \mathcal{Z}_\Gamma). \quad (2)$$

where each  $a_j$  is a positive integer, each  $f_j(X)$  is an irreducible polynomial of  $X \in \mathcal{Z}_\Gamma$ , and  $f_i \neq f_j$  if  $i \neq j$ .

The constants  $k_i, d_i, r_i$ , as well as  $\det \phi_i(X) = \det M_{\mathbb{K}_i}(x_i)$  are obtained by identification of the two expressions of  $\text{Det}(X)$  in (1) and (2).

## APPLICATION. Permutation Coloured (RCOP) saturated graphical models with **four** vertices

We will give explicitly constants  $(d_i, m_i, n_i)$  for all possible colorings on full graph with four vertices.

Every subgroup of the symmetric group  $\mathfrak{S}_4$  is conjugate to one of the groups  $\Gamma_k$ ,  $k = 1, \dots, 11$  given below. Namely, if  $\Gamma$  is a subgroup of  $\mathfrak{S}_4$ , then there exists  $k \in \{1, \dots, 11\}$  and  $\sigma \in \mathfrak{S}_4$  such that

$$\Gamma = \sigma \Gamma_k \sigma^{-1}.$$

We write  $\langle \sigma_1, \sigma_2, \dots, \sigma_n \rangle$  for the group generated by  $\sigma_1, \sigma_2, \dots, \sigma_n$ .

$(i_1 i_2 \dots i_n)$  denotes the cyclic permutation

$$i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_n \rightarrow i_1.$$

1.  $\Gamma_1 = \mathfrak{S}_4,$
2.  $\Gamma_2 = \langle (123), (12)(34) \rangle = A_4,$
3.  $\Gamma_3 = \langle (1324), (12) \rangle \simeq D_4,$
4.  $\Gamma_4 = \langle (123), (12) \rangle \simeq \mathfrak{S}_3,$
5.  $\Gamma_5 = \langle (1324) \rangle \simeq \mathbb{Z}/4\mathbb{Z},$
6.  $\Gamma_6 = \langle (12)(34), (13)(24) \rangle \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z},$
7.  $\Gamma_7 = \langle (12), (34) \rangle \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z},$
8.  $\Gamma_8 = \langle (123) \rangle \simeq \mathbb{Z}/3\mathbb{Z},$
9.  $\Gamma_9 = \langle (12)(34) \rangle \simeq \mathbb{Z}/2\mathbb{Z},$
10.  $\Gamma_{10} = \langle (12) \rangle \simeq \mathbb{Z}/2\mathbb{Z},$
11.  $\Gamma_{11} = \{e\}.$

Let us number all decreasing sequences of  $(\Gamma_k)_k$ :

1.  $\Gamma_1 \supset \Gamma_2 \supset \Gamma_6 \supset \Gamma_9 \supset \Gamma_{11},$

2.  $\Gamma_1 \supset \Gamma_2 \supset \Gamma_8 \supset \Gamma_{11}$

3.  $\Gamma_1 \supset \Gamma_3 \supset \Gamma_5 \supset \Gamma_9 \supset \Gamma_{11},$

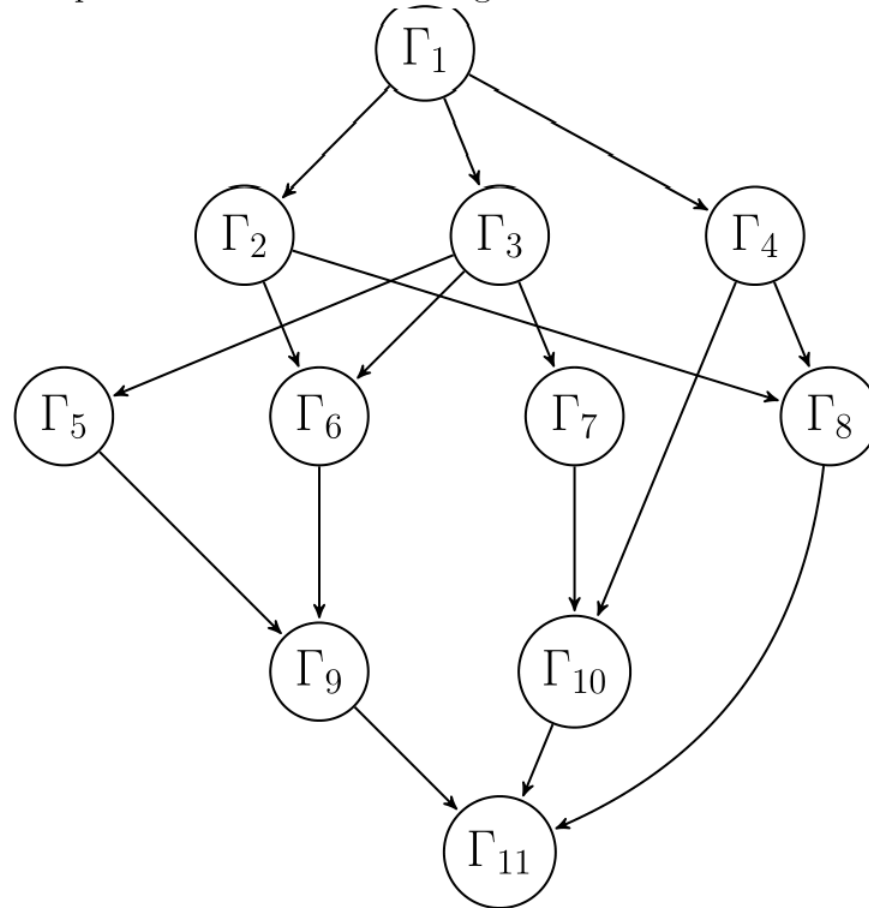
4.  $\Gamma_1 \supset \Gamma_3 \supset \Gamma_6 \supset \Gamma_9 \supset \Gamma_{11},$

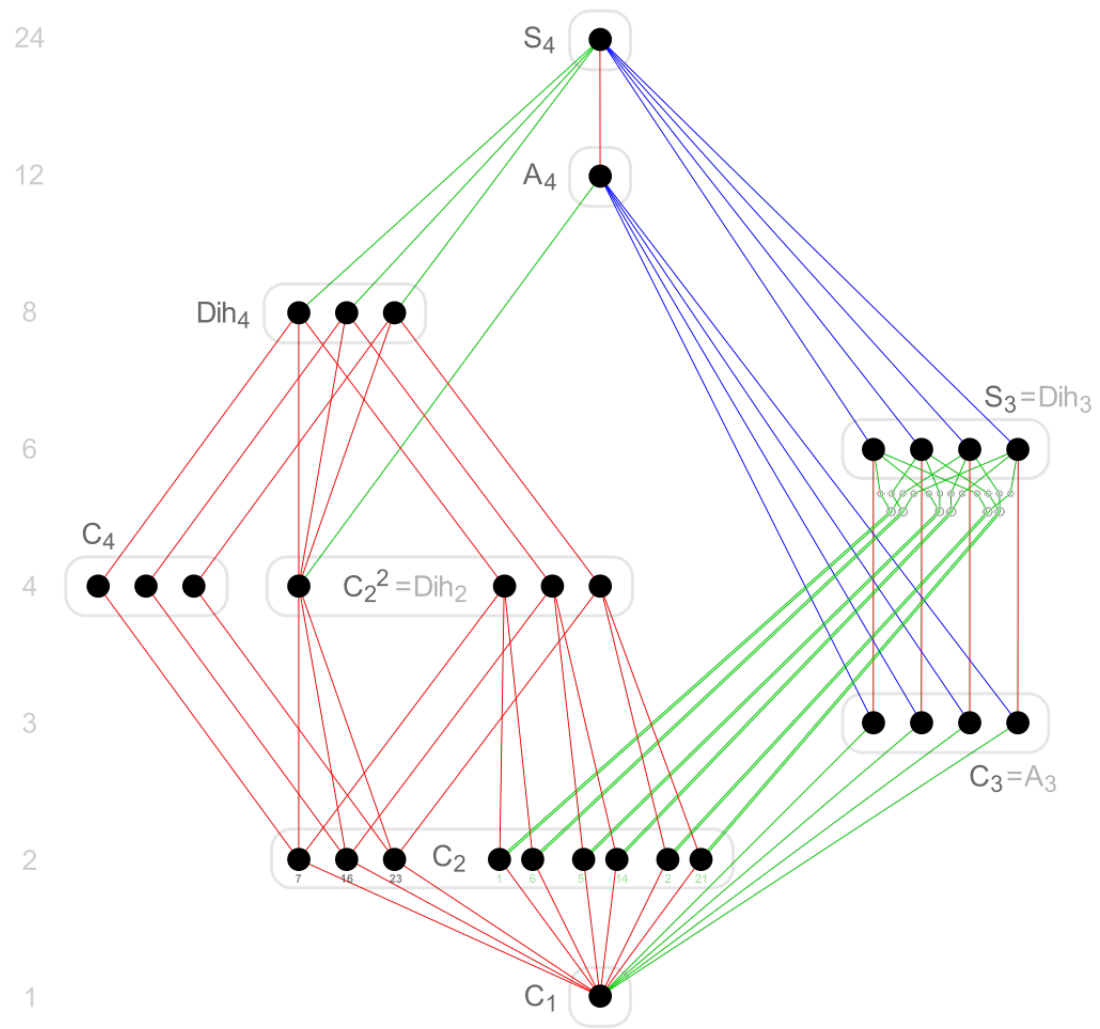
5.  $\Gamma_1 \supset \Gamma_3 \supset \Gamma_7 \supset \Gamma_{10} \supset \Gamma_{11},$

6.  $\Gamma_1 \supset \Gamma_4 \supset \Gamma_8 \supset \Gamma_{11},$

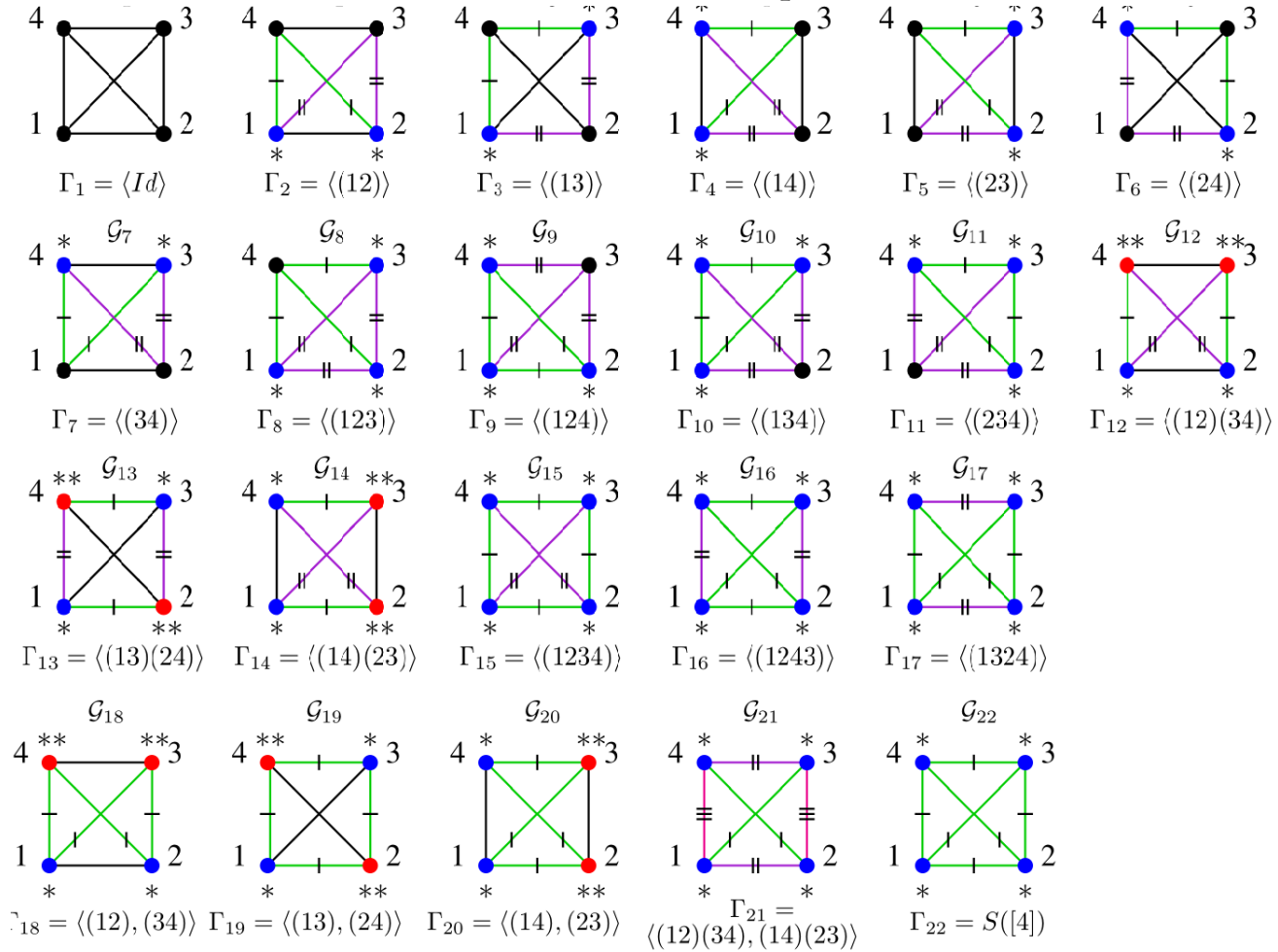
7.  $\Gamma_1 \supset \Gamma_4 \supset \Gamma_{10} \supset \Gamma_{11}.$

Below we present a Hasse diagram of the lattice of subgroups ( $\Gamma_k$ ) with the partial order relation being set inclusion.





# Gehrmann's list of full RCOPs, $p = 4$



The main goal is to compute Gamma integral over  $\mathcal{P}^\Gamma = \mathcal{Z}^\Gamma \cap \text{Sym}^+(4, \mathbb{R})$ .

We will explicit the block diagonalizations of  $\mathcal{Z}^\Gamma$ .

Let us define

$$\tilde{\mathcal{Z}}^\Gamma = \{P^{-1}XP : X \in \mathcal{Z}^\Gamma\},$$



where  $P$  is a common ON basis, chosen for 7 decreasing sequences of  $(\Gamma_k)_k$ , seen before, as follows

$$P_{(1),(3),(4)} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \end{pmatrix}$$

$$P_{(2),(6),(7)} = \begin{pmatrix} 1/2 & 1/\sqrt{12} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} & 0 \\ 1/2 & -3/\sqrt{12} & 0 & 0 \end{pmatrix}$$

$$P_{(5)} = \begin{pmatrix} 1/2 & 1/2 & 0 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 & -1/\sqrt{2} \\ 1/2 & -1/2 & 1/\sqrt{2} & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} & 0 \end{pmatrix}$$

It can be shown that

$$1. \tilde{\mathcal{Z}}^{\Gamma_1} = \{\text{diag}(A, B, B, B) : A, B \in \mathbb{R}\},$$

$$k = (1, 3), d = (1, 1), r = (1, 1),$$

$$2. \tilde{\mathcal{Z}}^{\Gamma_2} = \{\text{diag}(A, B, B, B) : A, B \in \mathbb{R}\},$$

$$k = (1, 3), d = (1, 1), r = (1, 1),$$

$$3. \tilde{\mathcal{Z}}^{\Gamma_3} = \{\text{diag}(A, B, C, C) : A, B, C \in \mathbb{R}\},$$

$$k = (1, 1, 2), d = (1, 1, 1), r = (1, 1, 1),$$

$$4. \tilde{\mathcal{Z}}^{\Gamma_4} = \left\{ \begin{pmatrix} A & C & & \\ C & B & & \\ & & D & \\ & & & D \end{pmatrix} : A, B, C, D \in \mathbb{R} \right\},$$

$$k = (1, 2), d = (1, 1), r = (2, 1),$$

5.  $\tilde{\mathcal{Z}}^{\Gamma_5} = \{\text{diag}(A, B, C, C) : A, B, C \in \mathbb{R}\}$  ,  
 $k = (1, 1, 2)$ ,  $d = (1, 1, 2)$ ,  $r = (1, 1, 1)$ ,
6.  $\tilde{\mathcal{Z}}^{\Gamma_6} = \{\text{diag}(A, B, C, D) : A, B, C, D \in \mathbb{R}\}$ ,  
 $k = (1, 1, 1, 1)$ ,  $d = (1, 1, 1, 1)$ ,  $r = (1, 1, 1, 1)$ .
7.  $\tilde{\mathcal{Z}}^{\Gamma_7} = \left\{ \begin{pmatrix} A & C & & \\ C & B & & \\ & & D & \\ & & & E \end{pmatrix} : A, B, C, D, E \in \mathbb{R} \right\}$ ,  
 $k = (1, 1, 1)$ ,  $d = (1, 1, 1)$ ,  $r = (2, 1, 1)$ ,

$$8. \tilde{\mathcal{Z}}^{\Gamma_8} = \left\{ \begin{pmatrix} A & C & & \\ C & B & & \\ & & D & \\ & & & D \end{pmatrix} : A, B, C, D \in \mathbb{R} \right\},$$

$$k = (1, 2), d = (1, 2), r = (2, 1),$$

$$9. \tilde{\mathcal{Z}}^{\Gamma_9} = \left\{ \begin{pmatrix} A & C & & \\ C & B & & \\ & & D & F \\ & & F & E \end{pmatrix} : A, B, C, D, E, F \in \mathbb{R} \right\},$$

$$k = (1, 1), d = (1, 1), r = (2, 2),$$

$$10. \tilde{\mathcal{Z}}^{\Gamma_{10}} = \left\{ \begin{pmatrix} A & E & F & \\ E & B & G & \\ F & G & C & \\ & & & D \end{pmatrix} : A, B, C, D, E, F, G \in \mathbb{R} \right\},$$

$$k = (1, 1), d = (1, 1), r = (3, 1),$$

11.  $\tilde{\mathcal{Z}}^{\Gamma_{11}} = \text{Sym}(4, \mathbb{R})$   
 $k = (1), d = (1), r = (4).$

## **Application: Frets Heads (1921)**

*Table 5.1.1* The measurements on the first and second adult sons in a sample of 25 families. (Data from Frets, 1921.)

First son		Second son	
Head length	Head breadth	Head length	Head breadth
191	155	179	145
195	149	201	152
181	148	185	149
183	153	188	149
176	144	171	142
208	157	192	152
189	150	190	149
197	159	189	152
188	152	197	159
192	150	187	151
179	158	186	148
183	147	174	147
174	150	185	152
190	159	195	157
188	151	187	158
163	137	161	130
195	155	183	158
186	153	173	148
181	145	182	146
175	140	165	137
192	154	185	152
174	143	178	147
176	139	176	143
197	167	200	158
190	163	187	150

Below we present the list of **12 best fitting models for Frets' heads, selected in former literature.**

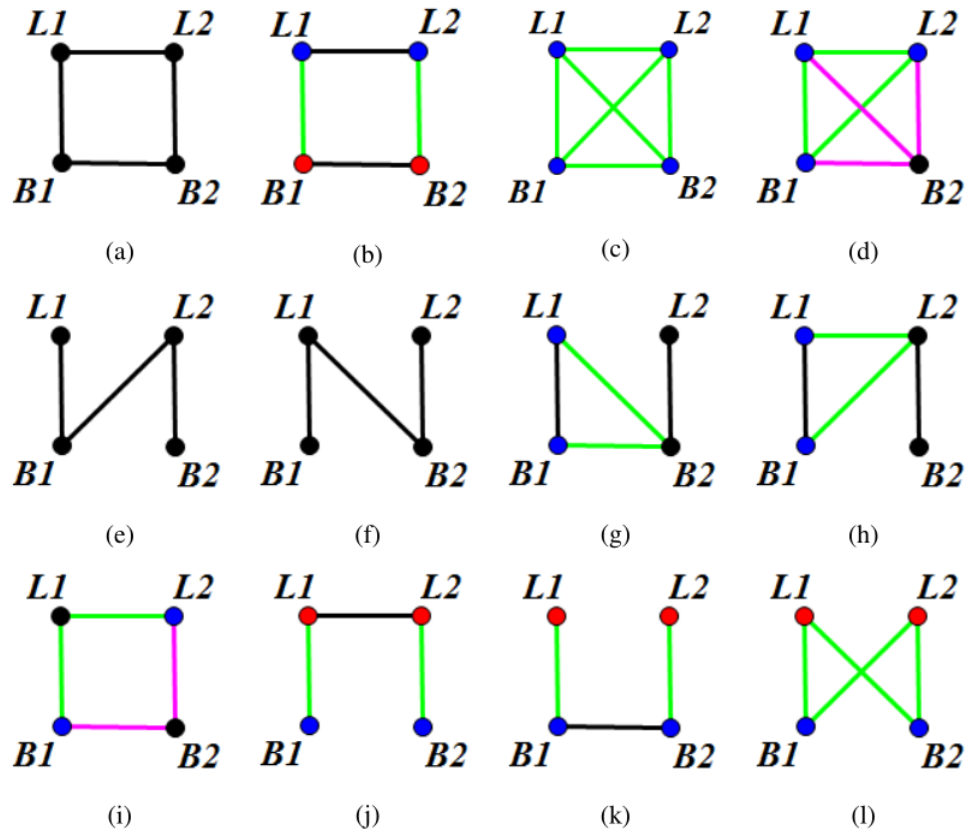


Fig. 7. Possible colored graphs supported by Fret's heads data. L1: The head length of the eldest son; B1: The head breadth of the eldest son; L2: The head length of the second son; B2: The head breadth of the second son.



Model (a) comes from Whittaker(1990) and is the smallest non-decomposable graph  $C_4$ .

Models (b,c) are selected by Højsgaard-Lauritzen(2008), by likelihood ratio test comparing each model to the saturated one.

Models (d)-(l) are the 9 minimally accepted models selected by Gehrman(2011) by Edwards-Havranek algorithm. The lowest BIC value is represented by the last graph (l).

All 12 models were considered by Massam-Li-Gao(2012) who applied to them Bayes Model selection based on prior  $G$ -Wishart distributions  $W_{s,D}$  and on an **approximative** (MC type) method of computation of constants  $I_G^\Gamma(s, D)$ .

For parameters  $s = 3$  or  $s = 10$  and for  $D = Id_4$ , the models with highest posterior probability were, respectively, (k),(b),(l) and (l),(k),(b).

## Choice of the best model among all complete RCOP models for 4 variables

The 22 complete RCOP models for 4 variables are shown on a previous slide, we use here the numbering of Gehrman.

We used as priors for  $K$  the Wishart laws  $W_{s,D}^\Gamma$  with  $s = 1$  and  $D = Id_4$

With Mathematica we obtained maximal posterior probability  $pp$  for the models:

$$\Gamma_{19} : pp = 0.4294$$

$$\Gamma_{13} : pp = 0.2273$$

$$\Gamma_8 : pp = 0.1793$$

$$\Gamma_{22} : pp = 0.0495$$

$$\Gamma_{16} : pp = 0.0329$$

All the other 17 models represent altogether  $pp = 0.08$ .

Note that the only complete RCOP model  $\Gamma_{10}$  selected by Gehrman by Edwards-Havranek algorithm among the 9 minimally accepted models (d)-(l) has  $pp = 0.0081$  which is 50 times less than  $\Gamma_{19}$ .