

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXIV

Model selection for colored graphical models for paired data

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Alberto Roverato

Dottoranda: Dung Ngoc NGUYEN

Date of submission: January 12th, 2022

Date of submission (after revision): February 28, 2022.

Abstract

A Gaussian graphical model (GGM) is a family of multivariate normal distributions whose conditional independence structure is represented by an undirected graph. The vertices of the graph represent the variables and every edge missing from the graph implies that the corresponding entry of the concentration matrix, which is the inverse of the covariance matrix, is equal to zero; see [Lauritzen \(1996\)](#).

The seminal paper by [Højsgaard and Lauritzen \(2008\)](#) introduced colored GGMs which are undirected graphical models with additional symmetry restrictions on the concentration matrix in the form of equality constraints on the parameters, which are then depicted on the dependence graph of the model by colorings of edges and vertices. The application of colored GGMs was motivated by the need of reducing the number of parameters when estimating covariance matrices of large dimensions with relatively few observations. On the other hand, there exist applied contexts where symmetry restrictions naturally follow from substantive research hypotheses of interest. A relevant instance is provided by the problem of joint learning of multiple graphical models, where the observations come from two or more groups sharing the same variables. The association structure of each group is represented by a network and it is expected that there are similarities between the groups. In this framework, the literature has mostly focused in the case where the groups correspond to independent experimental conditions so that every network is a distinct unit, disconnected from the other networks; see [Tsai *et al.* \(2021\)](#) for a review. On the other hand, in the case of paired data, the two groups are not independent because two sets of homologous variables are observed on every statistical unit. There is therefore the additional difficulty that the two groups are not independent. Hence, the respective networks need not to be disconnected ([Ranciati *et al.* 2021](#)), and there may also exist symmetries involving edges across the two groups.

In this thesis, we focus on the application of colored GGMs to the joint learning of graphical models for paired data that, in the following, we shortly call colored graphical models for paired data (PDCGMs). Although the symmetric restrictions implied by a colored GGM may usefully reduce the model dimensionality, the problem of model identification is much more challenging than with classical GGMs because both the dimensionality and the complexity of the search spaces highly increase. For the construction of efficient model selection methods it is therefore imperative to understand the structure of model classes. In this work, we consider PDCGMs and show that this class of models forms a non-distributive lattice with respect to *model inclusion order*, denoted by $\preceq_{\mathcal{C}}$. We then introduce a novel partial order, \preceq_{τ} , for this class of models and call it *the twin order*. Such order coincides with the model inclusion order if two models are model inclusion comparable but that also includes order relationships between certain models which are model inclusion incomparable. We show that the class of PDCGMs forms a complete distributive lattice with respect to the twin order and then we use this lattice to implement a coherent backward elimination stepwise model search procedure.

[Gabriel \(1969\)](#) introduced the following principle, termed coherence: “in any procedure involving multiple comparisons no hypothesis should be accepted if any hypothesis implied by it is rejected”. We remark that, for convenience, we say “accepted” instead of the more correct “non-rejected”. Consider some goodness-of-fit tests for testing models at a given level α so that for every model in a given class we can apply the test and determine whether the model is rejected or accepted. In this context, the coherence principle is typically implemented by requiring that we should not accept a model while rejecting a more general model; see, for instance, [Edwards and Havránek \(1987\)](#). Hence, under this formulation of the coherence principle, in a greedy search procedure if a model is rejected then all of its submodels are considered rejected without further testing, and the lattice with respect to model inclusion represents the natural framework to implement the procedure. However, we show that the model inclusion lattice for PDCGMs does not provide a proper structure of the coherence principle. On the other hand, the coherence principle can be properly implemented on our distributive lattice under the twin order that seems to represent the natural structure for the implementation of stepwise backward elimination procedures for PDCGMs. We therefore introduce a backward elimination stepwise procedure with local moves on our distributive lattice that satisfies the coherence principle. This procedure is implemented in the statistical programming language R (see [R Core Team 2021](#)) and its behavior is investigated on simulated data.

Finally, this procedure is applied to the identification of the brain network from fMRI data.

Sommario

Un modello grafico gaussiano (GGM) è una famiglia di distribuzioni normali multivariate la cui struttura di indipendenza condizionale viene rappresentata mediante un grafo non orientato. I vertici del grafo corrispondono alle variabili ed ogni arco assente dal grafo implica che il corrispondente elemento della matrice di concentrazione, ossia l'inversa della matrice di varianze e covarianze, è uguale a zero; si veda [Lauritzen \(1996\)](#).

I modelli grafici colorati, introdotti da [Højsgaard and Lauritzen \(2008\)](#), sono una famiglia di modelli grafici gaussiani con ulteriori vincoli di simmetria implementati come vincoli di uguaglianza negli elementi della matrice di concentrazione. Questi modelli sono rappresentati mediante grafi con vertici ed archi colorati assegnando, in modo opportuno, colore uguale ad archi e vertici tra loro “simmetrici”. L'utilizzo dei modelli grafici colorati fu motivato inizialmente dalla necessità di ridurre il numero di parametri nell'apprendimento di grafi con elevato numero di vertici in presenza di una limitata numerosità campionaria. Vi sono però contesti applicativi nei quali i vincoli di simmetria emergono naturalmente come quesiti scientifici di interesse. Un esempio rilevante è dato dall'apprendimento congiunto di network multipli nei quali le osservazioni provengono da due o più gruppi che condividono lo stesso insieme di variabili. In questo contesto, la ricerca si è focalizzata principalmente nel caso in cui i gruppi sono associati a contesti sperimentali indipendenti tra loro e quindi un network è associato ad ogni gruppo e i vari network sono sconnessi tra loro; si veda [Tsai *et al.* \(2021\)](#) per una rassegna. Tuttavia, nel caso di dati appaiati i due gruppi di interesse non sono indipendenti tra loro perché i due insiemi di variabili omologhe sono misurate sulla stessa unità osservata. Ne segue che, in questo caso, i due network non sono sconnessi tra loro ([Rancati *et al.* 2021](#)) e vi possono essere anche simmetrie che coinvolgono archi a cavallo tra i due network.

Oggetto di questa tesi è l'applicazione di modelli grafici colorati all'apprendimento congiunto di modelli grafici per dati appaiati. Sebbene i vincoli di simmetria implicino naturalmente una riduzione della dimensionalità del modello, il problema dell'apprendimento del modello dai dati è estremamente complesso dato che la dimensione dello spazio di ricerca è molto maggiore rispetto a quella dei tradizionali modelli grafici non colorati. Per la costruzione di procedure di ricerca che siano efficienti è fondamentale comprendere la struttura dello spazio di ricerca. In questo lavoro noi consideriamo i modelli grafici colorati per dati appaiati (PDCGM) e mostriamo che se si utilizza il tradizionale ordinamento basato sulla relazione di sottomodello (ordinamento *model inclusion*), questa famiglia forma un reticolo non-distributivo. Introduciamo quindi una nuova relazione ordine, che chiamiamo ordinamento *twin* che coincide con l'ordinamento *model inclusion* quando i due modelli considerati sono uno un sottomodello dell'altro, ma che introduce anche delle relazioni di ordine addizionali tra modelli che sono non comparabili in termini di *model inclusion*. Mostriamo quindi che la famiglia di PDCGM forma un reticolo distributivo rispetto all'ordinamento *twin* e quindi utilizziamo questa struttura per introdurre una procedura di apprendimento di tipo *stepwise*.

[Gabriel \(1969\)](#) ha introdotto il seguente principio detto principio di coerenza “in una procedura in cui vengono verificate ipotesi multiple, una qualunque ipotesi non dovrebbe essere accettata quando, al contempo, un'ipotesi implicata da questa viene rifiutata”. Si noti che, per brevità, in questa formulazione utilizziamo il termine “accettata” invece del termine più rigoroso “non-rifiutata”. Si consideri un test di livello α che può essere applicato al confronto di modelli in una procedura di apprendimento. In questo contesto, il principio di coerenza viene solitamente applicato richiedendo che non si deve accettare un qualunque modello quando un modello più generale è rifiutato; si veda, ad esempio, [Edwards and Havránek \(1987\)](#). Quindi, nell'implementazione di una procedura *stepwise backward elimination* coerente se un modello è rifiutato allora tutti i suoi sottomodelli sono automaticamente rifiutati senza necessità di confronto esplicito, e il reticolo ottenuto dalla relazione d'ordine *model inclusion* sembra essere la struttura più naturale per l'implementazione di questo tipo di procedura. Tuttavia, noi mostriamo che per la famiglia di modelli grafici colorati per dati appaiati l'applicazione del principio di coerenza richiede ragionamenti più sofisticati e che l'applicazione automatica di questo principio sulla base del reticolo *model inclusion* porta ad effettuare dei passi che violano il principio di coerenza. Invece, il reticolo basato sulla relazione *twin* permette di identificare tali passi non coerenti e sostituirli con dei passi che rispettano il principio di coerenza. Questa variazione conferisce inoltre efficienza alla procedura. La procedura è implementata nel linguaggio R ([veda R Core Team 2021](#)), le sue proprietà

sono illustrate mediante una serie di applicazioni a dati simulati ed, infine, utilizzata per l'identificazione di un *brain network* sulla base di dati fMRI.

Dedication

It is with genuine gratitude and warm regard that I dedicate this work to my family members and friends.

Acknowledgements

This thesis could not have been achieved without the support and help of my advisor, Ph.D. committee, colleagues, family, and friends whom I wish to thank wholeheartedly.

First and foremost, I would like to express my deepest appreciation to my supervisor Professor **Alberto Roverato**: “Without your dedicated assistance and immense expertise in formulating the research questions and methodologies, this thesis would have never been accomplished. I would like to thank you very much for your constant support and sympathetic understanding as well as plentiful experiences and sincere comments over these past years. You are very patient with my immaturity and greenness in research and steered me in the right direction whenever I got lost in my studies, or I ran into a trouble spot about my research or writing. I gratefully acknowledge all of the time you have discussed ideas, sometimes my crazy thoughts, and all of the effort that you have put into fixing my weaknesses, especially my bad grammar in writing and academic explanations. Your approach to research and science is reflected by your simple but clear writing style, which is something I hope to carry forward throughout my career. You have shared the pressure in every step throughout the process and accompanied me to overcome the challenges with rock steady support, especially in the final stretch. Thank you for all of the opportunities that you give me to further my research path”.

I wish to express my special thanks to Professor **Romain Abraham** for providing me the valuable master internship at Institut Denis Poisson, Université d’Orléans and supporting me during the time I looked for Ph.D. scholarships. I am also deeply indebted to Professor **Dang Duc Trong** for guiding me the bachelor thesis and orienting me on my academic career. I am also grateful to Professor **Xavier Gendre** who has motivated and supported me to pursue the doctoral program with a perfect blend of insight and humor short courses in Mathematics and Data Sciences.

Furthermore, I would like to extend my sincere thanks to **University of Padova** for organising a flexible and professional academic environment for people who want to continue studying and developing for the research future, and especially, awarding me a doctoral fellowship with the financial means. Particularly, I would like to express gratitude to Ph.D. committee, professors, researchers and lecturers in **Department of Statistical Sciences**. I had great opportunities and experiences to solidify my background in statistics as well as develop deeper knowledge and skills, from theory to practice, through

foundation courses and some other specialized short courses. Thanks for offering advice, support, and giving me wonderful experiences from expertise lessons to research activities under the guidance, teaching and training of ideal and dedicated professors and lecturers. Thanks should go to other staff members in the department and university for supporting administrative and accounting procedures, and problems related to information technology and general services during the last three years. Additionally, many thanks to my colleagues: “All of you have been unwavering in personal and professional support during the time I spent at the university. You are willing to give me honest suggestions and frank comments that helped me make some sense of the confusion, extend myself in teamwork and be bolder when asking questions. For many memorable lunch and evenings out and in, you gave me great experiences about Italian culture and showed me that you are really warm, wholehearted, intelligent, good-humored, and idealistic. I will miss the glasses of spritz and wine, the dinners at Peace’n Spice and sushi restaurants that we spent together, the football matches we cheered for the Italian team in EURO Championship, and other great things. Thank you”.

I also wish to thank **Dr. Duc Khanh To** and **Dr. Thi Kim Hue Nguyen** who helped me in the first days I went to Italy. Especially, Dr. Khanh introduced me about Ph.D. program in University of Padova and supported me about the application for the program as well as legal administrative paperwork when I arrived in Italy. Dr. Hue gave me valuable and useful advice and information in both academic and personal aspects, and provided me the caring and loving with the warm meals and outing.

Most importantly, I must express my very profound gratitude to members in my family, especially my parents who gave birth to me in the first place, offered sincere encouragement through phone calls every week, and supported me spiritually with wholehearted love throughout my life. Special thanks to my boyfriend **Dr. Trung Tin Nguyen**: “It is your discipline, effort, and passion in doing research that inspire and impact me throughout years of study and through the process of researching. Your insightful feedback and valuable discussion pushed me to sharpen my thinking and brought my work to a higher level. Your companionship and love are extremely precious to me. Thank you”.

Last but not least, the acknowledgments are the place to express gratitude to my friends and others who always stood behind me and gave me moments of warmth through phone calls and messages, for calming me down by sincere encouragement.

Padova, January 12, 2022

Dung Ngoc Nguyen.

Contents

List of Figures	xix
List of Tables	xxiii
Introduction	3
Overview	3
Main contributions of the thesis	4
1 Background	7
1.1 Partial orders and lattices	7
1.1.1 Partially ordered sets and partial orders	7
1.1.2 Hasse diagram	9
1.1.3 Lattices	10
1.1.4 The set inclusion lattice	12
1.1.5 The set partition lattice	13
1.2 Graphs, color classes and colored graphs	14
1.2.1 Undirected graphs	14
1.2.2 The lattice structure of undirected graphs	16
1.2.3 Color classes and colored graphs	17
1.2.4 The lattice structure of colored graphs	18
1.3 Covariance matrices, concentration matrices and concentration graphs	20
1.4 Graphical models	24
1.4.1 Multivariate Gaussian distribution	24
1.4.2 Gaussian graphical models	26
1.4.3 RCON models: equality restrictions on <u>concentration</u>	27
1.5 Structure of model spaces	29
1.5.1 Undirected graphical model space	30
1.5.2 Colored graphical model space	30
2 Colored graphs and colored graphical models for paired data	33
2.1 Colored graphical models for paired data	33
2.2 Colored graphs for paired data	34
2.3 The set partition lattice of PDCGs	36
2.3.1 Colored graphs for paired data on a given fixed uncolored graph	36
2.3.2 General case for colored graphs for paired data	41

2.4	The novel partial order for the set of PDCGs and its properties	43
2.4.1	Notations and terminologies of important quantities	43
2.4.2	Associated lattice structure for the set of PDCGs and its structural properties	51
3	Model search over the twin lattice	57
3.1	Dimensions of the search spaces and related works	57
3.2	Principle of coherence	61
3.3	Greedy search on the twin lattice	63
3.3.1	The backward elimination stepwise procedure on the twin lattice	64
3.3.2	R implementation	69
3.4	Greedy search on the model inclusion lattice	71
3.4.1	The backward elimination stepwise procedure on the model inclusion lattice	71
3.4.2	R implementation	72
3.5	Comparison and implementation details	73
3.5.1	Numerical simulations	73
3.5.2	Application to fMRI data	80
3.6	Other approaches	88
3.6.1	Forward inclusion stepwise procedure on the twin lattice	88
3.6.2	Penalized likelihood approach	89
	Concluding remarks	92
	Bibliography	95

List of Figures

1.1	Graphical representations for the division relation $ $ on the set $\{3, 18, 30, 45, 90\}$ by (a) the directed graph, and (b) Hasse diagram.	10
1.2	Hasse diagram of a complete poset that is not a lattice.	10
1.3	Graphical representations for lattices containing (a) the diamond structure, and (b) the pentagon structure.	11
1.4	Hasse diagram of the lattice $\langle \mathcal{P}(\{1, 2, 3\}), \subseteq \rangle$	12
1.5	Hasse diagram of the lattice of partitions of a set A of three distinct elements ordered by the refinement \leq . A red colored region indicates a subset of A that forms a member of the enclosing partition. Uncolored dots indicate single-element subsets or atomic subsets.	14
1.6	Example of (a) an undirected graph, (b) a complete graph, and (c) an empty graph. The three graphs here use a common vertex set $V = [4]$	15
1.7	An undirected graph $G = (V, E)$ with the vertex set $V = [5]$	15
1.8	Example of the partial order between two undirected graphs G and H in $U_{[4]}$	16
1.9	A colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ based on the uncolored undirected graph $G = (V, E)$ shown in Figure 1.6a.	17
1.10	Example of the partial order between two colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{C}_{[4]}$	19
1.11	Colored graphs represent (a) the <i>zero</i> and (b) the <i>unit</i> in the lattice $\langle \mathcal{C}_{[4]}, \preceq_c \rangle$	20
2.2	A colored graph for paired data on the vertex set $V = [6]$ where V is partitioned into $L = \{1, 2, 3\}$ and $R = \{4, 5, 6\}$. Every symmetric twin vertices or every symmetric twin edges are depicted in the same color, and marked by the same symbol. The black color indicates vertices or edges which stand alone in the color classes.	35
2.3	Hasse diagram of the lattice structure of the set of colored graphs for paired data, \mathcal{S}_G , based on the given fixed uncolored graph G , where $G = ([4], E(G))$ is represented at the top.	40
2.4	Representations of (b) the zero and (c) the unit on the lattice of PDCGs based on (a) the given fixed uncolored graph G , $\langle \mathcal{S}_G, \preceq_c \rangle$	41
2.5	A counterexample of the non-distributivity of PDCGs \mathcal{G}, \mathcal{H} , and \mathcal{K} on the lattice $\langle \mathcal{S}_{[6]}, \preceq_c \rangle$	42
2.6	The diamond structure of the sublattice of $\langle \mathcal{S}_{[4]}, \preceq_c \rangle$	42
2.7	The vertex set $V = [6]$ is split into two disjoint subsets $L = \{1, 2, 3\}$ and $R = \{4, 5, 6\}$	43

2.8	Relationship between the set L and the set R is presented by the twin function.	44
2.9	Pairs of twin edges connecting vertices within L and R in graph with 6 vertices, namely, (a) $(1, 2), (1', 2')$, (b) $(2, 3), (2', 3')$, and (c) $(1, 3), (1', 3')$	45
2.10	Pairs of twin edges connecting vertices across L and R in the graph with 6 vertices, namely, (a) $(1, 2'), (2, 1')$, (b) $(1, 3'), (3, 1')$, and (c) $(2, 3'), (3, 2')$	45
2.11	Examples of pairs of edges which are not twin edges, namely, (a) $(1, 2), (1', 3')$, (b) $(1, 2), (1, 2')$, and (c) $(1, 2'), (3, 1')$	46
2.12	Relationship between the set F_L and the set F_R is presented by the twin function.	47
2.13	The partition of $F_{[6]}$ into (a) F_L , (b) F_R and (c) F_T	48
2.14	An undirected graph $G = (V, E)$	49
2.16	A colored graph for paired data $\mathcal{G} = (\mathcal{V}, \mathcal{E})$	51
2.17	Example of the partial order \preceq_τ between two colored graphs for paired data \mathcal{G} and \mathcal{H} in $\mathcal{S}_{[4]}$	52
2.18	Sublattices of colored graphs for paired data in \mathcal{S}_V ordered by (a) the twin order \preceq_τ and (b) the model inclusion order \preceq_c	54
2.19	Colored graphs represent (a) the zero, and (b) the unit in $\langle \mathcal{S}_{[4]}, \preceq_\tau \rangle$	54
3.1	Perfectly-paired uncolored graphs with (a)-(b) one pair of homologous edges, and (c) two pairs of homologous edges on the vertex set of 4 vertices.	58
3.2	Logarithmic scale of the number of “perfectly-paired” (uncolored) graphical models (illustrated in black) and the number of “perfectly-paired” colored graphical models for paired data (illustrated in blue) with different even number of vertices p from 2 to 20.	59
3.3	Logarithmic scale of numbers of colored graphical models for paired data represented by the complete uncolored graphs (illustrated in red), and the “perfectly-paired” uncolored graphs (illustrated in blue), with different even number of vertices p from 2 to 20.	60
3.4	Sublattice structure of colored graphs for paired data with the set of 4 vertices under the model inclusion \preceq_c	62
3.5	Sublattice structure of colored graphs for paired data with the set of 4 vertices under the twin order \preceq_τ	63
3.6	Example of how to split the set of neighbors of a certain colored graphical model for paired data \mathcal{M}^* in $\mathcal{S}_{[6]}$ into two layers where the nearest neighbors of \mathcal{M}^* are models belonging to the red shaded area.	68
3.7	The set of neighbor submodels of the saturated model with 4 vertices is split into two layers by the twin order \preceq_τ where models in the upper layer and the lower layer can be found at Steps 4.a and 5.a.	70
3.8	The set of neighbor submodels of the saturated model with 4 vertices by the model inclusion order \preceq_c	73
3.9	Averaged seconds on elapsed time from the backward elimination procedures based on the twin order \preceq_τ (illustrated in red) and the model inclusion \preceq_c (illustrated in blue) of two scenarios A (on the left) and B (on the right).	77

3.10	Averaged number of fitted models from the backward elimination procedures based on the twin order \preceq_τ (illustrated in red) and the model inclusion \preceq_c (illustrated in blue) of two scenarios A (on the left) and B (on the right).	77
3.11	The analyzed colored graphs of the PDCGM for the subject 14 obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_τ , with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. Figures (a) and (b) present asymmetric edges “on the left” and “on the right”, respectively. Figures (c) and (d) present symmetric twin edges between and across two hemispheres, respectively.	86
3.12	The analyzed colored graphs of the PDCGM for the subject 15 obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_τ , with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. Figures (a) and (b) present asymmetric edges “on the left” and “on the right”, respectively. Figures (c) and (d) present symmetric twin edges between and across two hemispheres, respectively.	87
3.13	Sublattice structure of colored graphs for paired data with the set of 4 vertices under the twin order \preceq_τ	89

List of Tables

3.1	The number of symmetric twin vertices and edges for scenaria A and B for different numbers of vertices p from 8 to 20.	74
3.2	Generated colored graphs for paired data for scenarios A and B (from left to right) with different number of vertices p (from top to bottom).	75
3.3	Performance measures of the model selection procedure for the lattice structure equipped by the partial orders \preceq_τ and \preceq_C . Results are recorded as mean (and standard deviation) computed across the 20 replicated datasets for each scenario. Here, #edges and #sym are computed as the average number of edges, average number of pairs symmetric edges, and their standard deviations are rounded, while Time is on the average of seconds.	79
3.4	Graphical representation of the colored graphical models for paired data obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_τ , proposed in Section 3.3, with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. From left to right: anterior temporal lobe and the frontal lobe. From top to bottom: subject 14 and subject 15.	83
3.5	Graphical representation of the colored graphical models for paired data obtained from the application of the stepwise backward elimination procedure based on the model inclusion \preceq_C , described in Section 3.4, with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. From left to right: anterior temporal lobe and the frontal lobe. From top to bottom: subject 14 and subject 15.	84
3.6	Recorded results from the application to fMRI data of the backward elimination stepwise procedure on the twin lattice, proposed in Section 3.3, and similar approach on the model inclusion lattice, described in Section 3.4.	85
3.7	Recorded results from the application to fMRI data of the graphical glasso (glasso) approach.	91

Introduction

Overview

Chapter 1: Background

This thesis deals with colored Gaussian graphical models and, more specifically, with the lattice structure which can be used to represent the model spaces formed by these models under a particular partial order. This chapter provides the required background, which is designed as follows. Firstly, we briefly overview the theory of partially ordered sets and lattices. We use Hasse diagrams to represent them diagrammatically and consider two relevant cases given by the set inclusion lattice and the set partition lattice. Secondly, we introduce undirected graphs and colored graphs; moreover, the associated lattice structures of these graph spaces are also described. Thirdly, we review the literature of graphical models and present the associated model structures for the families of uncolored undirected graphical models and colored graphical models.

Chapter 2: Colored graphs and colored graphical models for paired data

This chapter is devoted to the analysis of the lattice structures of the search spaces. Particularly, we identify the family of colored graphical models for paired data as a subfamily of the general colored graphical models. Accordingly, we define colored graphs for paired data and investigate the structure of the set of these graphs. The structure of Chapter 2 is designed as follows. We consider graphical models for paired data and, in the first two sections of this chapter, we formally define this specific class of models. The family of graphical models for paired data is associated with the family of colored graphs for paired data, and we then provide a comprehensive analysis of the structure of the family of such graphs. More concretely, in the third section, we firstly consider the model inclusion order, as in [Gehrmann \(2011\)](#), and show that colored graphs for paired

data form a complete non-distributive lattice. On the other hand, we also identify a relevant sublattice of the colored graphs for paired data based on a given fixed uncolored graph that is complete and distributive under the model inclusion order.

The main result of this chapter is the introduction of a novel order that we call the *twin order*, shown in the fourth section. The connections of this novel order with the traditional model inclusion order are also described. Furthermore, we show that, under the twin order, the family of colored graphs for paired data forms a complete distributive lattice. The *meet* and the *join* operations on this lattice can be efficiently computed because they coincide with the set intersection and the set union operations.

Chapter 3: Model search over the twin lattice

The selection of colored graphical models from data is a major challenge because the number of different models grows super-exponentially with the number of variables. In this chapter, we focus on greedy search methods which perform local moves on the lattice structure of model space. In this framework, it is crucial to deal efficiently with the lattice structure. Thus, we implement a backward elimination stepwise procedure on the twin lattice and show that it is more efficient than an equivalent procedure on the model inclusion lattice. Furthermore, we show that the use of the twin lattice allows us to avoid an incoherent step in the model search. The performance of the proposed procedure is evaluated on simulated data and applied to the identification of a brain network from functional MRI data.

Main contributions of the thesis

The novel contributions in this thesis are given in Chapter 2 and Chapter 3.

The results of Chapter 2 are theoretical and concern the structure and properties of the search spaces, and in more detail,

1. we formally define the family of colored graphs for paired data and the associated family of colored graphical models for paired data;
2. we consider the model inclusion order and show that the search space of colored graphs for paired data forms a complete non-distributive lattice; moreover, the sublattice obtained by considering the subset of the colored graphs for paired data with the fixed uncolored graph forms a complete distributive lattice;
3. we introduce a novel partial order for the family of colored graphs for paired data and call it the *twin order*;

4. we show the connections existing between the model inclusion order and the twin order;
5. we show that, under the twin order, the family of colored graphs for paired data forms a complete distributive lattice where the *meet* and the *join* operations on such lattice are efficiently specified because they coincide to the set intersection and the set union operations.

The results of Chapter 3 concern the construction of a backward elimination stepwise procedure on the twin lattice that is implemented in the R language. In more detail,

1. we consider the principle of coherence and show that a backward elimination stepwise procedure on the model inclusion lattice includes an incoherent step;
2. we specify a backward elimination stepwise procedure on the lattice based on the novel twin order that avoids the incoherent step;
3. the novel procedure based on the twin order is shown to be more efficient than a similar approach on the model inclusion lattice from the simulation experiments and the application to the fMRI data;
4. we implement the model selection procedure for PDCGMs under the twin order on simulated data and real data by the scripts written in the R programming language, found at <https://github.com/NgocDung-NGUYEN/backwardCGM-PD>.

Chapter 1

Background

The aim of this chapter is threefold. In the first part, we provide a brief overview of the theory of partially ordered sets and lattices, which play a major role in the description of the structure of model spaces. We also learn how to represent them diagrammatically and consider two relevant specific cases given by the set inclusion and the set partition lattices. In the second part, we introduce undirected graphs and colored graphs. Moreover, the lattice structures associated with these graph spaces are also described. Finally, we shall give some basic views of the literature on graphical models and present the associated model structures for the families of undirected graphical models and colored graphical models.

1.1 Partial orders and lattices

In this section, we introduce partial orders and their graphical representation by means of an Hasse diagram. Then, we describe lattices and consider two relevant cases: the set inclusion and the set partition. For a more detailed account of partial orders and lattices, see [Birkhoff \(1940\)](#), [Schechter \(1996, Chapter 3\)](#), and [Grätzer \(2002\)](#).

1.1.1 Partially ordered sets and partial orders

The *Cartesian product* of two sets A and B is the set of all possible ordered pairs (a, b) consisting of elements a in A and b in B ; more specifically, it is defined by

$$A \times B = \{(a, b) \mid a \in A, b \in B\}. \quad (1.1)$$

The term *relation* is used as a mathematical structure to trace the relationship between the elements of two or more sets, or between the elements on the same set. A *binary relation* over sets A and B is a subset of the Cartesian product $A \times B$.

Example 1.1. Let $A = \{1, 2, 3\}$ be a set. Then, the Cartesian product of the set $A \times A$ is given by

$$A \times A = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}.$$

The relation “greater than”, denoted by $>$, on the set $A = \{1, 2, 3\}$ is the binary relation over $A \times A$ given by $\{(3, 1), (3, 2), (2, 1)\} \subseteq A \times A$.

Let A be a non-empty set. We consider a binary relation \preceq so that for a pair of elements $a, b \in A$, if $a \preceq b$ holds then we say that a is *related* to b . Moreover, the set A is called a *partially ordered set* or *poset* and \preceq a *partial ordering relation* if \preceq is

- reflexive: $a \preceq a$;
- antisymmetric: $a \preceq b$ and $b \preceq a$ implies $a = b$;
- transitive: $a \preceq b$ and $b \preceq c$ implies $a \preceq c$,

for all $a, b, c \in A$. We denote a poset A with partial ordering relation \preceq by $\langle A, \preceq \rangle$.

Example 1.2. We consider the divisibility relation, i.e. $\{(a, b) \mid b \text{ is a divisor of } a\}$, denoted by $|$, on the set $A = \{3, 18, 30, 45, 90\}$. Then, the set A equipped by $|$ is a poset and we can write $\langle A, | \rangle$.

Let $\langle A, \preceq \rangle$ be a poset and $a, b \in A$. We say $a \prec b$ if $a \preceq b$ and $a \neq b$. When it holds that neither $a \preceq b$ nor $b \preceq a$ then a and b are *incomparable*. Furthermore, a is *covered* by b (or b *covers* a), denoted by $a \prec b$, if $a \prec b$ and there is no element $c \in A$ such that $a \prec c \prec b$.

Example 1.2 (continued). Continue the previous example of the poset $\langle A, | \rangle$, the elements 18 and 30 are incomparable by the order $|$, however, both 18 and 30 cover 3.

In addition, let H a subset of A , then $a \in A$ is an *upper bound* of H if $h \preceq a$ for all $h \in H$. We call a the *least upper bound* or *supremum* of H if every upper bound b of H satisfies $a \preceq b$ and hence we write $a = \sup H$. Similarly, $a \in A$ is a *lower bound* of H if $a \preceq h$ for all $h \in H$. We call a the *greatest lower bound* or *infimum* of H if every lower bound b of H satisfies $b \preceq a$ and then we write $a = \inf H$. Remark that, the *supremum* (or *infimum*) of a set is unique if it exists.

An element a in a poset $\langle A, \preceq \rangle$ is called *minimal* if it is not greater than any other element in A . If there is one unique minimal element a , we call it the *minimum element* or the *smallest element*. Similarly, an element a is called the *maximal element* if it is not smaller than any other element in A . If the maximal element is unique, we call it the *maximum element* or the *largest element*.

It is a convention that $\sup \emptyset$ is the smallest element in A , called *zero*, if it exists, and $\inf \emptyset$ is the largest element in A , called *unit*, if it exists. To shorten the notations, we write $\hat{0}$ to denote the zero and $\hat{1}$ to denote the unit.

Example 1.2 (continued). *The zero and the unit in the poset $\langle A, | \rangle$ are determined as $\hat{0} = 3$ and $\hat{1} = 90$.*

1.1.2 Hasse diagram

It is useful to represent a partial order graphically. Since a partial order is a binary relation, it can be represented by means of a directed graph where the vertices are elements of A and there is a directed edge from a to b if a is related to b through the partial order \preceq , i.e. $a \preceq b$, for any elements a, b in A .

When we deal with a partial order, we know that the relation is reflexive, transitive, and antisymmetric. This allows us to simplify the graphical representation of a partially ordered set $\langle A, \preceq \rangle$ by taking the following steps:

- **(reflexivity)** remove all edges from a vertex to itself,
- **(transitivity)** remove all transitive edges, i.e. if $a \preceq b \preceq c$, we delete edge from a to c but retain the other two edges from a to b and from b to c ,
- **(antisymmetry)** if $a \preceq b$ then the vertex b appears above the vertex a , hence we remove directions of edges assuming that they are oriented upwards

for $a, b, c \in A$. The resulting graph looks far simpler and it is called a *Hasse diagram*.

Example 1.2 (continued). *Let's continue with the consideration of the poset $\langle A, | \rangle$. The directed graph corresponding to this relation is shown in Figure 1.1a, on the other hand, Figure 1.1b displays the corresponding Hasse diagram. As it can be seen, the Hasse diagram is a useful tool which completely describes the associated partial order.*

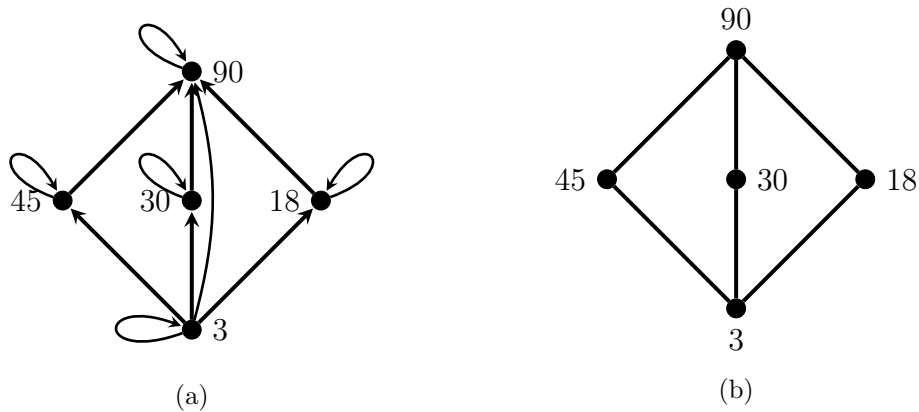


FIGURE 1.1: Graphical representations for the division relation $|$ on the set $\{3, 18, 30, 45, 90\}$ by (a) the directed graph, and (b) Hasse diagram.

1.1.3 Lattices

Many important properties of an ordered set A are expressed in terms of the existence of certain upper bounds or lower bounds of subsets of A . One of the most important classes of ordered sets defined in this way is lattices introduced in Birkhoff (1940); see also Grätzer (2002), and Davey and Priestley (2002). In particular, a poset $\langle A, \preceq \rangle$ is a *lattice* if $\inf H$ and $\sup H$ exist for every finite non-empty subset H of A . Moreover, a poset $\langle A, \preceq \rangle$ is called *complete* if $\inf H$ and $\sup H$ also exist for $H = \emptyset$.

Example 1.3. Figure 1.2 shows a complete poset; however, this poset is not a lattice since c and d are upper bounds of $\{a, b\}$ but c and d are not comparable.

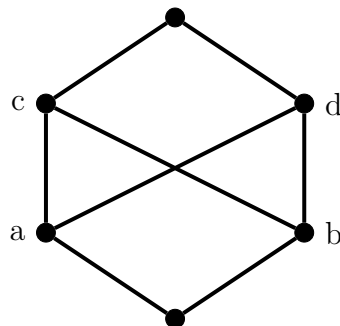


FIGURE 1.2: Hasse diagram of a complete poset that is not a lattice.

For a lattice A and $a, b \in A$, we might write

$$a \wedge b = \inf\{a, b\} \quad \text{and} \quad a \vee b = \sup\{a, b\},$$

and refer to \wedge as the *meet operation* and to \vee as the *join operation*. Furthermore, a non-empty subset H of a lattice A is a *sublattice* of A if

$$a, b \in H \implies a \wedge b \in H \text{ and } a \vee b \in H.$$

A lattice A is *distributive* if for all $a, b, c \in A$,

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c). \quad (1.2)$$

We can understand that a distributive lattice is a lattice in which the join operation \vee and the meet operation \wedge distribute over each other, i.e. the distributivity law is satisfied by the meet and the join operations. Distributivity is a fundamental property that facilitates the implementation of efficient procedures and representation in lattices; see, among others, [Habib *et al.* \(2001\)](#), [Davey and Priestley \(2002\)](#), [Munro and Sinnamon \(2018\)](#). The Hasse diagram of Figure 1.1b is called a “diamond”, because of its structure. The following theorem shows that a distributive lattice cannot contain a diamond as a sublattice.

Theorem 1.1 ([Davey and Priestley \(2002, Chapter 4.10\)](#)). *Let A be a lattice. A is not distributive if and only if A has a sublattice isomorphic to the diamond or the pentagon as given in Figure 1.3.*

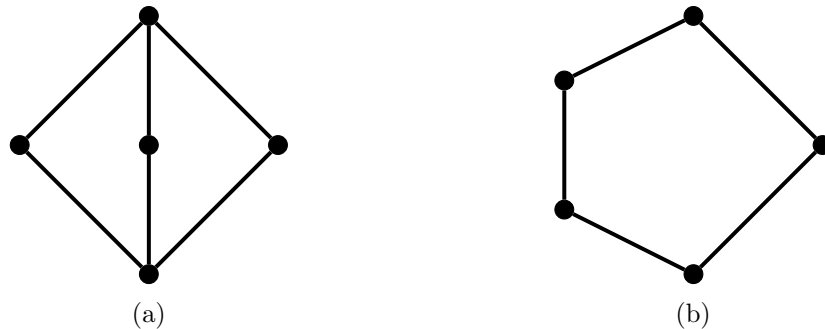


FIGURE 1.3: Graphical representations for lattices containing (a) the diamond structure, and (b) the pentagon structure.

Example 1.2 (continued). *Consider the poset represented in Figure 1.1b. It can be checked that this poset is a complete lattice where $\hat{0} = 3$ and $\hat{1} = 90$. However, this is not distributive by means of a counter example:*

$$45 \vee (30 \wedge 18) = 45 \vee 3 = 45 \quad \text{that is different to} \quad (45 \vee 30) \wedge (45 \vee 18) = 90 \wedge 90 = 90.$$

Another way to prove the non-distributivity of this lattice is that the diamond is contained as a sublattice in its structure.

1.1.4 The set inclusion lattice

For any finite set A , the *power set* of A , denoted by $\mathcal{P}(A)$, is the set of all subsets of A including the empty set \emptyset and A itself. The power set carries a natural order, namely *set inclusion* \subseteq . More formally, for any sets $A_1, A_2 \in \mathcal{P}(A)$, the set A_1 is a *subset* of the set A_2 , denoted $A_1 \subseteq A_2$ if all elements of A_1 are also elements of A_2 . Hence, the power set $\mathcal{P}(A)$ equipped by \subseteq is a poset and we can write $\langle \mathcal{P}(A), \subseteq \rangle$.

Example 1.4. *The power set of $A = \{1, 2, 3\}$ is the set*

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\},$$

which is equipped by inclusion relation \subseteq . Thus, in the poset $\langle \mathcal{P}(\{1, 2, 3\}), \subseteq \rangle$, the sets $\{1, 2\}$ and $\{1, 3\}$ are incomparable; moreover, both $\{1, 2\}$ and $\{1, 3\}$ cover $\{1\}$.

Dealing with this lattice is especially straightforward because the meet and join operations coincide with the set intersection “ \cap ” and the set union “ \cup ”, respectively. Hence, this is a distributive lattice because the union and the intersection distribute to each other. Moreover, it is complete because the zero is the empty set and the unit is the set A .

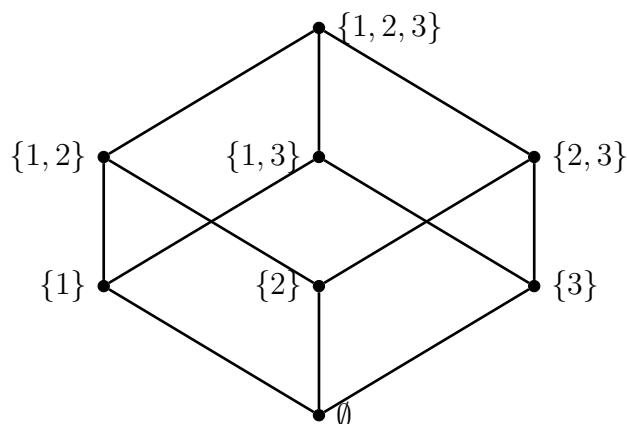


FIGURE 1.4: Hasse diagram of the lattice $\langle \mathcal{P}(\{1, 2, 3\}), \subseteq \rangle$.

1.1.5 The set partition lattice

Let A be a set, and a *partition* of A is a collection of non-empty subsets A_1, \dots, A_m of A such that each element of A is in exactly one of those subsets; more formally,

$$A = A_1 \cup \dots \cup A_m, \quad A_i \cap A_j = \emptyset, \quad (1.3)$$

for every $i, j \in \{1, \dots, m\}, i \neq j$. Thus, the sets A_1, \dots, A_m are pairwise disjoint sets, and their union is A . The subsets A_1, \dots, A_m are called the *parts* or *classes* of the partition (see e.g. [Brualdi 1977](#), Chapter 2). A class with a single element is called *atomic* and a class which is not atomic is *composite*.

The set of partitions of a finite set is a well-known instance for the case of non-distributive lattice. Before demonstrating it, we need to specify the ordering relationship for the set of partitions of A , called $P(A)$. More specifically, for any finite set A , a partition P_1 is a *refinement* of a partition P_2 ; or,

$$P_1 \leq P_2 \quad \text{for } P_1, P_2 \in P(A)$$

whenever P_1 is *finer* than P_2 , or, put differently, whenever P_2 is *coarser* than P_1 , i.e., if every set in P_2 can be expressed as a union of sets in P_1 . This relation \leq on the set of partitions of A is a partial order. The meet and the join operations are more convoluted than in the set inclusion lattice. More concretely,

- (A) the meet $P_1 \wedge P_2$ is the partition whose sets are the pairwise intersections of sets of P_1 and P_2 , and it is the coarsest which refines both P_1 and P_2 ;
- (B) the join $P_1 \vee P_2$ is the partition whose sets are exactly a union of sets from both P_1 and P_2 , and it is the finest which is simultaneously refined by P_1 and P_2 .

We refer the readers to [Canfield \(2001\)](#) and [Pittel \(2000\)](#) for more details and discussion.

Example 1.5. We consider two partitions P_1 and P_2 of a set $\{1, 2, 3, 4, 5, 6\}$ where

- $P_1 = \{\{1, 3\}, \{2\}, \{4, 5\}, \{6\}\}$, and
- $P_2 = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6\}\}$.

By (A), the meet \wedge of P_1 and P_2 is determined as the partition $\{\{1\}, \{2\}, \{3\}, \{4, 5\}, \{6\}\}$ that is coarser than $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. On the other hand, by (B), the partition $\{\{1, 3\}, \{2\}, \{4, 5, 6\}\}$ is specified as the join \vee of P_1 and P_2 which is finest among upper bound partitions $\{\{1, 2, 3\}, \{4, 5, 6\}\}$, $\{\{1, 3\}, \{2, 4, 5, 6\}\}$, and $\{\{2\}, \{1, 3, 4, 5, 6\}\}$.

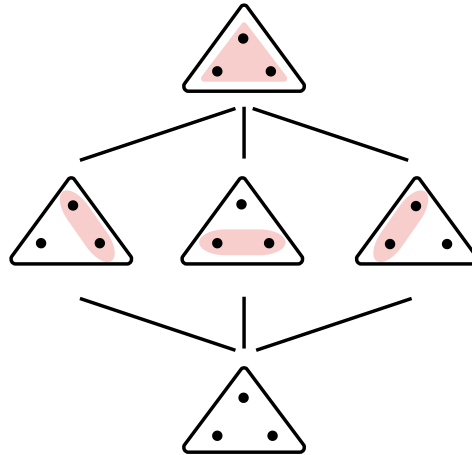


FIGURE 1.5: Hasse diagram of the lattice of partitions of a set A of three distinct elements ordered by the refinement \leq . A red colored region indicates a subset of A that forms a member of the enclosing partition. Uncolored dots indicate single-element subsets or atomic subsets.

Under these operations, the poset $\langle P(A), \leq \rangle$ is a complete lattice where the zero is the partition of A into $|A|$ atomic classes and the unit is the partition made up of one composite class containing all of possible elements in A . Furthermore, the set partition lattice is typically non-distributive, as shown in the following example.

Example 1.6. *We consider A a set of three distinct elements. Thus, A has five partitions displayed in Figure 1.5 where each shaded (red) area a subset of a specified partition of A . The partition on the top is a single set of three elements whereas the partition on the bottom has three single-element subsets. As we can see, the structure of $P(A)$ in this example contains a diamond and this is sufficient to show that $\langle P(A), \leq \rangle$ is a non-distributive lattice.*

1.2 Graphs, color classes and colored graphs

In this section, we introduce the families of undirected graphs and colored graphs together with their associated lattice structures. For more detail, we refer the reader to [Gross and Yellen \(2003\)](#), [Bondy and Murty \(2008\)](#).

1.2.1 Undirected graphs

An undirected graph, denoted by $G = (V, E)$, consists of a set of vertices V and a set of edges $E \subseteq \{(i, j) \in V \times V, i < j\}$. If one wishes to make clear which graph is under consideration, $V(G)$ and $E(G)$ can be used to denote the vertex set and edge set of the graph G . $V = \{1, 2, \dots, p\}$ is shortened as $V = [p]$.

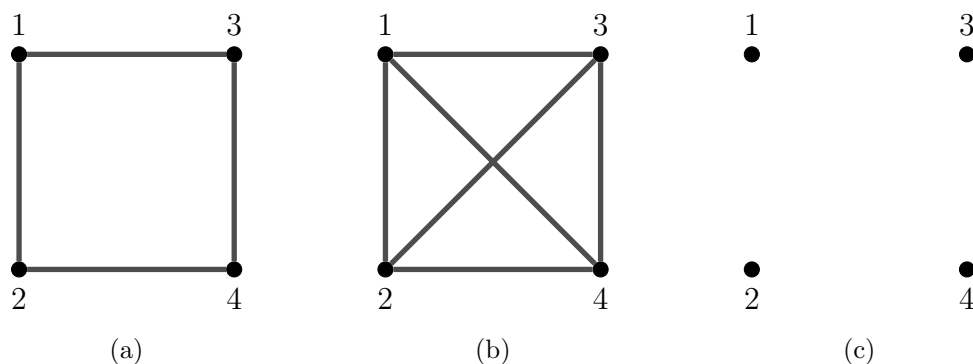


FIGURE 1.6: Example of (a) an undirected graph, (b) a complete graph, and (c) an empty graph. The three graphs here use a common vertex set $V = [4]$.

The vertices of a graph are drawn as black dots or circles and edges are drawn as straight lines between the vertices. A graph is *complete* if all possible edges are present and *empty* if it has no edges. To calculate the number of edges in a complete graph with p vertices, we think about the situation starting from the perspective of the first node: This node is connected to $p - 1$ other nodes; If we move to the second node, it adds $p - 2$ more connections, and so forth. So we have $\sum_{i=1}^p (p - i) = p(p - 1)/2$ edges.

Example 1.7. An example of a graph $G = (V, E)$ is given in Figure 1.6a where the vertex set $V = \{1, 2, 3, 4\}$ and the edge set $E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$. Moreover, illustrations of the complete graph and the empty graph with $V = [4]$ are displayed in Figures 1.6b and 1.6c, respectively.

A *path* of length k from a vertex v to a vertex u is formed by a sequence of distinct vertices ($v \equiv v_0, v_1, \dots, v_k \equiv u$) such that (v_i, v_{i+1}) is an edge for all $0 \leq i < (k - 1)$.

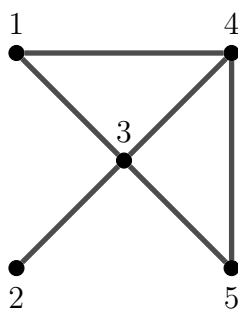


FIGURE 1.7: An undirected graph $G = (V, E)$ with the vertex set $V = [5]$.

If there is an edge (u, v) between vertices u and v , then the edge (u, v) is said to be *incident* with u and v and vertices u, v are *adjacent*.

In a graph G , the *degree* of a vertex v is the number of edges of G which are incident to v and is denoted by $d(v)$. A vertex with degree zero is an *isolated* vertex.

The *neighbourhood* of a vertex v in G , denoted by $\text{ne}(v)$, is the set of vertices which are adjacent to v . Whereas $\text{cl}(v) = \{v\} \cup \text{ne}(v)$ is the *closure* of v . Therefore, the degree of v in G is equal to the cardinality of its neighbourhood.

Example 1.8. An example for the graph in Figure 1.7 is considered where the vertex 4 in G is adjacent to vertices 1, 3, and 5, therefore, its degree is 3 with neighbourhood set $\text{ne}(4) = \{1, 3, 5\}$. Furthermore, as we can see, there are three paths between vertices 1 and 2, namely $(1, 3, 2)$, $(1, 4, 3, 2)$, and $(1, 4, 5, 3, 2)$; however, there is no path of length $k = 1$ because vertices 1 and 2 are not adjacent.

1.2.2 The lattice structure of undirected graphs

Consider the family of undirected graphs with the vertex set V , denoted by U_V . A natural partial order within this family is given by the *edge set inclusion*. More formally, for two graphs $G = (V, E(G)), H = (V, E(H)) \in U_V$, the partial order of graphs G and H is defined on the edge set inclusion, that is

$$G \leq H \quad \text{whenever} \quad E(G) \subseteq E(H) \quad (1.4)$$

so that the *meet operation* and the *join operation* are specified as follows

$$G \wedge H = (V, E(G) \cap E(H)) \quad \text{and} \quad G \vee H = (V, E(G) \cup E(H)), \quad (1.5)$$

respectively. Therefore, it is straightforward to see that $\langle U_V, \leq \rangle$ is a complete distributive lattice where the *zero* is the empty graph and the *unit* is the complete graph.



FIGURE 1.8: Example of the partial order between two undirected graphs G and H in $U_{[4]}$.

Example 1.9. We consider two graphs $G = (V, E(G))$ and $H = (V, E(H))$ in $U_{[4]}$ which are displayed in Figure 1.8 where

$$E(G) = \{(1, 2), (1, 3), (2, 4)\}, \quad E(H) = \{(1, 2), (1, 3), (2, 4), (3, 4)\}.$$

It is apparent that $E(G) \subseteq E(H)$ and therefore $G \leq H$.

1.2.3 Color classes and colored graphs

There are many general and widely applicable concepts of graph colorings (see e.g. Gross and Yellen 2003, chapter 5). One of them deals with partitioning the underlying set of a structure into parts where each of which satisfies a given requirement. In particular, let $G = (V, E)$ be an undirected graph. A *vertex coloring* of G is a partition $\mathcal{V} = \{\mathcal{V}_1, \dots, \mathcal{V}_k\}$ of V where $\mathcal{V}_1, \dots, \mathcal{V}_k$ are referred as *vertex color classes*. Similarly, an *edge coloring* of G is a partition $\mathcal{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_h\}$ of E where $\mathcal{E}_1, \dots, \mathcal{E}_h$ are referred as *edge color classes*. We call $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a *colored graph*.

On the other hand, the *uncolored version* of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the undirected graph $G = (V, E)$ obtained by removing colors from \mathcal{G} . More formally, the vertex set V is obtained by merging vertex color classes, i.e. $V = \bigcup_{i=1}^k \mathcal{V}_i$, and the edge set E is obtained by merging edge color classes, i.e. $E = \bigcup_{j=1}^h \mathcal{E}_j$.

Similar to the case of (uncolored) undirected graphs, $\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G})$ are used to indicate respectively the vertex coloring and the edge coloring of the specified graph \mathcal{G} . For any vertices (respectively, edges) belonging to the same color class, those vertices (respectively, edges) are depicted in the same color. Note that, the atomic color classes are displayed in black color, in other words, the black color is used for vertices and edges standing alone in a color class. In order to make color classes readable also in black and white printing, we use different symbols to denote them. Remark that, the black vertices or the black edges with no symbols marked on them are not to be interpreted as being in the same color class.

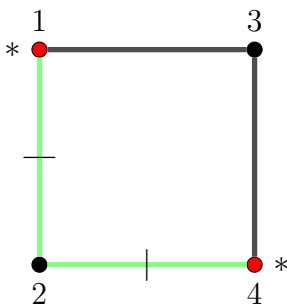


FIGURE 1.9: A colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ based on the uncolored undirected graph $G = (V, E)$ shown in Figure 1.6a.

Example 1.10. One possible colorings of the graph $G = (V, E)$ in Figure 1.6a is displayed by Figure 1.9. In this colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the vertex set V is partitioned into three vertex color classes $\mathcal{V} = \{\{1, 4\}, \{2\}, \{3\}\}$ where vertices 1 and 4 are depicted

in red and they are marked with an asterisk. Moreover, the edge coloring is a partition of E into three edge color classes $\mathcal{E} = \{(1, 2), (2, 4)\}, \{(1, 3)\}, \{(3, 4)\}$ where $(1, 2)$ and $(2, 4)$ are depicted in green and they are marked with a dash. Since vertices 2 and 3 form two distinct atomic color classes, thus they are depicted in black and left unmarked. Similarly, edges $(1, 3)$ and $(3, 4)$ also stand alone in two distinct atomic color classes, thus they are drawn via black straight lines without any symbols marked on them.

1.2.4 The lattice structure of colored graphs

The lattice structure of undirected graphs is based on the set inclusion ordering relation and it inherits the important properties of this type of lattice. On the other hand, colored graphs are specified by partitions of the vertex set and the edge set, for this reason, [Gehrmann \(2011\)](#) introduced a lattice of colored graphs based on the set partition order.

Consider the set \mathcal{C}_V of colored graphs with the vertex set V , then following the definition from [Gehrmann \(2011\)](#), for $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G})), \mathcal{H} = (\mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{H})) \in \mathcal{C}_V$ defined on underlying uncolored graphs $G = (V, E(G))$ and $H = (V, E(H))$, we say $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$ whenever

$$(C.i) \quad G \leq H;$$

$$(C.ii) \quad \mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{H});$$

$$(C.iii) \quad \text{every edge color class in } \mathcal{E}(\mathcal{G}) \text{ is a union of color classes in } \mathcal{E}(\mathcal{H}).$$

It can be readily checked that “ $\preceq_{\mathcal{C}}$ ” defines a partial order on \mathcal{C}_V .

Example 1.11. Consider the colored graphs $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ and $\mathcal{H} = (\mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{H}))$ in $\mathcal{C}_{[4]}$ represented in [Figure 1.10](#) where

$$\mathcal{V}(\mathcal{G}) = \{\{1, 4\}, \{2, 3\}\}, \quad \mathcal{E}(\mathcal{G}) = \{(1, 2), (1, 3), (2, 4)\}$$

and

$$\mathcal{V}(\mathcal{H}) = \{\{1, 4\}, \{2\}, \{3\}\}, \quad \mathcal{E}(\mathcal{H}) = \{(1, 2), (2, 4)\}, \{(1, 3)\}, \{(3, 4)\}.$$

with their (uncolored) underlying graphs $G = (V, E(G))$ and $H = (V, E(H))$ shown in [Figure 1.8](#). It holds that $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$ because all of the conditions (C.i)-(C.iii) is satisfied; specifically,

$$(i) \quad G \leq H \text{ since } E(G) \subseteq E(H) \text{ (see Example 1.9);}$$

- (ii) $\mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{H})$ since the subset $\{2, 3\}$ in $\mathcal{V}(\mathcal{G})$ is a union of subsets $\{2\}, \{3\}$ in $\mathcal{V}(\mathcal{H})$, where the relation “ \geq ” between sets $\mathcal{V}(\mathcal{G})$ and $\mathcal{V}(\mathcal{H})$ is specified in Section 1.1.5;
- (iii) the edge color class $\{(1, 2), (1, 3), (2, 4)\}$ in $\mathcal{E}(\mathcal{G})$ is a union of color classes $\{(1, 3)\}$ and $\{(1, 2), (2, 4)\}$ in $\mathcal{E}(\mathcal{H})$.

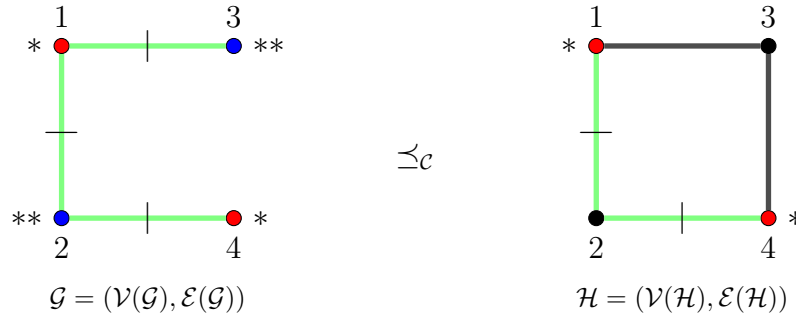


FIGURE 1.10: Example of the partial order between two colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{C}_{[4]}$.

Moreover, [Gehrmann \(2011\)](#) proved that $\langle \mathcal{C}_V, \preceq_c \rangle$ is a complete lattice with meet and join operations given by

$$\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}(\mathcal{G}) \vee \mathcal{V}(\mathcal{H}), \mathcal{E}^*(\mathcal{G}) \vee \mathcal{E}^*(\mathcal{H})), \quad (1.6)$$

and

$$\mathcal{G} \vee \mathcal{H} = (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H}), \mathcal{E}^{**}(\mathcal{G}) \wedge \mathcal{E}^{**}(\mathcal{H})), \quad (1.7)$$

where \wedge and \vee are the meet and the join operations between partitions defined in (A) and (B), respectively, in Section 1.1.5. Moreover, $\mathcal{E}^*(\mathcal{G}) \subseteq \mathcal{E}(\mathcal{G}), \mathcal{E}^*(\mathcal{H}) \subseteq \mathcal{E}(\mathcal{H})$ are maximal with the property that they are partitions of the same set of edges inside $E(\mathcal{G}) \cap E(\mathcal{H})$, and $\mathcal{E}^{**}(\mathcal{G}) = \mathcal{E}(\mathcal{G}) \cup \{\{E(\mathcal{H}) \setminus E(\mathcal{G})\}\}, \mathcal{E}^{**}(\mathcal{H}) = \mathcal{E}(\mathcal{H}) \cup \{\{E(\mathcal{G}) \setminus E(\mathcal{H})\}\}$.

The *zero* in $\langle \mathcal{C}_V, \preceq_c \rangle$ is given by the empty graph in which all vertices are of the same color and the *unit* is the complete graph with atomic color classes. Figure 1.11 illustrates the *zero* and the *unit* in $\langle \mathcal{C}_{[4]}, \preceq_c \rangle$.

A weak point of the lattice of colored graphs with respect to the lattice of undirected graphs is that it does not hold the important property of the distributivity. This follows from the fact that the set partition lattice is not distributive as shown below.

Let $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E})$, $\mathcal{H} = (\mathcal{V}(\mathcal{H}), \mathcal{E})$, and $\mathcal{K} = (\mathcal{V}(\mathcal{K}), \mathcal{E})$, where edge colorings of \mathcal{G}, \mathcal{H}

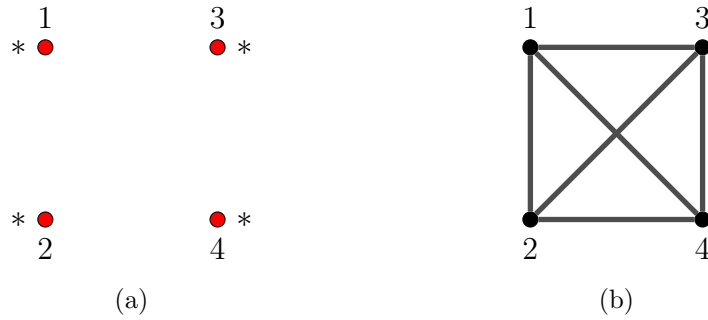


FIGURE 1.11: Colored graphs represent (a) the *zero* and (b) the *unit* in the lattice $\langle \mathcal{C}_{[4]}, \preceq_c \rangle$.

and \mathcal{K} are all equal to a common edge coloring \mathcal{E} , i.e.

$$\mathcal{E}(\mathcal{G}) = \mathcal{E}(\mathcal{H}) = \mathcal{E}(\mathcal{K}) = \mathcal{E},$$

with \mathcal{E} a specified partition from the set E . If the distributivity law was satisfied in $\langle \mathcal{C}_V, \preceq_c \rangle$ then, we would have

$$\mathcal{G} \vee (\mathcal{H} \wedge \mathcal{K}) = (\mathcal{G} \vee \mathcal{H}) \wedge (\mathcal{G} \vee \mathcal{K})$$

for all $\mathcal{V}(\mathcal{G}), \mathcal{V}(\mathcal{H}), \mathcal{V}(\mathcal{K}) \in P(V)$, $\mathcal{E} \in P(E)$. Now,

$$\begin{aligned} \mathcal{G} \vee (\mathcal{H} \wedge \mathcal{K}) &= (\mathcal{V}(\mathcal{G}) \wedge (\mathcal{V}(\mathcal{H}) \vee \mathcal{V}(\mathcal{K})), \mathcal{E}) \\ (\mathcal{G} \vee \mathcal{H}) \wedge (\mathcal{G} \vee \mathcal{K}) &= ((\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H})) \vee (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{K})), \mathcal{E}), \end{aligned}$$

due to (1.6) and (1.7). However,

$$\mathcal{V}(\mathcal{G}) \wedge (\mathcal{V}(\mathcal{H}) \vee \mathcal{V}(\mathcal{K})) \neq (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H})) \vee (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{K})),$$

for all $\mathcal{V}(\mathcal{G}), \mathcal{V}(\mathcal{H}), \mathcal{V}(\mathcal{K}) \in P(V)$ since $\langle P(V), \leq \rangle$ is not distributive. Thus $\langle \mathcal{C}_V, \preceq_c \rangle$ is non-distributive either.

1.3 Covariance matrices, concentration matrices and concentration graphs

Let $V = \{1, \dots, p\}$ be the finite set with the cardinality $|V| = p$, and $\mathbf{Y} \equiv \mathbf{Y}_V$ a vector of continuous random variables indexed by V . The *covariance matrix* of \mathbf{Y} , denoted by Σ , is a symmetric and positive definite matrix giving the covariances between each pair

of elements of \mathbf{Y} . More specifically,

$$\Sigma = \text{Var}(\mathbf{Y}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix},$$

where

$$\sigma_{ii} = \text{Var}(Y_i), \quad \sigma_{ij} = \text{Cov}(Y_i, Y_j)$$

for all $i, j \in V$.

We can split $\mathbf{Y}_V \in \mathbb{R}^p$ into two subvectors $\mathbf{Y}_A \in \mathbb{R}^{p_1}$ and $\mathbf{Y}_B \in \mathbb{R}^{p_2}$,

$$\mathbf{Y}_V = \begin{pmatrix} \mathbf{Y}_A \\ \mathbf{Y}_B \end{pmatrix}, \quad \text{with } p_1 + p_2 = p.$$

Hence, A and B are two subsets of V such that

$$V = A \cup B, \quad A \cap B = \emptyset,$$

where the cardinalities of A and B are p_1 and p_2 , respectively. It is realized that the covariance matrix of \mathbf{Y}_V has been partitioned into four parts as shown below

$$\text{Var}(\mathbf{Y}_V) = \Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix},$$

where Σ_{AA} and Σ_{BB} are the covariance matrices for the random vector \mathbf{Y}_A and \mathbf{Y}_B , respectively. And Σ_{AB} is the covariances between the elements of \mathbf{Y}_A and \mathbf{Y}_B .

We write $\Sigma_{AA|B}$ to denote the covariance matrix of $\mathbf{Y}_A \mid \mathbf{Y}_B$, which is the residual vector deriving from the linear least squares predictor of \mathbf{Y}_A on \mathbf{Y}_B , see e.g. [Whitaker \(1990\)](#). Hence, the covariance matrix $\Sigma_{AA|B}$ can be computed as

$$\Sigma_{AA|B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA},$$

see [Chatfield and Collins \(2018\)](#), [Krzanowski \(2000\)](#). $\Sigma_{AA|B}$ is also called the *partial covariance matrix* and $\{\sigma_{ij|B}\}_{i,j \in A}$ are its elements.

For $i, j \in A$, the *partial correlation coefficient* between two variables Y_i and Y_j given \mathbf{Y}_B , denoted by $\rho_{ij|B}$, can be computed in terms of $\sigma_{ij|B}$ as follows

$$\rho_{ij|B} = \frac{\sigma_{ij|B}}{\sqrt{\sigma_{ii|B}\sigma_{jj|B}}}. \quad (1.8)$$

A partial correlation coefficient is a measure of the linear dependence of a pair of random variables, belonging to a collection of random variables, in the linearly adjusted influence of the remaining variables in the collection.

The inverse of the covariance matrix is called the *concentration matrix* or *precision matrix* of \mathbf{Y} , and we denote it by $\Theta = \Sigma^{-1}$, with elements $\{\theta_{ij}\}_{i,j \in V}$. Suppose that we partition the concentration matrix in a corresponding manner as above, that is

$$\Theta = \begin{pmatrix} \Theta_{AA} & \Theta_{AB} \\ \Theta_{BA} & \Theta_{BB} \end{pmatrix};$$

by the inverse of a block partitioned matrix, we get the result

$$\Theta_{AA}^{-1} = \Sigma_{AA|B}.$$

The relation between the partial correlation coefficients and the concentration matrix is expressed by the following lemma.

Lemma 1.2. *Let \mathbf{Y}_V be a random vector indexed by $V = [p]$ and $\Theta = \{\theta_{ij}\}_{i,j \in V} = \Sigma^{-1}$ the concentration matrix of \mathbf{Y}_V . Then, the partial correlation coefficient between Y_i and Y_j , given all others $\mathbf{Y}_{V \setminus \{i,j\}}$, can be computed from the entries of the matrix Θ as*

$$\rho_{ij|V \setminus \{i,j\}} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}}, \quad \forall i, j \in V. \quad (1.9)$$

Moreover, the diagonal elements θ_{ii} of Θ are reciprocals of the conditional variances given the remaining variables, i.e.

$$\theta_{ii} = \text{Var}(Y_i | \mathbf{Y}_{V \setminus \{i\}})^{-1}, \quad \forall i \in V. \quad (1.10)$$

The extended proof of this lemma can be found in [Lauritzen \(1996\)](#) and it is proved compactly here below.

Proof. If \mathbf{Y} is partitioned into $(\mathbf{Y}_A, \mathbf{Y}_B)^T$ such that $\mathbf{Y}_A = (Y_i, Y_j)$ and $\mathbf{Y}_B = \mathbf{Y}_{V \setminus \{i,j\}}$ for any $i, j \in V$, then the partial correlation between variables Y_i and Y_j given the remaining variables can be written as

$$\rho_{ij|V \setminus \{i,j\}} = \frac{(\Theta_{AA}^{-1})_{12}}{\sqrt{(\Theta_{AA}^{-1})_{11}(\Theta_{AA}^{-1})_{22}}} = -\frac{(\Theta_{AA})_{12}}{\sqrt{(\Theta_{AA})_{22}(\Theta_{AA})_{11}}}. \quad (1.11)$$

The second equality holds by the simple inversion formula of 2×2 matrix

$$\begin{pmatrix} (\Theta_{AA}^{-1})_{11} & (\Theta_{AA}^{-1})_{12} \\ (\Theta_{AA}^{-1})_{21} & (\Theta_{AA}^{-1})_{22} \end{pmatrix} = \Theta_{AA}^{-1} = \frac{1}{|\Theta_{AA}|} \begin{pmatrix} (\Theta_{AA})_{22} & -(\Theta_{AA})_{12} \\ -(\Theta_{AA})_{21} & (\Theta_{AA})_{11} \end{pmatrix}.$$

Remember that Θ_{AA} is a concentration matrix of (Y_i, Y_j) , thus, Θ_{AA} is a matrix made up of elements

$$\Theta_{AA} = \begin{pmatrix} \theta_{ii} & \theta_{ij} \\ \theta_{ji} & \theta_{jj} \end{pmatrix}.$$

Hence, by replacing $(\Theta_{AA})_{12} = \theta_{ij}$, $(\Theta_{AA})_{11} = \theta_{ii}$, $(\Theta_{AA})_{22} = \theta_{jj}$ into the equation (1.11), we obtain (1.9).

Furthermore, if $A = \{i\}$ and $B = V \setminus \{i\}$, then

$$\theta_{ii} = \text{Var}(Y_i | \mathbf{Y}_{V \setminus \{i\}})^{-1}, \quad \forall i \in V,$$

by the relation

$$\Theta_{AA}^{-1} = \Sigma_{AA|B} = \text{Var}(\mathbf{Y}_A | \mathbf{Y}_B).$$

□

Let $G = (V, E)$ be an undirected graph with the vertex set $V = \{1, \dots, p\}$ and Θ the concentration matrix of random vector $\mathbf{Y}_V = (Y_1, \dots, Y_p)^T$ with the corresponding set of vertices V . We say that Θ is *adapted* to G if

$$(i, j) \notin E(G) \quad \text{implies} \quad \theta_{ij} = 0 \tag{1.12}$$

for any $i \neq j$ and $i, j \in V$. The graph G is called a *concentration graph* of \mathbf{Y}_V .

The concentration graph G is called a *conditional independence graph* or *independence graph* of \mathbf{Y}_V when there is no edge between the vertices if the random variables corresponding to the vertices are conditional independent given the rest of the random variables. More specifically, the concentration graph $G = (V, E)$ is a independence graph of \mathbf{Y}_V with $V = [p]$ if

$$(i, j) \notin E(G) \quad \text{implies} \quad Y_i \perp\!\!\!\perp Y_j | \mathbf{Y}_{V \setminus \{i, j\}}. \tag{1.13}$$

Example 1.12. We consider the graph $G = (V, E)$ in Figure 1.6a and $\mathbf{Y}_{[4]} = (Y_1, Y_2, Y_3, Y_4)$ is a random vector. We obtain the concentration matrix Θ of $\mathbf{Y}_{[4]}$ that is adapted to G

as

$$\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & 0 \\ * & \theta_{22} & 0 & \theta_{24} \\ * & * & \theta_{33} & \theta_{34} \\ * & * & * & \theta_{44} \end{pmatrix},$$

where $\theta_{14} = \theta_{23} = 0$ since the edges $(1, 4)$ and $(2, 3)$ are missing in G ; moreover, G is called the independent graph of $\mathbf{Y}_{[4]}$, i.e. $Y_1 \perp\!\!\!\perp Y_4 \mid (Y_2, Y_3)$ and $Y_2 \perp\!\!\!\perp Y_3 \mid (Y_1, Y_4)$.

Due to the symmetric nature of the concentration matrix Θ , only the upper triangular portion is usually defined. Hence, the entries with the star “*” has the same value as the opposite corner of the matrix.

1.4 Graphical models

The *graphical model* for random variables \mathbf{Y} is a single family of probability distributions for \mathbf{Y} with the intuitive representation of relationships between variables given by graphs. When the family of distributions is multivariate normal then we speak of the *Gaussian graphical models* or *covariance selection models* (see e.g. [Dempster \(1972\)](#), [Whitakker \(1990\)](#)). In [Section 1.4.1](#), we will begin from the family of multivariate Gaussian models where no statistical independencies are assumed to hold. Next, in [Section 1.4.2](#), we will present the overview of Gaussian graphical models, which is a popular used family of probabilistic models by restricting of conditional independence of selected pairs of variables, given the remaining ones, in term of the undirected graph. A special case of equality constraints for the Gaussian distribution is symmetry constraints. The model can be represented by vertex and edge colored graphs, where parameters associated with equally colored vertices or edges are restricted to being identical. We call the family of these models *colored graphical models*. Hence, [Section 1.4.3](#) is devoted to one of the types of such restrictions is equality between specified elements of the concentration matrix (RCON), introduced in [Højsgaard and Lauritzen \(2008\)](#). For each section above, we also review recent results on maximum likelihood estimation of such models.

1.4.1 Multivariate Gaussian distribution

Let $\mathbf{Y} \equiv \mathbf{Y}_V$ be a vector of continuous random variables indexed by $V = \{1, \dots, p\}$. We say that \mathbf{Y} follows a multivariate normal (or Gaussian) distribution with a mean

vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and a covariance matrix Σ if \mathbf{Y} has the density function

$$f(\mathbf{y}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) / 2 \right\}, \quad (1.14)$$

where $|\Sigma|$ is the determinant of Σ . Then, we write $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Moreover, the *conditional distribution* of $\mathbf{Y}_A \mid \mathbf{Y}_B = \mathbf{y}_B$ is a multivariate normal with the covariance matrix, denoted by $\Sigma_{AA|B}$,

$$\Sigma_{AA|B} = \text{Var}(\mathbf{Y}_A \mid \mathbf{Y}_B = \mathbf{y}_B).$$

In the multivariate normal distribution, the conditional independence is simply reflected in the concentration matrix of the distribution through zero entries. Namely, we have the following proposition.

Proposition 1.3. *Consider the random vector $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, then it holds for $i, j \in V$ with $i \neq j$ that*

$$Y_i \perp\!\!\!\perp Y_j \mid \mathbf{Y}_{V \setminus \{i, j\}} \iff \rho_{ij|V \setminus \{i, j\}} = 0 \iff \theta_{ij} = 0.$$

Indeed, in the case of the jointly normal distribution, the partial correlation $\rho_{ij|V \setminus \{i, j\}}$ is zero if and only if Y_i is conditionally independent from Y_j given $\mathbf{Y}_{V \setminus \{i, j\}}$. This property does not hold in the general case.

Our interest is for the structure of Σ (or Θ) and for the rest of the section we assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$. Then the probability density function of \mathbf{Y} can be rewritten in term of Θ as

$$\begin{aligned} f(\mathbf{y}|\Theta) &= (2\pi)^{-p/2} |\Theta|^{1/2} \exp \left\{ -\mathbf{y}^T \Theta \mathbf{y} / 2 \right\} \\ &= (2\pi)^{-p/2} |\Theta|^{1/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Theta \mathbf{y} \mathbf{y}^T) \right\}. \end{aligned} \quad (1.15)$$

The expression (1.15) shows that the normal distribution with $\boldsymbol{\mu} = \mathbf{0}$ belongs to the regular exponential family with the canonical parameter Θ and the canonical statistic $-\mathbf{y} \mathbf{y}^T / 2$. We can find more detail in [Lauritzen \(1996\)](#), [Rencher and Christensen \(2012\)](#), [Tong \(2012\)](#), and [Chatfield and Collins \(2018\)](#).

Then, given a sample $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}$ of independent random vectors from a multivariate normal distribution with covariance matrix Σ , the log-likelihood function for $\Theta = \Sigma^{-1}$

can be written as

$$l(\Theta) = \frac{n}{2} \log |\Theta| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}^{(i)})^T \Theta \mathbf{y}^{(i)} / 2 \quad (1.16)$$

$$= \frac{n}{2} \log |\Theta| - \frac{1}{2} \operatorname{tr} \left\{ \Theta \left(\sum_{i=1}^n \mathbf{y}^{(i)} (\mathbf{y}^{(i)})^T \right) \right\} \quad (1.17)$$

$$= \frac{n}{2} \log |\Theta| - \frac{n}{2} \operatorname{tr}(S\Theta), \quad (1.18)$$

where $S = n^{-1} \sum_{i=1}^n \mathbf{y}^{(i)} (\mathbf{y}^{(i)})^T = \mathbf{y}^T \mathbf{y} / n$ is the *sample covariance matrix*.

Following Lauritzen (1996), in the multivariate normal model with zero mean, the maximum likelihood estimate of the unknown concentration matrix exist if and only if S is positive definite. This happens with probability one if $n \geq p$ and never when $n < p$. When the maximum likelihood estimate of Θ exists, it is given by

$$\hat{\Theta} = S^{-1}. \quad (1.19)$$

1.4.2 Gaussian graphical models

Definition 1.1. Let $G = (V, E)$ be an undirected graph with the vertex set $V = \{1, \dots, p\}$ and \mathbf{Y}_V a random vector indexed by V . Then, \mathbf{Y}_V is said to satisfy the *Gaussian graphical model* with the graph $G = (V, E)$ if

- (i) \mathbf{Y}_V follows the jointly normal distribution with $\Sigma = \operatorname{Var}(\mathbf{Y}_V)$, and
- (ii) $\Theta = \Sigma^{-1}$ is adapted to G .

In other words, as in Lauritzen (1996, section 5.2), Maathuis *et al.* (2018, section 8.5), and Cox and Wermuth (2014, chapter 2), the *Gaussian graphical model* for \mathbf{Y}_V with the graph G is given by assuming that \mathbf{Y}_V follows a multivariate normal distribution which obeys the undirected pairwise Markov property with respect to G .

Let $\mathcal{S}^+(G)$ be the set of symmetric and positive definite matrices which have zero elements whenever there is no edge in G , i.e.

$$\mathcal{S}^+(G) = \{M^T = M, M \succ 0 \text{ s.t. } M_{ij} = 0 \text{ whenever } (i, j) \notin E(G)\}. \quad (1.20)$$

Then the *Gaussian graphical model* for \mathbf{Y} can be compactly described as

$$\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \Sigma), \quad \Theta = \Sigma^{-1} \in \mathcal{S}^+(G). \quad (1.21)$$

Therefore, the sparsity pattern of the inverse covariance matrix Θ gives the conditional independence relationships among the variables. That is,

$$(i, j) \notin E(G) \implies \theta_{ij} = \rho_{ij|V \setminus \{i, j\}} = 0 \iff Y_i \perp\!\!\!\perp Y_j \mid \mathbf{Y}_{V \setminus \{i, j\}}.$$

The model where no conditional independence restrictions are assumed to hold is called the *saturated model*.

Example 1.13. For example, let $\mathbf{Y}_{[4]} = (Y_1, Y_2, Y_3, Y_4)^T$ be the vector of four random variables and $\text{Var}(\mathbf{Y}) = \Sigma$ where

$$\Sigma = \begin{pmatrix} 7/6 & 1/3 & 1/3 & 1/6 \\ * & 7/6 & 1/6 & 1/3 \\ * & * & 7/6 & 1/3 \\ * & * & * & 7/6 \end{pmatrix} \text{ then } \Theta = \begin{pmatrix} 1 & -1/4 & -1/4 & 0 \\ * & 1 & 0 & -1/4 \\ * & * & 1 & -1/4 \\ * & * & * & 1 \end{pmatrix}.$$

From the concentration matrix Θ , we get the structure of the graph $G = (V, E)$ associated with $\mathbf{Y}_{[4]}$, namely, the vertex set $V = \{1, 2, 3, 4\}$ and the edge set $E = \{(1, 2), (1, 3), (2, 3), (3, 4)\}$ as in Figure 1.6a.

Since Gaussian graphical models only pose constraints on the concentration matrix Θ , i.e. $\Theta \in \mathcal{S}^+(G)$, it follows that the Gaussian graphical model is itself a regular exponential model with canonical statistic equals $-S(G)/2$ where $S(G)$ is an incomplete matrix with entries

$$S(G)_{ij} = s_{ij}, \quad \text{for } (i, j) \in E \text{ or } i = j,$$

and the entries $S(G)_{ij}$ for $i \neq j$ and $(i, j) \notin E$ are left unspecified. Here $\{s_{ij}\}_{i, j \in V}$ are elements of the sample covariance matrix S . The maximum likelihood estimate of Σ is unique and it is denoted by $\hat{\Sigma} = \{\hat{\sigma}_{ij}\}_{i, j \in V}$ such that $\hat{\Sigma}^{-1} \in \mathcal{S}^+(G)$ and

$$\hat{\sigma}_{ij} = s_{ij} = S(G)_{ij}$$

for every $(i, j) \in E(G)$ or $i = j$.

1.4.3 RCON models: equality restrictions on concentration

RCON models are Gaussian graphical models which place equality constraints on the entries of the concentration matrix $\Theta = \Sigma^{-1}$. For a model whose conditional independence structure is represented by the graph $G = (V, E)$, the restrictions can be represented by a graph coloring $(\mathcal{V}, \mathcal{E})$ with the vertex coloring \mathcal{V} representing constraints on the

entries on the diagonal of Θ and the edge coloring \mathcal{E} representing constraints in the off-diagonal entries. In particular, as in [Højsgaard and Lauritzen \(2008\)](#),

Definition 1.2. An RCON model on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a Gaussian graphical model with some additional equality constraints, namely that

$$\theta_{uu} = \theta_{vv} \quad \text{if } u, v \in \mathcal{V}_j \text{ for some } j, \quad (1.22)$$

and for $u \neq v$, $w \neq t$,

$$\theta_{uv} = \theta_{wt} \quad \text{if } (u, v), (w, t) \in \mathcal{E}_k \text{ for some } k. \quad (1.23)$$

Roughly speaking, whenever two vertices $u, v \in V$ belong to the same vertex color class, the corresponding two diagonal entries θ_{uu} and θ_{vv} are restricted to being identical. Similarly, two edges $(u, v), (w, t) \in E$ of the same color represent the constraint $\theta_{uv} = \theta_{wt}$. Suppose that the colored graph \mathcal{G} is made up of $(\mathcal{V}, \mathcal{E})$ such that

$$\mathcal{V} = \{\mathcal{V}_1, \dots, \mathcal{V}_r\} \quad \text{and} \quad \mathcal{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_s\},$$

for $r, s \geq 1$. For each vertex color class $\mathcal{V}_k \in \mathcal{V}$, $k = 1, \dots, r$, let T^k be the $p \times p$ diagonal matrix with entries $T_{uu}^k = 1$ if the vertex $u \in \mathcal{V}^k$ and zero otherwise. Similarly for each edge color class $\mathcal{E}_h \in \mathcal{E}$, $h = 1, \dots, s$, let T^h be a symmetric $p \times p$ matrix with $T_{uv}^h = 1$ if the edge $(u, v) \in \mathcal{E}_h$ and zero otherwise. Therefore, given a colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the concentration matrix Θ in RCON model can be rewritten as

$$\Theta = \sum_{c \in \mathcal{V} \cup \mathcal{E}} \theta_c T^c, \quad \theta \in \mathbb{R}^{r+s} \quad (1.24)$$

where c refers to a color class in $\mathcal{V} \cup \mathcal{E}$.

Let $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ be the set of symmetric and positive definite matrices which are written as the sum in (1.24). Formally, the distribution of a random vector \mathbf{Y} is said to lie in the RCON model represented by the colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if

$$\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \Sigma), \quad \Theta = \Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E}). \quad (1.25)$$

Example 1.14. By the graph coloring $(\mathcal{V}, \mathcal{E})$ of the graph \mathcal{G} in [Figure 1.9](#) where

$$\mathcal{V} = \{\{\mathbf{1}, \mathbf{4}\}, \{\mathbf{2}, \mathbf{3}\}\}, \quad \mathcal{E} = \{\{(1, 2), (1, 3), (2, 4)\}, \{(3, 4)\}\},$$

for each color class in $\mathcal{V} \cup \mathcal{E}$, we have that

$$T^{\text{red}} = \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix}, \quad T^{\text{blue}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T^{\text{violet}} = \begin{pmatrix} 0 & \mathbf{1} & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad T^{34} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Hence, the RCON model represented by the graph \mathcal{G} in Figure 1.9 is specified as

$$\mathbf{Y} \sim \mathcal{N}_4(0, \Sigma), \quad \Theta = \Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$$

where

$$\mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \left\{ M^T = M, M \succ 0 \text{ s.t. } M = \begin{pmatrix} \theta & \theta & \theta & 0 \\ \theta & 0 & \theta & 0 \\ 0 & \theta & \theta_{34} & 0 \\ 0 & 0 & 0 & \theta \end{pmatrix} \right\}.$$

By (1.24) and (1.25), the density function of \mathbf{Y} can be written as

$$f(\mathbf{y} \mid \Theta) = (2\pi)^{-p/2} |\Theta|^{1/2} \exp \left\{ -\frac{1}{2} \sum_{c \in \mathcal{V} \cup \mathcal{E}} \theta_c \text{tr}(T^c \mathbf{y} \mathbf{y}^T) \right\}.$$

Here θ is a $(r+s)$ dimensional vector of canonical parameters and $t = (-t_1/2, \dots, -t_{r+s}/2)$ are determined as the canonical statistics with $t_c = \text{tr}(T^c \mathbf{y} \mathbf{y}^T)$ for $c \in \mathcal{V} \cup \mathcal{E}$.

Following Højsgaard and Lauritzen (2008), since the restrictions are linear in the concentration matrix, an RCON model is a linear exponential model and the maximum likelihood estimates of the unknown parameters are uniquely determined by equating the canonical sufficient statistics to their expectation.

1.5 Structure of model spaces

Typically, a family of statistical models can be naturally embedded with a partial order induced by the model inclusion relation; more concretely, given two models M_1 and M_2 , if M_1 is contained by M_2 , denoted $M_1 \subseteq M_2$, then we can say that M_1 is “smaller” than M_2 , or equivalently, M_1 is a *submodel* of M_2 . In order to perform model selection, it

is useful to understand the properties of model spaces. Therefore, in this section, the model structures of the families of (uncolored) undirected graphical models and colored graphical models are investigated by studying the properties on the lattice structure of the associated model space.

1.5.1 Undirected graphical model space

Consider the family of undirected graphical models. Every model is characterized by the associated undirected graph and model inclusion coincides with edge set inclusion. Hence, the search space of undirected graphical models is a complete distributive lattice. More formally, let M_V be the family of undirected graphical models represented by undirected graphs in U_V . Then $\langle M_V, \subseteq \rangle$ is a complete distributive lattice with the submodel relation induced by the edge set inclusion on U_V , that is

$$M(G) \subseteq M(H) \quad \text{if and only if} \quad G \leq H, \quad (1.26)$$

for $M(G), M(H) \in M_V$ represented by graphs $G, H \in U_V$. The meet is represented by the model associated with graph $(V, E(G) \cap E(H))$ and the join is represented by the model associated with graph $(V, E(G) \cup E(H))$. The zero is the graphical model where all of the selected variables are assumed to be independent pairwise which is represented by the empty graph, and the unit is the undirected graphical model represented by the complete graph where no statistical independencies between any pair of variables hold.

Example 1.15. *In Example 1.9, it is apparent that $G \leq H$ since $E(G) \subseteq E(H)$, therefore, we can say that the undirected graphical model represented by G is a submodel of the undirected graphical model represented by H .*

1.5.2 Colored graphical model space

Consider the family of colored graphical models. Each model is represented in terms of the colored graph, and the model inclusion is inherited by the set partition relation. Therefore, the search space of colored graphical models is the complete non-distributive lattice. More precisely, let \mathcal{M}_V be the family of colored graphical models represented by colored graphs in \mathcal{C}_V . Then, $\langle \mathcal{M}_V, \subseteq \rangle$ is a complete non-distributive lattice with the submodel relation induced by the partial order of \mathcal{C}_V based on the set partition ordering relation, that is

$$\mathcal{M}(\mathcal{G}) \subseteq \mathcal{M}(\mathcal{H}) \quad \text{if and only if} \quad \mathcal{G} \preceq_c \mathcal{H}. \quad (1.27)$$

for two colored graphical models $\mathcal{M}(\mathcal{G}), \mathcal{M}(\mathcal{H}) \in \mathcal{M}_V$ represented by colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{C}_V$. The zero is the colored graphical model where all parameters corresponding to the vertices are assumed to be identical and parameters corresponding to the edges are assumed to be all zero. The unit coincides to the unit of the family of undirected graphical models where there are no symmetry constraints on the parameters associated to vertices and edges.

Example 1.16. *We assume that \mathcal{G} and \mathcal{H} are two colored graphs specified in Example 1.11 where $\mathcal{G} \preceq_c \mathcal{H}$. This implies that the colored graphical model $\mathcal{M}(\mathcal{G})$ represented by \mathcal{G} is a submodel of the colored graphical model $\mathcal{M}(\mathcal{H})$ represented by \mathcal{H} through the model inclusion.*

Chapter 2

Colored graphs and colored graphical models for paired data

This chapter is devoted to the analysis of the lattice structures of the search spaces. Particularly, we consider graphical models for paired data and, in the first two sections of this chapter, we formally define this specific class of models. The family of graphical models for paired data is associated with the family of colored graphs for paired data, we then provide a comprehensive analysis of the structure of the family of such graphs. More concretely, in the third section, we consider the model inclusion order, as in [Gehrmann \(2011\)](#), and show that colored graphs for paired data form a complete non-distributive lattice. After that, we identify a relevant sublattice of the colored graphs for paired data based on a given fixed uncolored graph that is complete and distributive under the model inclusion order.

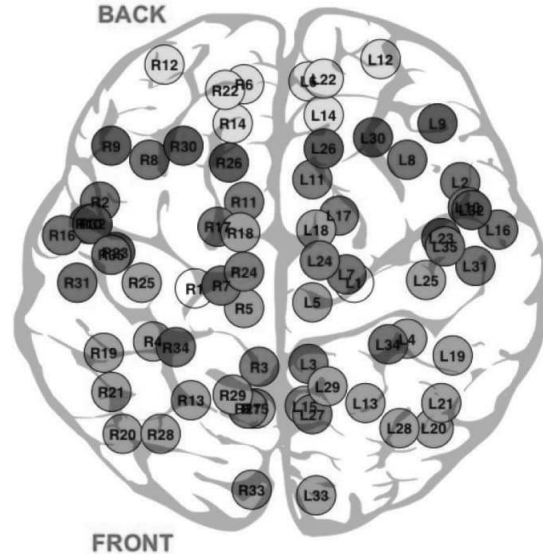
The main result of this chapter is the introduction of a novel order that we call the *twin order*, shown in the fourth section. The connections of this novel order with the traditional model inclusion order are also described. Furthermore, we show that, under the twin order, the family of colored graphs for paired data forms a complete distributive lattice. The meet and the join operations on this lattice can be efficiently computed because they coincide with the set intersection and the set union operations.

2.1 Colored graphical models for paired data

We consider the case of paired data where the variables on every statistical unit are measured in “pairs” in the sense that every variable is uniquely associated with an *homologous*, or *twin*, variable.

Example 2.1. Consider the brain network from functional magnetic resonance imaging (fMRI) data such that:

- variables Y_i 's are associated with regions of interest (ROIs);
- for every variable Y_i on the left hemisphere, there is an homologous variable on the right hemisphere.



Hence, for paired data, the vector of random variables \mathbf{Y}_V can be partitioned as $\mathbf{Y}_V = (\mathbf{Y}_L, \mathbf{Y}_R)^T$ and every variable in \mathbf{Y}_L corresponds to a variable in \mathbf{Y}_R . In this way, in a graphical model analysis, the graph associated with \mathbf{Y}_V can be split, without loss of generality, into a “L”eft part and a “R”ight part and it is of interest to explicitly consider symmetries both between and across the left and the right parts of the network. More concretely, in paired data, there are specific types of symmetries which are naturally of interest: (1) a vertex might be symmetric to its homologous vertex; (2) one edge might be symmetric to its homologous edge, where two edges are homologous if the endpoints of one edge are vertices that is homologous to the endpoints of another. Note that, if an edge has both vertices in the left part of the network then its homologous edge has both vertices on the right part. On the other hand, if one edge is across the left and the right parts, also its homologous edge has a vertex on the left and another vertex on the right part.

So far, we have used the term “symmetry” in an informal way. There exist different ways to formally implement symmetries. In colored graphical models, symmetries are implemented as equality constraints on the parameters and therefore they may imply equal correlations, equal partial correlations or equal concentrations.

For the moment, we restrain from adopting a specific approach and more generally deal with colored graphs and color classes encoding the symmetries for paired data.

2.2 Colored graphs for paired data

We formally introduce colored graphs for paired data as follows.

Definition 2.1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a colored graph on the vertex set $V = [p]$. We say

that \mathcal{G} is a colored graph for paired data (PDCG) if there exists a partition $V = L \cup R$ of V with $L \cap R = \emptyset$ such that

1. $|L| = |R|$ and every vertex in L is uniquely associated with a vertex in R ;
2. every vertex color class is made up of either a single vertex or a pair of homologous vertices;
3. every edge color class is made up of either a single edge or a pair of homologous edges.

Example 2.2. Consider the colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ shown in Figure 2.2 on the vertex set $V = [6]$ where the network is divided into two parts (in the shaded grey areas) such that $V = L \cup R$ with $L = \{1, 2, 3\}$ and $R = \{4, 5, 6\}$. In this network, vertices 1, 2, and 3 in L are uniquely associated with vertices 4, 5, and 6 in R , respectively. Moreover, the vertex coloring is made up of atomic classes $\{1\}$ and $\{4\}$, and composite classes containing two homologous vertices, namely $\{2, 5\}$ and $\{3, 6\}$. Similarly, the edge coloring is also made up of atomic classes $\{(2, 3)\}$ and $\{(3, 5)\}$, and composite classes $\{(1, 2), (4, 5)\}$ and $\{(1, 6), (3, 4)\}$ where each of which contains a pair of homologous edges. Therefore, with such characteristics, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is called a colored graph for paired data.

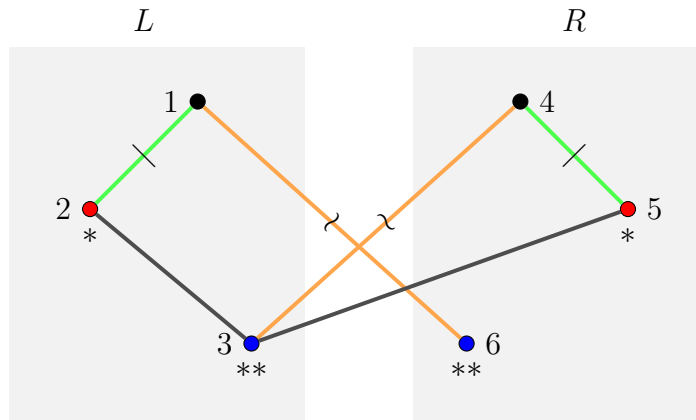


FIGURE 2.2: A colored graph for paired data on the vertex set $V = [6]$ where V is partitioned into $L = \{1, 2, 3\}$ and $R = \{4, 5, 6\}$. Every symmetric twin vertices or every symmetric twin edges are depicted in the same color, and marked by the same symbol. The black color indicates vertices or edges which stand alone in the color classes.

In the analysis of paired data problems, the partition of V into L and R is always known, and therefore, this is the case we consider and assume without loss of generality that $|L| = |R| = p/2 = q$ and $L = \{1, \dots, q\}$ and $R = \{q + 1, \dots, p\}$. Hence, we denote that \mathcal{S}_V the family of PDCGs with the vertex set V .

2.3 The set partition lattice of PDCGs

Clearly, \mathcal{S}_V is strictly contained by \mathcal{C}_V in the sense that the set of PDCGs \mathcal{S}_V is a subfamily of the set of colored graphs \mathcal{C}_V . Therefore, we consider the partial order $\preceq_{\mathcal{C}}$ based on the set partition order of Section 1.2.4, and, with such order, we investigate the structural properties for the lattice structure on the set \mathcal{S}_V . In fact, we consider two following PDCGs:

- (a) the family of colored graphs for paired data on the given fixed uncolored graph,
- (b) the more general case where both the uncolored graphs and the symmetries are unknown,

and show that, under the partial order $\preceq_{\mathcal{C}}$, the family of models in (a) forms a complete distributive lattice whereas in (b) it forms a complete lattice with non-distributivity.

2.3.1 Colored graphs for paired data on a given fixed uncolored graph

Consider an undirected graph $G = (V, E)$ and let \mathcal{S}_G be the set of colored graphs for paired data with the uncolored version G . If \mathcal{G}, \mathcal{H} is a pair of colored graphs in \mathcal{S}_G , the partial order $\preceq_{\mathcal{C}}$ can be simplified as follows: $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$ if and only if

$$(\mathcal{C}_{\mathcal{S}_G}.i) \quad \mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{H});$$

$$(\mathcal{C}_{\mathcal{S}_G}.ii) \quad \mathcal{E}(\mathcal{G}) \geq \mathcal{E}(\mathcal{H}),$$

where \geq is the partial order in the set partition lattice defined in Section 1.1.5. Indeed, since \mathcal{G} and \mathcal{H} have a fixed uncolored version then the edge set inclusion between two underlying undirected graphs is always satisfied. Moreover, for that reason, we can write compactly the condition about the edge color classes of two considered graphs by the order on the poset $\langle P(E), \leq \rangle$ since $\mathcal{E}(\mathcal{G}), \mathcal{E}(\mathcal{H})$ in this case are partitions of a given fixed edge set E .

The following proposition states that the set of the colored graphs for paired data on a given fixed uncolored graph under the partial order “ $\preceq_{\mathcal{C}}$ ” forms a complete distributive lattice. More specifically,

Proposition 2.1. *The poset $\langle \mathcal{S}_G, \preceq_{\mathcal{C}} \rangle$ with the partial order “ $\preceq_{\mathcal{C}}$ ” defined by conditions $(\mathcal{C}_{\mathcal{S}_G}.i)$ and $(\mathcal{C}_{\mathcal{S}_G}.ii)$ is a complete distributive lattice where the meet operation is*

$$\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}(\mathcal{G}) \vee \mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{G}) \vee \mathcal{E}(\mathcal{H})),$$

and the join operation is

$$\mathcal{G} \vee \mathcal{H} = (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{G}) \wedge \mathcal{E}(\mathcal{H})),$$

where \wedge and \vee denote the meet and the join operations of two partitions defined in Section 1.1.5.

The unit is the uncolored graph G . The zero is the colored graph in which there are $|V|/2$ composite vertex color classes where each class contains exactly two twin vertices, and pairs of twin edges presenting in the graph G are blocked into composite edge color classes.

Proof. We start by proving that the special case of the set of partitions of a set A , where each partition is made up of either atomic sets or composite sets including exactly two (homologous) components, forms a distributive lattice where the meet and the join are the intersection and the union of atomic sets and composite sets, respectively. Then, we apply this idea to the case of colored graphs for paired data on the given fixed uncolored graph in \mathcal{S}_G .

1. Firstly, we shall declare the special case of partitions of a set for a paired data problem, namely, there are two types of color classes: atomic color classes containing only one vertex or one edge, and composite color classes containing either two twin vertices or two twin edges, if they exist in the underlying graph. Therefore, we denote, for a partition P of a set A ,

$$\text{atom}(P) = \{\text{all atomic sets in } P\},$$

$$\text{comp}(P) = \{\text{all composite sets in } P\}.$$

This implies that $P = \text{atom}(P) \cup \text{comp}(P)$. We now show how to compute the join $P_\vee = P_1 \vee P_2$ for any $P_1, P_2 \in P(A)$; in this way, it follows that P_\vee exists. Obviously $P_1 \leq P_\vee$ and $P_2 \leq P_\vee$. By the definition of the partial order for the set of partitions of A in Section 1.1.5, we have that:

- for any $a \in \text{atom}(P_\vee)$, it holds that $a \in \text{atom}(P_1)$ and $a \in \text{atom}(P_2)$ so that

$$\text{atom}(P_\vee) \subseteq \text{atom}(P_1) \cap \text{atom}(P_2);$$

- for any $a \in \text{comp}(P_1)$, it holds that $a \in \text{comp}(P_\vee)$, and for any $a \in \text{comp}(P_2)$, it also holds that $a \in \text{comp}(P_\vee)$. Hence,

$$\text{comp}(P_\vee) \supseteq \text{comp}(P_1) \cup \text{comp}(P_2).$$

Since P_\vee is the smallest partition which is larger than both P_1 and P_2 , P_\vee is chosen as fine as possible, that is

$$\text{atom}(P_\vee) = \text{atom}(P_1) \cap \text{atom}(P_2) \quad \text{and} \quad \text{comp}(P_\vee) = \text{comp}(P_1) \cup \text{comp}(P_2). \quad (2.1)$$

Similarly, if we compute the meet $P_\wedge = P_1 \wedge P_2$ for any $P_1, P_2 \in P(A)$, then we can imply that P_\wedge exists. Clearly, $P_\wedge \leq P_1$ and $P_\wedge \leq P_2$. By the definition of the partial order \leq in $P(A)$, we have that:

- for any subset $a \in \text{atom}(P_1)$, $a \in \text{atom}(P_\wedge)$; moreover, for any subset $a \in \text{atom}(P_2)$, $a \in \text{atom}(P_\wedge)$ also. Hence, it implies that

$$\text{atom}(P_\wedge) \supseteq \text{atom}(P_1) \cup \text{atom}(P_2);$$

- for any subset $a \in \text{comp}(P_\wedge)$, $a \in \text{comp}(P_1)$ and $a \in \text{comp}(P_2)$ then

$$\text{comp}(P_\wedge) \subseteq \text{comp}(P_1) \cap \text{comp}(P_2);$$

Since P_\wedge is the largest partition which is smaller than both P_1 and P_2 , P_\wedge is chosen as coarse as possible, that is

$$\text{atom}(P_\wedge) = \text{atom}(P_1) \cup \text{atom}(P_2) \quad \text{and} \quad \text{comp}(P_\wedge) = \text{comp}(P_1) \cap \text{comp}(P_2). \quad (2.2)$$

2. We apply (2.1) and (2.2) for partitions of the vertex V and the edge set E . As a consequence, for two colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{S}_G$, the meet operation of \mathcal{G} and \mathcal{H} is the colored graph $(\mathcal{V}_\wedge, \mathcal{E}_\wedge)$ where atomic and composite classes in \mathcal{V}_\wedge and \mathcal{E}_\wedge are computed in (2.1), namely

$$\begin{aligned} \mathcal{V}_\wedge &= \mathcal{V}(\mathcal{G}) \vee \mathcal{V}(\mathcal{H}) = (\text{atom}(\mathcal{V}(\mathcal{G})) \cap \text{atom}(\mathcal{V}(\mathcal{H}))) \cup (\text{comp}(\mathcal{V}(\mathcal{G})) \cup \text{comp}(\mathcal{V}(\mathcal{H}))) \\ \mathcal{E}_\wedge &= \mathcal{E}(\mathcal{G}) \vee \mathcal{E}(\mathcal{H}) = (\text{atom}(\mathcal{E}(\mathcal{G})) \cap \text{atom}(\mathcal{E}(\mathcal{H}))) \cup (\text{comp}(\mathcal{E}(\mathcal{G})) \cup \text{comp}(\mathcal{E}(\mathcal{H}))); \end{aligned}$$

and the join operation of \mathcal{G} and \mathcal{H} is the colored graph $(\mathcal{V}_\vee, \mathcal{E}_\vee)$ where atomic and composite classes in \mathcal{V}_\vee and \mathcal{E}_\vee are computed in (2.2), namely

$$\begin{aligned}\mathcal{V}_\vee &= \mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H}) = (\text{atom}(\mathcal{V}(\mathcal{G})) \cup \text{atom}(\mathcal{V}(\mathcal{H}))) \cup (\text{comp}(\mathcal{V}(\mathcal{G})) \cap \text{comp}(\mathcal{V}(\mathcal{H}))) \\ \mathcal{E}_\vee &= \mathcal{E}(\mathcal{G}) \wedge \mathcal{E}(\mathcal{H}) = (\text{atom}(\mathcal{E}(\mathcal{G})) \cup \text{atom}(\mathcal{E}(\mathcal{H}))) \cup (\text{comp}(\mathcal{E}(\mathcal{G})) \cap \text{comp}(\mathcal{E}(\mathcal{H}))).\end{aligned}$$

The zero is the smallest colored graph $(\mathcal{V}_0, \mathcal{E}_0)$ in \mathcal{S}_G , with the vertex coloring and the edge coloring as coarse as possible, so that

$$\begin{aligned}\text{atom}(\mathcal{V}_0) &= \emptyset, \quad \text{comp}(\mathcal{V}_0) = \{\{i, j\} \text{ for all } i, j \text{ are twin in } V\}; \\ \text{comp}(\mathcal{E}_0) &= \{\{(i, j), (r, s)\} \text{ for any } (i, j), (r, s) \text{ are twin in } E\},\end{aligned}$$

and the atomic classes containing edges that are not in any subsets in $\text{comp}(\mathcal{E}_0)$. The unit is the largest colored graph $(\mathcal{V}_1, \mathcal{E}_1)$ in \mathcal{S}_G , which has the vertex coloring and the edge coloring are as fine as possible, it means that

$$\begin{aligned}\text{atom}(\mathcal{V}_1) &= \{\{i\} \text{ for all } i \in V\}, \quad \text{comp}(\mathcal{V}_1) = \emptyset; \\ \text{atom}(\mathcal{E}_1) &= \{\{(i, j)\} \text{ for all } (i, j) \in E\}, \quad \text{comp}(\mathcal{E}_1) = \emptyset.\end{aligned}$$

Moreover, in this special case, the distributive law is satisfied since the meet and the join operations are translated by the union and the intersection of atomic classes and composite classes, particularly in (2.1) and (2.2).

□

The complexity in the computation of the meet and the join for the set partition lattice, as given in (A) and (B) of Section 1.1.5, for the set of colored graphs for paired data on the given fixed uncolored graphs \mathcal{S}_G , makes their implementation not straightforward. A side results of Proposition 2.1 are more efficient ways to carry out these operations by means of the set intersection and the set union. Specifically, the following corollary gives more details about the forms of the meet and the join of PDCGs on the given fixed uncolored graph in \mathcal{S}_G .

Corollary 2.2. *Under the conditions of Proposition 2.1, the meet operation can be computed as*

$$\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}(\mathcal{G}) \vee \mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{G}) \vee \mathcal{E}(\mathcal{H}))$$

where

$$\begin{aligned}\mathcal{V}(\mathcal{G}) \vee \mathcal{V}(\mathcal{H}) &= (\text{atom}(\mathcal{V}(\mathcal{G})) \cap \text{atom}(\mathcal{V}(\mathcal{H}))) \cup (\text{comp}(\mathcal{V}(\mathcal{G})) \cup \text{comp}(\mathcal{V}(\mathcal{H}))) \\ \mathcal{E}(\mathcal{G}) \vee \mathcal{E}(\mathcal{H}) &= (\text{atom}(\mathcal{E}(\mathcal{G})) \cap \text{atom}(\mathcal{E}(\mathcal{H}))) \cup (\text{comp}(\mathcal{E}(\mathcal{G})) \cup \text{comp}(\mathcal{E}(\mathcal{H}))).\end{aligned}$$

And the join operation can be computed

$$\mathcal{G} \vee \mathcal{H} = (\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H}), \mathcal{E}(\mathcal{G}) \wedge \mathcal{E}(\mathcal{H}))$$

where

$$\begin{aligned}\mathcal{V}(\mathcal{G}) \wedge \mathcal{V}(\mathcal{H}) &= (\text{atom}(\mathcal{V}(\mathcal{G})) \cup \text{atom}(\mathcal{V}(\mathcal{H}))) \cup (\text{comp}(\mathcal{V}(\mathcal{G})) \cap \text{comp}(\mathcal{V}(\mathcal{H}))) \\ \mathcal{E}(\mathcal{G}) \wedge \mathcal{E}(\mathcal{H}) &= (\text{atom}(\mathcal{E}(\mathcal{G})) \cup \text{atom}(\mathcal{E}(\mathcal{H}))) \cup (\text{comp}(\mathcal{E}(\mathcal{G})) \cap \text{comp}(\mathcal{E}(\mathcal{H}))).\end{aligned}$$

Here, $\text{atom}(P) = \{\text{all atomic sets in the partition } P\}$ and $\text{comp}(P) = \{\text{all composite sets in the partition } P\}$.

Proof. This follows immediately from the proof of Proposition 2.1. □

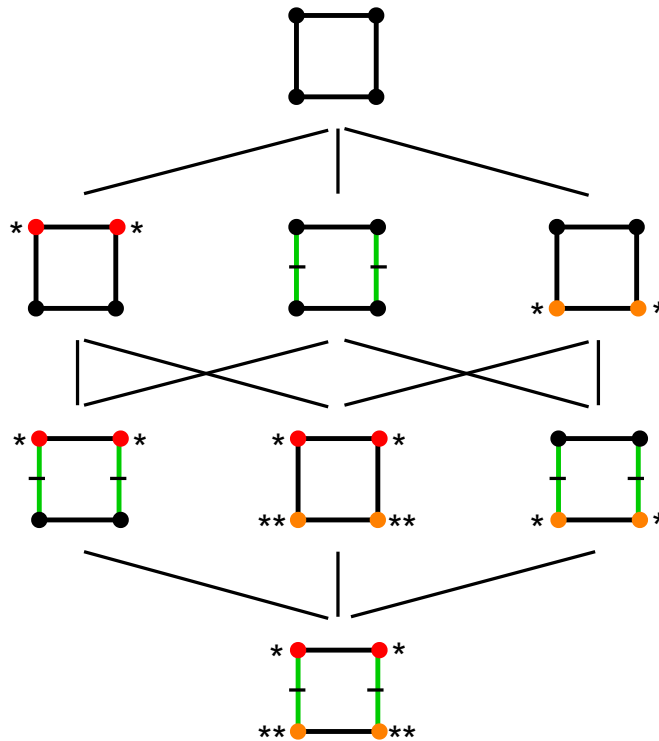


FIGURE 2.3: Hasse diagram of the lattice structure of the set of colored graphs for paired data, \mathcal{S}_G , based on the given fixed uncolored graph G , where $G = ([4], E(G))$ is represented at the top.

Example 2.3. Figure 2.4 displays an undirected graph $G = ([6], E)$ together with the zero and the unit of \mathcal{S}_G . Note that, the zero is the colored graph where all of twin vertices and twin edges are symmetric whereas the unit is the uncolored graph G where vertices and edges stand alone in atomic color classes.

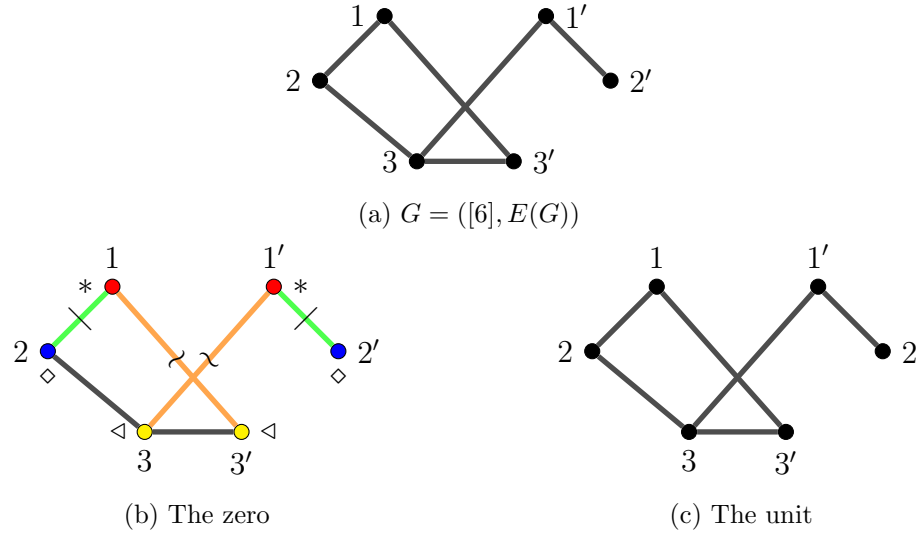


FIGURE 2.4: Representations of (b) the zero and (c) the unit on the lattice of PDCGs based on (a) the given fixed uncolored graph G , $\langle \mathcal{S}_G, \preceq_c \rangle$.

2.3.2 General case for colored graphs for paired data

We consider now the set of colored graphs for paired data on the vertex set V with no restrictions on the (uncolored) graph structure denoted by \mathcal{S}_V . This forms a proper subfamily of the family of colored graphs. Hence, we can consider the family of \mathcal{S}_V and the partial order \preceq_c . In this way, we obtain a complete lattice where the zero is the empty graph with $|V|/2$ composite vertex color classes, and each class contains exactly two twin vertices, and the unit is the uncolored complete graph. However, the lattice in this case is not distributive as shown in the following counterexamples.

Example 2.4. We consider three colored graphs for paired data \mathcal{G}, \mathcal{H} and \mathcal{K} on the vertex set $V = [6]$ shown in the first row of Figure 2.5. It is clear that the distributive law does not hold in this case because

$$\mathcal{G} \vee (\mathcal{H} \wedge \mathcal{K}) \neq (\mathcal{G} \vee \mathcal{H}) \wedge (\mathcal{G} \vee \mathcal{K})$$

demonstrated in Figure 2.5. One can prove that $\langle \mathcal{S}_V, \preceq_c \rangle$ is a non-distributive lattice.

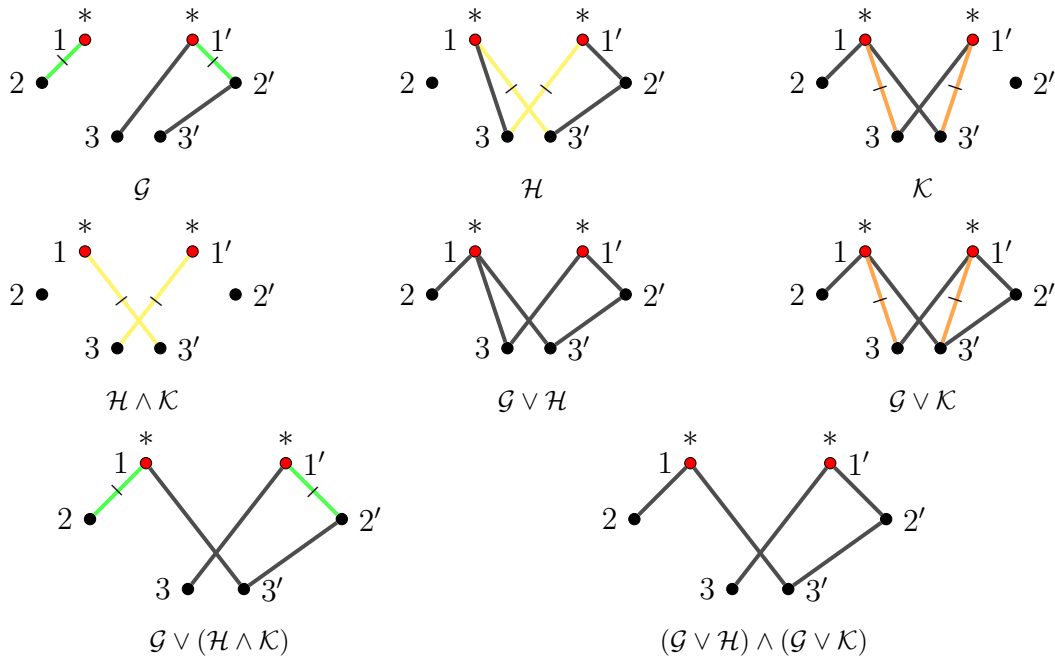


FIGURE 2.5: A counterexample of the non-distributivity of PDCGs $\mathcal{G}, \mathcal{H},$ and \mathcal{K} on the lattice $\langle \mathcal{S}_{[6]}, \preceq_c \rangle$.

Example 2.5. Figure 2.6 shows intuitively another counterexample about the non-distributivity of the lattice $\langle \mathcal{S}_V, \preceq_c \rangle$. Particularly, the diamond structure contained as a sublattice on the set of the considered PDCGs with four vertices shown in Theorem 1.1.

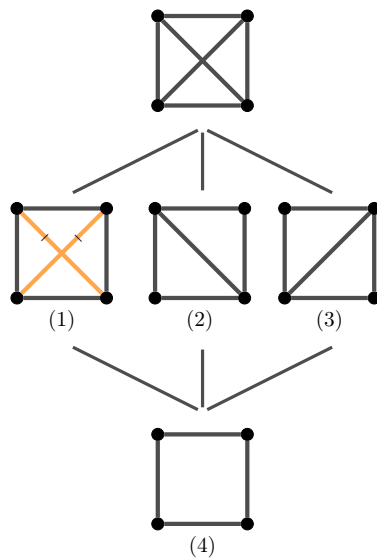


FIGURE 2.6: The diamond structure of the sublattice of $\langle \mathcal{S}_{[4]}, \preceq_c \rangle$.

2.4 The novel partial order for the set of PDCGs and its properties

In this section, we introduce some notations and terminologies that facilitate the search for a novel partial order and the subsequent proofs on the set of colored graphs for paired data. We then define a novel partial order for the class of colored graphs for paired data and show that, with such order, the family of colored graphs for paired data is a complete distributive lattice. Moreover, the connections between the model inclusion and the novel partial order are also described.

2.4.1 Notations and terminologies of important quantities

In a paired data problem, the graph G has an even number of vertices p and we set $q = p/2$. Hence, for every vertex $i \in V$ with $i \leq q$, there exists an homologous, or twin, vertex $i + q \in V$. Then, we partition the vertex set V into two disjoint subsets L and R such that

$$V = L \cup R \quad \text{with} \quad L \cap R = \emptyset,$$

where we let $L = \{1, \dots, q\}$ and $R = \{q + 1, \dots, p\}$. The application considered in this thesis concerns the paired data from the left and the right hemispheres of the brain.

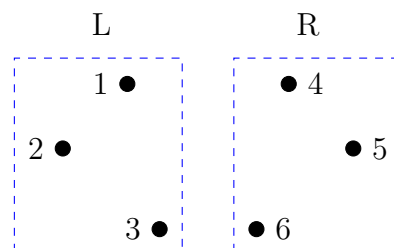


FIGURE 2.7: The vertex set $V = [6]$ is split into two disjoint subsets $L = \{1, 2, 3\}$ and $R = \{4, 5, 6\}$.

Next, we will introduce the *twin correspondence* which identifies whether or not a pair of vertices (edges) is a *twin*. More formally, the *twin correspondence*, denoted by $\tau(\cdot)$, is the function defined as

$$\tau(i) = \begin{cases} i + q, & \text{if } i = 1, \dots, q, \\ i - q, & \text{if } i = q + 1, \dots, p. \end{cases} \quad (2.3)$$

Note that, $\tau = \tau^{-1}$, in other words, $\tau(\tau(i)) = i$ for every $i \in \{1, \dots, p\}$. Furthermore, for a non-empty subset $A \subseteq V$, we write

$$\tau(A) = \{\tau(a), a \in A\}.$$

In order to label the correspondence between elements of L and R on a graph, for every $i \in L$, we put $i' = \tau(i)$ so that the set R can be written as $R = \{1', \dots, q'\}$. Therefore, by the property of $\tau(\cdot)$, for every $i \in L$, we have that

$$\tau(i) = i' \quad \text{and} \quad \tau(i') = i.$$

We say that two vertices $i, j \in V$ are *twin vertices* of G if

$$\tau(i) = j, \quad \text{or, equivalently} \quad \tau(j) = i.$$

By this definition, we can see that there is a twin correspondence τ between elements in L and elements in R which is illustrated by Figure 2.8. where

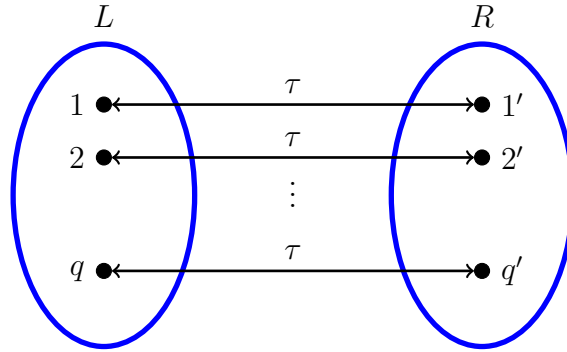


FIGURE 2.8: Relationship between the set L and the set R is presented by the twin function.

$$\tau(L) = \{\tau(1), \dots, \tau(q)\} = \{1', \dots, q'\} = R,$$

corresponding to,

$$\tau(R) = \{\tau(1'), \dots, \tau(q')\} = \{1, \dots, q\} = L.$$

Therefore, in a graph with p vertices, there will be $p/2$ pairs of twin vertices.

We also apply the function τ to ordered pairs of vertices, that is to potential edges,

$$\tau((i, j)) = (\tau(i), \tau(j)), \quad \text{for every } i, j \in V.$$

Similar to the case of vertices, we apply $\tau(\cdot)$ with sets of edges as follows. We say that a pair of edges (i, j) and (r, s) are *twin edges* if

$$\tau((i, j)) = (r, s) \quad \text{or, equivalently} \quad \tau((r, s)) = (i, j).$$

Example 2.6. In the complete graph of 6 vertices, there are six possible pairs of twin edges. Figure 2.9 displays twin edges connecting vertices within sets L and R , namely $(1, 2), (1', 2'), (1, 3), (1', 3')$ and $(2, 3), (2', 3')$, and, Figure 2.10 displays twin edges connecting vertices across L and R , namely $(1, 2'), (2, 1'), (1, 3'), (3, 1')$ and $(2, 3'), (3, 2')$. On the other hand, Figure 2.11 illustrates some pairs of edges that are not twin. Therefore, in general, on the graph of p vertices, there are $p(p-2)/4$ possible pairs of twin edges.

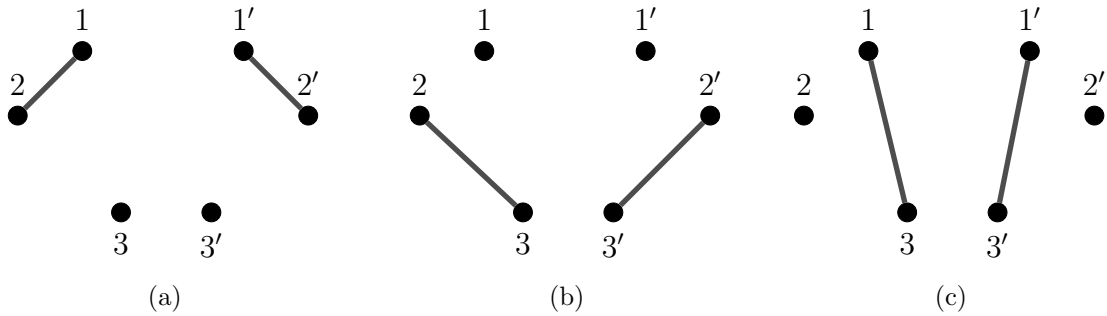


FIGURE 2.9: Pairs of twin edges connecting vertices within L and R in graph with 6 vertices, namely, (a) $(1, 2), (1', 2')$, (b) $(2, 3), (2', 3')$, and (c) $(1, 3), (1', 3')$.

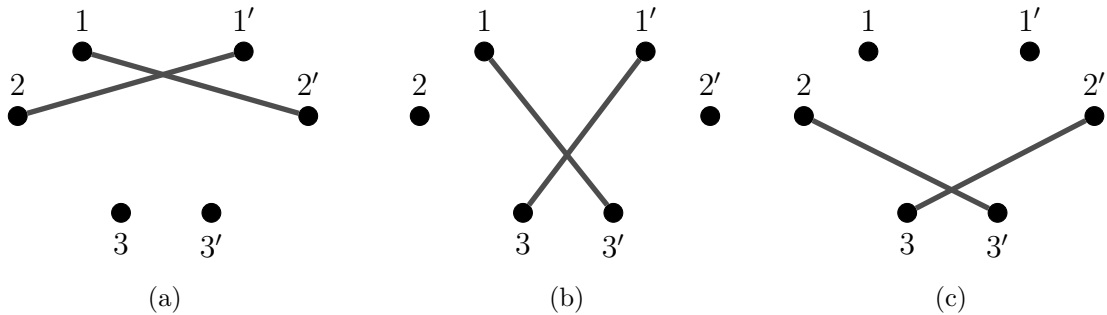


FIGURE 2.10: Pairs of twin edges connecting vertices across L and R in the graph with 6 vertices, namely, (a) $(1, 2'), (2, 1')$, (b) $(1, 3'), (3, 1')$, and (c) $(2, 3'), (3, 2')$.

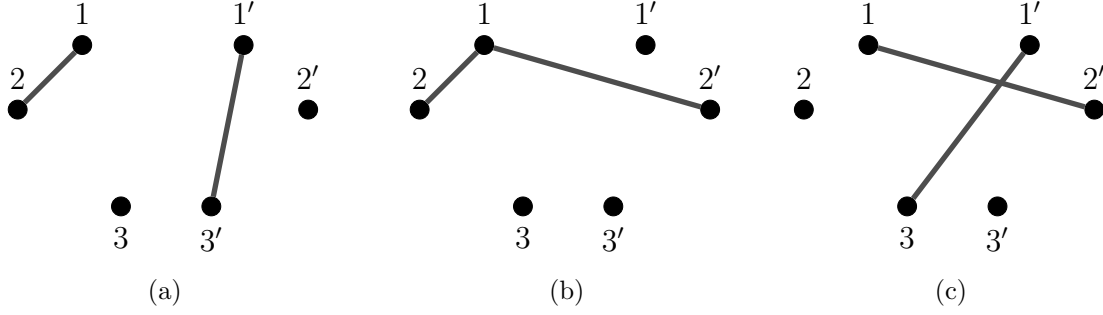


FIGURE 2.11: Examples of pairs of edges which are not twin edges, namely, (a) $(1,2), (1',3')$, (b) $(1,2), (1,2')$, and (c) $(1,2'), (3,1')$.

Let $F_V = \{(i, j) \in V \times V, i < j\}$ be a set of all possible edges in the complete graph on vertex set $V = [p]$ with $|F_V| = p(p-1)/2$. Therefore, the edge set E of a graph G will be a subset of F_V , i.e. $E \subseteq F_V$. We partition F_V into pairwise disjoint subsets F_L, F_R and F_T such that

$$F_V = F_L \cup F_R \cup F_T, \quad \text{with} \quad F_L \cap F_R = F_L \cap F_T = F_R \cap F_T = \emptyset$$

where

- F_L is the set of all possible edges “on the left”

$$F_L = \{(i, j) \in F_V \mid i, j \in L\} \cup \{(i, j) \in F_V \mid i \in L, j \in R, i < \tau(j)\}; \quad (2.4)$$

- F_R is the set of all possible edges “on the right”

$$F_R = \{(i, j) \in F_V \mid i, j \in R\} \cup \{(i, j) \in F_V \mid i \in L, j \in R, i > \tau(j)\}; \quad (2.5)$$

- and F_T is the set of all possible edges connecting twin vertices

$$F_T = \{(i, j) \in F_V \mid i \in L, j \in R, i = \tau(j)\}. \quad (2.6)$$

Note that, for $V = [p]$, it holds that

$$|F_T| = \frac{p}{2}, \quad |F_L| = |F_R| = \frac{p(p-2)}{4}.$$

Moreover, for any subset A of the edge set F_V ,

$$\tau(A) = \{(\tau(i), \tau(j)) \text{ for all } (i, j) \in A\}.$$

And hence, the sets F_L and F_R are connected by a twin correspondence, that is

$$F_L = \tau(F_R), \quad F_R = \tau(F_L). \quad (2.7)$$

which is displayed in Figure 2.12. Furthermore, it is clear that

$$\tau(F_T) = F_T.$$

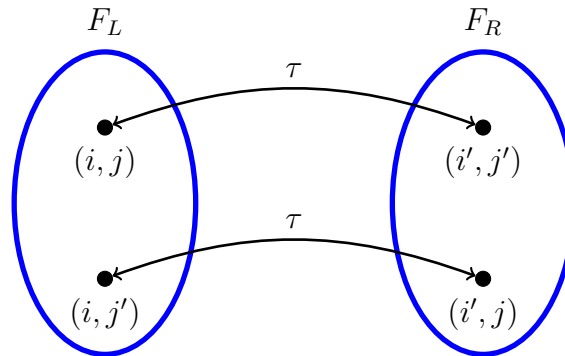


FIGURE 2.12: Relationship between the set F_L and the set F_R is presented by the twin function.

Example 2.7. We consider a complete graph where the vertex set $V = [6]$ is partitioned into $L \cup R$ where $L = \{1, 2, 3\}$ and $R = \{1', 2', 3'\}$. The set F_V is therefore determined by listing $6(6 - 1)/2 = 15$ possible edges in this graph such that $F_V = F_L \cup F_R \cup F_T$ where

$$F_L = \{(1, 2), (1, 3), (1, 2'), (1, 3'), (2, 3), (2, 3')\},$$

$$F_R = \{(1', 2'), (1', 3'), (1', 2), (1', 3), (2', 3'), (2', 3)\},$$

$$F_T = \{(1, 1'), (2, 2'), (3, 3')\},$$

which are all displayed by graphs in Figure 2.13. The connection between F_L and F_R is specified in this example as follows

$$\begin{array}{ll}
 F_L & F_R \\
 (1, 2) \xrightarrow{\tau} (1', 2'), & (1', 2') \xrightarrow{\tau} (1, 2) \\
 (1, 3) \xrightarrow{\tau} (1', 3'), & (1', 3') \xrightarrow{\tau} (1, 3) \\
 (1, 2') \xrightarrow{\tau} (2, 1'), & (2, 1') \xrightarrow{\tau} (1, 2') \\
 (1, 3') \xrightarrow{\tau} (3, 1'), & (3, 1') \xrightarrow{\tau} (1, 3') \\
 (2, 3) \xrightarrow{\tau} (2', 3'), & (2', 3') \xrightarrow{\tau} (2, 3) \\
 (2, 3') \xrightarrow{\tau} (3, 2'), & (3, 2') \xrightarrow{\tau} (2, 3').
 \end{array}$$

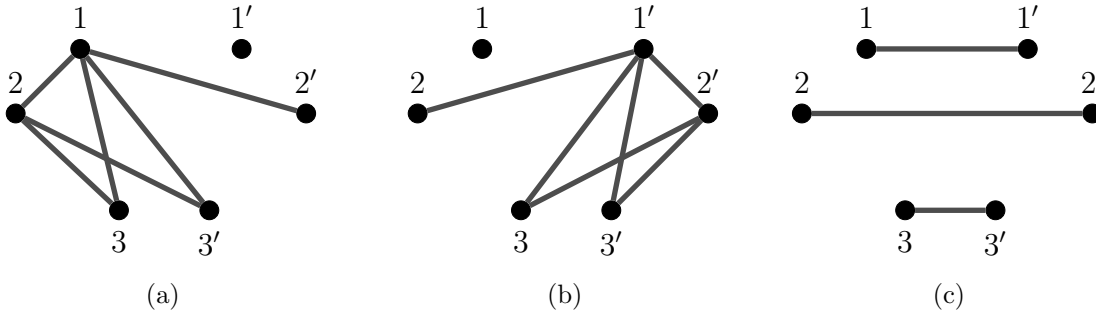


FIGURE 2.13: The partition of $F_{[6]}$ into (a) F_L , (b) F_R and (c) F_T .

For a graph $G = (V, E)$, the partition of F_V into F_L, F_R and F_T naturally induces a partition of E into

$$E_L = E \cap F_L, \quad E_R = E \cap F_R, \quad E_T = E \cap F_T, \quad (2.8)$$

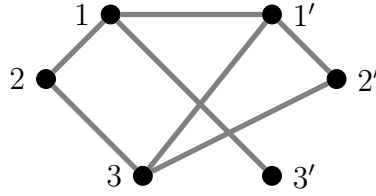
so that

$$E = E_L \cup E_R \cup E_T \quad \text{with} \quad E_L \cap E_R = E_L \cap E_T = E_R \cap E_T = \emptyset.$$

An edge (i, j) in E_L is a twin edge when $\tau((i, j))$ is an element of E_R or, equivalently, if $(i, j) \in \tau(E_R)$. Hence, the subset of E_L of edges with a twin is given by $E_L \cap \tau(E_R)$.

Example 2.8. From Figure 2.14, we get the edge set E of the graph as

$$E = \{(1, 2), (2, 3), (1, 3'), (1', 3), (1', 2'), (2', 3), (1, 1')\}$$

FIGURE 2.14: An undirected graph $G = (V, E)$.

and by (2.8), sets E_L , E_R , and E_T in this graph are determined as below

$$E_L = \{(1, 2), (2, 3), (1, 3')\},$$

$$E_R = \{(1', 2'), (3, 2'), (3, 1')\},$$

$$E_T = \{(1, 1')\}.$$

Moreover, we can see that $(1, 2)$, $(1', 2')$ and $(1, 3')$, $(3, 1')$ are pairs of twin edges shown in Figure 2.14. Therefore, if we consider edges $(1, 2)$ and $(1, 3')$, which are representations for twin edges, they are elements of the set $E_L \cap \tau(E_R)$ whereas $(2, 3)$ is not because $(2', 3')$ is missing in the graph G .

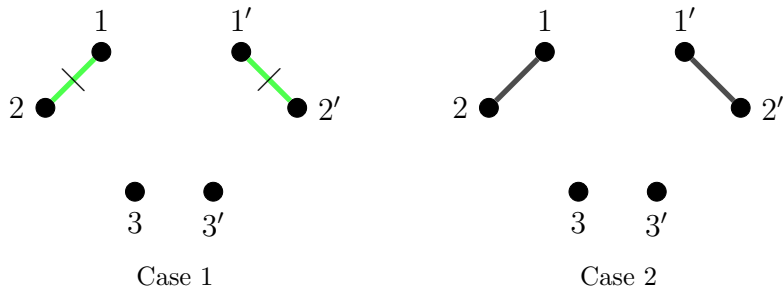
When we consider an edge $(i, j) \in E_L \cap \tau(E_R)$, there are two possibilities of the color class for (i, j) and its twin $\tau(i, j)$, that is

Case 1: $(i, j), \tau(i, j)$ form a color class, i.e.

$$\{(i, j), \tau(i, j)\} \in \mathcal{E};$$

Case 2: $(i, j), \tau(i, j)$ forms two color classes, i.e.

$$\{(i, j)\}, \{\tau(i, j)\} \in \mathcal{E}.$$



Case 1

Case 2

The set of edges $(i, j) \in E_L \cap \tau(E_R)$ which fall into the second case forms the set $\mathbb{E}_L \subseteq E_L$, more formally,

$$\mathbb{E}_L = \{(i, j) \in E_L \cap \tau(E_R) \mid \{(i, j)\} \in \mathcal{E}\}. \quad (2.9)$$

Obviously, the following sequence of inclusions hold for $\mathbb{E}_L \subseteq E_L \cap \tau(E_R) \subseteq E_L \subseteq E$. Additionally, edges in $E_L \cap \tau(E_R)$ belonging to the same edge color class with their twins are elements of the set $(E_L \cap \tau(E_R)) \setminus \mathbb{E}_L$.

A witty explanation for the use of notation \mathbb{E}_L is that the double strokes on the letter refer to twin edges which are present in a graph; and the black color on the double strokes is also the color represented for these twin edges in the graph because they belong to the atomic color classes.

Similarly, we call \mathbb{L} a set of vertices $i \in L$ under the condition that $i, \tau(i)$ are in the atomic color classes. More formally,

$$\mathbb{L} = \{i \in L \mid \{i\} \in \mathcal{V}\}, \quad (2.10)$$

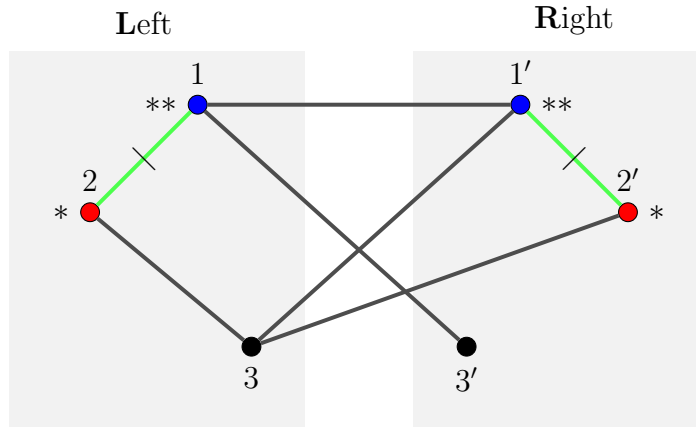
and hence, we get that $\mathbb{L} \subseteq L \subseteq V$.

Example 2.9. For the PDCG of Figure 2.16 it holds that $\mathcal{V} = \{\{1, 1'\}, \{2, 2'\}, \{3\}, \{3'\}\}$ and $\mathcal{E} = \{\{(1, 2), (1', 2')\}, \{(1, 1')\}, \{(1, 3')\}, \{(2, 3)\}, \{(3, 1')\}, \{(3, 2')\}\}$. The vertex set V is split into $L = \{1, 2, 3\}$ and $R = \{1', 2', 3'\}$. Thus, $\mathbb{L} \subseteq L$ and $\mathbb{L} = \{3\}$ since $\{3\}, \{3'\}$ forms atomic classes in \mathcal{V} , and $L \setminus \mathbb{L} = \{1, 2\}$ presents vertices that belong to the composite color classes with their twin. Furthermore, the edge set E can be obtained by merging color classes in \mathcal{E} , therefore, from E , we get that $E_L = \{(1, 2), (1, 3'), (2, 3)\}$, $E_R = \{(1', 2'), (3, 1'), (3, 2')\}$ and $E_T = \{(1, 1')\}$. The edges having their twin in the graphs form $E_L \cap \tau(E_R) = \{(1, 2), (1, 3')\}$. By convention above, \mathbb{E}_L contains edges in $E_L \cap \tau(E_R)$ that form atomic classes, namely, $\mathbb{E}_L = \{(1, 3')\}$; and, $E_L \cap \tau(E_R) \setminus \mathbb{E}_L$ refers to edges belonging to composite classes together with the twin, which equals $\{(1, 2)\}$ in this case. For the rest of atomic classes in \mathcal{E} , they can be found by $E \setminus E_L \cap \tau(E_R)$ and $E \setminus \tau(E_L) \cap E_R$.

These sets allow us to give an efficient representation of PDCGs.

Proposition 2.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a PDCG with uncolored version $G = (V, E)$. Then the sets $(V, E, \mathbb{L}, \mathbb{E}_L)$ provide an alternative, equivalent, representation of \mathcal{G} .

Proof. A PDCG \mathcal{G} is made up of two components: vertex color classes \mathcal{V} and edge color classes \mathcal{E} , such that each color class in \mathcal{V} or \mathcal{E} forms an atomic class or a composite

FIGURE 2.16: A colored graph for paired data $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

class with maximum two homologous elements. About the vertex color classes, we will show that (V, \mathbb{L}) is an one-to-one representation for \mathcal{V} . Particularly, we assume that $V = [p]$ and it is split into L and R where $L = [q]$ with $q = p/2$. By (2.10), $\mathbb{L} = \{i \in L \mid \{i\} \in \mathcal{V}\}$, or, equivalently, $\mathbb{L} = \{i \in L \mid \{i\}, \{\tau(i)\} \in \mathcal{V}\}$, which presents atomic classes in \mathcal{V} . Moreover, $L \setminus \mathbb{L} = \{i \in L \mid \{i, \tau(i)\} \in \mathcal{V}\}$ presents composite classes in \mathcal{V} .

About the edge color class, with a similar argument, we shall prove that (E, \mathbb{E}_L) is an one-to-one representation for \mathcal{E} . We firstly note that, Equation (2.9) can be written fully as $\mathbb{E}_L = \{(i, j) \in E_L \cap \tau(E_R) \mid \{(i, j)\}, \{\tau(i, j)\} \in \mathcal{E}\}$. Furthermore, if an edge $(i, j) \in E$ has its twin in the graph, \mathcal{E} contains either $\{(i, j)\}, \{\tau(i, j)\}$ or $\{(i, j), \tau(i, j)\}$; otherwise, it is presented as an atomic class in \mathcal{E} . Thus, \mathbb{E}_L , $E \setminus E_L \cap \tau(E_R)$, and $E \setminus \tau(E_L) \cap E_R$ present atomic classes and $E_L \cap \tau(E_R) \setminus \mathbb{E}_L$ refers to composite classes in \mathcal{E} , where E_L, E_R are induced from E by steps described above. \square

2.4.2 Associated lattice structure for the set of PDCGs and its structural properties

In this section, we will define the novel partial order for PDCGs on the vertex set V such that it inherits the efficient properties from the set inclusion lattice.

Definition 2.2. For two PDCGs $\mathcal{G}, \mathcal{H} \in \mathcal{S}_V$, we say $\mathcal{G} \preceq_{\tau} \mathcal{H}$ if and only if

$$(\tau.i) \quad E(\mathcal{G}) \subseteq E(\mathcal{H}),$$

$$(\tau.ii) \quad \mathbb{L}(\mathcal{G}) \subseteq \mathbb{L}(\mathcal{H}),$$

$$(\tau.iii) \quad \mathbb{E}_L(\mathcal{G}) \subseteq \mathbb{E}_L(\mathcal{H}).$$

Then, “ \preceq_τ ” defines a partial order on the set of PDCGs on the vertex set V , \mathcal{S}_V , and we call it *the twin order*.

Example 2.10. Consider two PDCGs $\mathcal{G}, \mathcal{H} \in \mathcal{S}_{[4]}$ displayed in Figure 2.17, where

Graph \mathcal{G} : $\mathbb{L}(\mathcal{G}) = \{2\}$, $E(\mathcal{G}) = \{(1, 2), (1', 2'), (1, 1'), (1, 2')\}$, $\mathbb{E}_L(\mathcal{G}) = \emptyset$;

Graph \mathcal{H} : $\mathbb{L}(\mathcal{H}) = \{1, 2\}$, $E(\mathcal{H}) = \{(1, 2), (1', 2'), (1, 1'), (1, 2'), (1', 2)\}$, $\mathbb{E}_L(\mathcal{H}) = \{(1, 2)\}$.

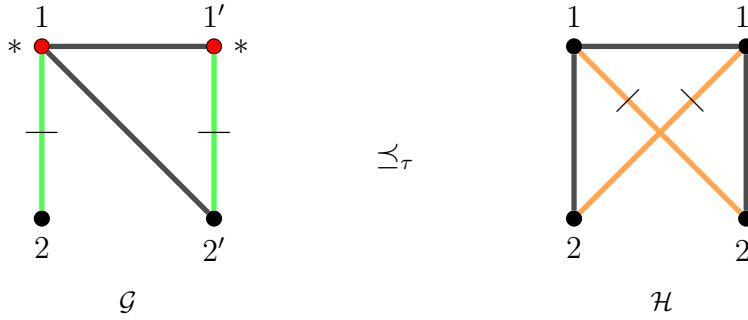


FIGURE 2.17: Example of the partial order \preceq_τ between two colored graphs for paired data \mathcal{G} and \mathcal{H} in $\mathcal{S}_{[4]}$.

As we can see, $\mathcal{G} \preceq_\tau \mathcal{H}$ because all of conditions $(\tau.i)$, $(\tau.ii)$, and $(\tau.iii)$ are satisfied for \mathcal{G} and \mathcal{H} in this case. Nevertheless, \mathcal{G} and \mathcal{H} are incomparable by the order \preceq_c since it violates the third condition, namely, the color class $\{(1, 2')\} \in \mathcal{E}(\mathcal{G})$ is not the union of $\{(1, 2'), (1', 2)\} \in \mathcal{E}(\mathcal{H})$.

If we restrict to the set of PDCGs on the given fixed uncolored version then the orders \preceq_τ and \preceq_c coincide. Precisely, this is stated in the following proposition.

Proposition 2.4. The partial order \preceq_c coincides with the partial order \preceq_τ in the class of colored graphs \mathcal{S}_G .

Proof. To show Proposition 2.4, for any pair of colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{S}_G$, we should prove that conditions defined in the partial order \preceq_c are equivalent to conditions defined in the partial order \preceq_τ . In particular,

1. $(\mathcal{C}_{S_G}.i) \implies (\tau.ii)$. If $\mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{H})$, it means that every color class in $\mathcal{V}(\mathcal{G})$ is a union of color classes in $\mathcal{V}(\mathcal{H})$, then for any vertex $i \in \mathbb{L}(\mathcal{G})$, i.e. $\{i\}, \{i'\} \in \mathcal{V}(\mathcal{G})$, it follows $i \in \mathbb{L}(\mathcal{H})$.
- $(\mathcal{C}_{S_G}.i) \iff (\tau.ii)$. $\mathcal{V}(\mathcal{G}) \geq \mathcal{V}(\mathcal{H})$ holds for those vertices in $\mathbb{L}(\mathcal{G})$ or $L \setminus \mathbb{L}(\mathcal{H})$. And in the case of $i \in \mathbb{L}(\mathcal{H}) \setminus \mathbb{L}(\mathcal{G})$, this means $\{i, i'\} \in \mathcal{V}(\mathcal{G})$ and $\{i\}, \{i'\} \in \mathcal{V}(\mathcal{H})$, so condition $(\mathcal{C}_{S_G}.i)$ is satisfied.

2. $(\mathcal{C}_{\mathcal{S}_{\mathcal{G}}}.ii) \implies (\tau.iii)$. It can be proved similarly to the previous instance of $(\mathcal{C}_{\mathcal{S}_{\mathcal{G}}}.i) \implies (\tau.ii)$.

$(\mathcal{C}_{\mathcal{S}_{\mathcal{G}}}.ii) \longleftarrow (\tau.iii)$. Since $E(\mathcal{G}) = E(\mathcal{H}) = E$, we consider the two following cases:

- (a) for any $(i, j) \notin E_L \cap \tau(E_R)$, the condition $\mathcal{E}(\mathcal{G}) \geq \mathcal{E}(\mathcal{H})$ is satisfied;
- (b) for any $(i, j) \in E_L \cap \tau(E_R)$, the condition $\mathcal{E}(\mathcal{G}) \geq \mathcal{E}(\mathcal{H})$ is also true with the same argument, as in the proof above.

□

In the general case, the lattice under the twin order \preceq_{τ} respects the model inclusion order $\preceq_{\mathcal{C}}$ as shown in the following proposition.

Proposition 2.5. *For any $\mathcal{G}, \mathcal{H} \in \mathcal{S}_V$, if $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$ then $\mathcal{G} \preceq_{\tau} \mathcal{H}$.*

Proof. Consider two colored graphs $\mathcal{G}, \mathcal{H} \in \mathcal{S}_V$ such that $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$, this means that conditions $(\mathcal{C}.i)$ - $(\mathcal{C}.iii)$ hold. We need to prove that conditions $(\tau.i)$ - $(\tau.iii)$ also hold for a pair of \mathcal{G} and \mathcal{H} . Indeed, $(\mathcal{C}.i)$ coincides with $(\tau.i)$ by the definition of the partial order of undirected graphs in Section 1.2.2. Moreover, $(\mathcal{C}.ii) \iff (\tau.ii)$ holds since it is proved in the first point of the proof of Proposition 2.4. Therefore, it is sufficient to show that if the condition $(\mathcal{C}.iii)$ is satisfied then the condition $(\tau.iii)$ is also satisfied, which is shown in the following.

$(\mathcal{C}.iii) \implies (\tau.iii)$. For any edge $(i, j) \in \mathbb{E}_L(\mathcal{G})$, it holds that (i, j) and $\tau(i, j)$ belong to $E(\mathcal{G})$ so that $\{(i, j)\}$ and $\{\tau(i, j)\}$ are atomic classes in $\mathcal{E}(\mathcal{G})$. Since $\mathcal{G} \preceq_{\mathcal{C}} \mathcal{H}$ then $(\mathcal{C}.iii)$ is true, i.e. every class in $\mathcal{E}(\mathcal{G})$ is a union of classes in $\mathcal{E}(\mathcal{H})$. It follows that (i, j) and $\tau(i, j)$ must be in $E(\mathcal{H})$ so that $\{(i, j)\}$ and $\{\tau(i, j)\}$ are also atomic classes in $\mathcal{E}(\mathcal{H})$, which implies $(i, j) \in \mathbb{E}_L(\mathcal{H})$. Hence, we can conclude that $\mathbb{E}_L(\mathcal{G}) \subseteq \mathbb{E}_L(\mathcal{H})$. □

With the conditions $(\tau.i)$, $(\tau.ii)$, and $(\tau.iii)$ in the definition of the partial order in $\langle \mathcal{S}_V, \preceq_{\tau} \rangle$ defined above, we have the following results.

Theorem 2.6. *The family of PDCGs on the vertex set V , \mathcal{S}_V , equipped with the partial order \preceq_{τ} is a complete distributive lattice where*

1. *the meet operation is the colored graph*

$$\mathcal{G} \wedge \mathcal{H} = (V, E(\mathcal{G}) \cap E(\mathcal{H}), \mathbb{L}(\mathcal{G}) \cap \mathbb{L}(\mathcal{H}), \mathbb{E}_L(\mathcal{G}) \cap \mathbb{E}_L(\mathcal{H}));$$

2. *the join operation is the colored graph*

$$\mathcal{G} \vee \mathcal{H} = (V, E(\mathcal{G}) \cup E(\mathcal{H}), \mathbb{L}(\mathcal{G}) \cup \mathbb{L}(\mathcal{H}), \mathbb{E}_L(\mathcal{G}) \cup \mathbb{E}_L(\mathcal{H}));$$

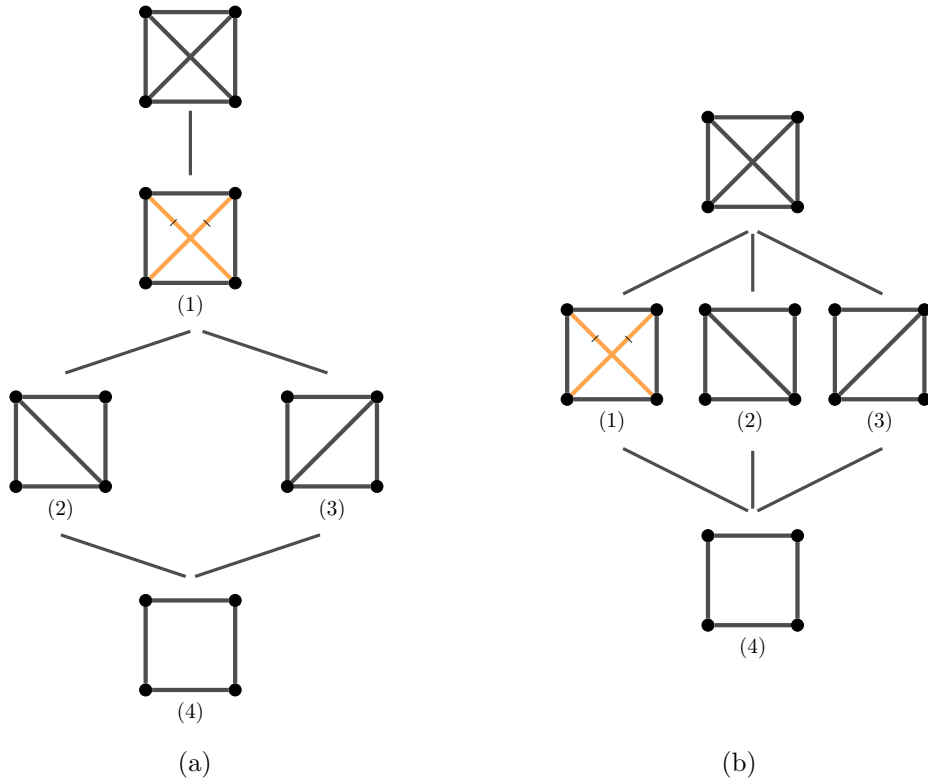


FIGURE 2.18: Sublattices of colored graphs for paired data in \mathcal{S}_V ordered by (a) the twin order \preceq_τ and (b) the model inclusion order \preceq_C .

- 3. the zero is an empty graph with $\mathbb{L}(\hat{0}) = \emptyset$ and $\mathbb{E}_L(\hat{0}) = \emptyset$, and
- 4. the unit is a complete graph with atomic color classes in vertices and edges, i.e. $\mathbb{L}(\hat{1}) = L$ and $\mathbb{E}_L(\hat{1}) = F_L$.

We call the lattice \mathcal{S}_V with the twin order \preceq_τ the twin lattice.

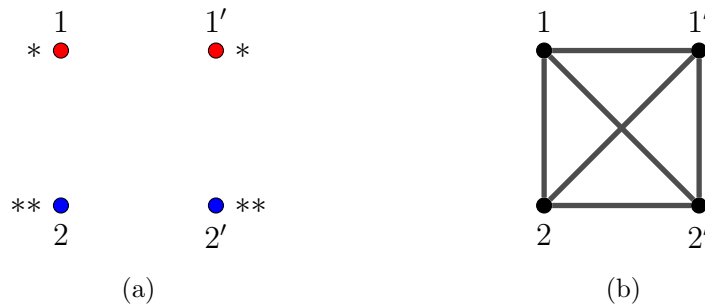


FIGURE 2.19: Colored graphs represent (a) the zero, and (b) the unit in $\langle \mathcal{S}_{[4]}, \preceq_\tau \rangle$.

Proof. The proof of the theorem is designed as follows: first, we prove $\langle \mathcal{S}_V, \preceq_\tau \rangle$ is a lattice by determining the explicit forms of the corresponding meet and join operations, where the meet and the join are defined by the set intersection and the set union

operations, respectively. As a consequence, we find the zero and the unit on the lattice structure of \mathcal{S}_V , from which the completeness holds. Finally, we prove that the lattice $\langle \mathcal{S}_V, \preceq_\tau \rangle$ is distributive since the meet and join operations distribute to each other.

1. Let $\mathcal{G} \in \mathcal{S}_V$ be a colored graph such that

$$\mathbb{L}(\mathcal{G}) = \mathbb{L}(\mathcal{G}_1) \cap \mathbb{L}(\mathcal{G}_2), \quad E(\mathcal{G}) = E(\mathcal{G}_1) \cap E(\mathcal{G}_2), \quad \mathbb{E}_L(\mathcal{G}) = \mathbb{E}_L(\mathcal{G}_1) \cap \mathbb{E}_L(\mathcal{G}_2),$$

for any $\mathcal{G}_1, \mathcal{G}_2 \in \mathcal{S}_V$. Obviously, \mathcal{G} is a lower bound of \mathcal{G}_1 and \mathcal{G}_2 according to the \preceq_τ relation. Suppose that \mathcal{G}' is the infimum of \mathcal{G}_1 and \mathcal{G}_2 , i.e. $\mathcal{G}' = \mathcal{G}_1 \wedge \mathcal{G}_2$. This means that \mathcal{G}' is a lower bound of $\mathcal{G}_1, \mathcal{G}_2$ and $\mathcal{G} \preceq_\tau \mathcal{G}'$. Moreover, since $\mathcal{G}' \preceq_\tau \mathcal{G}_1$ and $\mathcal{G}' \preceq_\tau \mathcal{G}_2$ then by the definition of the partial order \preceq_τ on \mathcal{S}_V , we have

- (a) $\mathbb{L}(\mathcal{G}') \subseteq \mathbb{L}(\mathcal{G}_1)$ and $\mathbb{L}(\mathcal{G}') \subseteq \mathbb{L}(\mathcal{G}_2)$,
- (b) $E(\mathcal{G}') \subseteq E(\mathcal{G}_1)$ and $E(\mathcal{G}') \subseteq E(\mathcal{G}_2)$,
- (c) $\mathbb{E}_L(\mathcal{G}') \subseteq \mathbb{E}_L(\mathcal{G}_1)$ and $\mathbb{E}_L(\mathcal{G}') \subseteq \mathbb{E}_L(\mathcal{G}_2)$,

which implies that

$$\mathbb{L}(\mathcal{G}') \subseteq \mathbb{L}(\mathcal{G}_1) \cap \mathbb{L}(\mathcal{G}_2), \quad E(\mathcal{G}') \subseteq E(\mathcal{G}_1) \cap E(\mathcal{G}_2), \quad \mathbb{E}_L(\mathcal{G}') \subseteq \mathbb{E}_L(\mathcal{G}_1) \cap \mathbb{E}_L(\mathcal{G}_2).$$

Hence, $\mathcal{G}' \preceq_\tau \mathcal{G}$. By the antisymmetry of the partial order, we get that $\mathcal{G}' \equiv \mathcal{G}$.

2. A similar argument can be applied for the proof of the join operation. Particularly, let $\mathcal{G} \in \mathcal{S}_V$ be a colored graph such that

$$\mathbb{L}(\mathcal{G}) = \mathbb{L}(\mathcal{G}_1) \cup \mathbb{L}(\mathcal{G}_2), \quad E(\mathcal{G}) = E(\mathcal{G}_1) \cup E(\mathcal{G}_2), \quad \mathbb{E}_L(\mathcal{G}) = \mathbb{E}_L(\mathcal{G}_1) \cup \mathbb{E}_L(\mathcal{G}_2),$$

for any $\mathcal{G}_1, \mathcal{G}_2 \in \mathcal{S}_V$. Obviously, \mathcal{G} is an upper bound of \mathcal{G}_1 and \mathcal{G}_2 by \preceq_τ relation. Suppose that \mathcal{G}' is the supremum of \mathcal{G}_1 and \mathcal{G}_2 , i.e. $\mathcal{G}' = \mathcal{G}_1 \vee \mathcal{G}_2$. This means that \mathcal{G}' is an upper bound of $\mathcal{G}_1, \mathcal{G}_2$ and $\mathcal{G}' \preceq_\tau \mathcal{G}$. Moreover, since $\mathcal{G}_1 \preceq_\tau \mathcal{G}'$ and $\mathcal{G}_2 \preceq_\tau \mathcal{G}'$ then by the definition of the partial order \preceq_τ on \mathcal{S}_V , we have

- (a) $\mathbb{L}(\mathcal{G}') \supseteq \mathbb{L}(\mathcal{G}_1)$ and $\mathbb{L}(\mathcal{G}') \supseteq \mathbb{L}(\mathcal{G}_2)$,
- (b) $E(\mathcal{G}') \supseteq E(\mathcal{G}_1)$ and $E(\mathcal{G}') \supseteq E(\mathcal{G}_2)$,
- (c) $\mathbb{E}_L(\mathcal{G}') \supseteq \mathbb{E}_L(\mathcal{G}_1)$ and $\mathbb{E}_L(\mathcal{G}') \supseteq \mathbb{E}_L(\mathcal{G}_2)$,

which implies that

$$\mathbb{L}(\mathcal{G}') \supseteq \mathbb{L}(\mathcal{G}_1) \cup \mathbb{L}(\mathcal{G}_2), \quad E(\mathcal{G}') \supseteq E(\mathcal{G}_1) \cup E(\mathcal{G}_2), \quad \mathbb{E}_L(\mathcal{G}') \supseteq \mathbb{E}_L(\mathcal{G}_1) \cup \mathbb{E}_L(\mathcal{G}_2).$$

Hence, $\mathcal{G} \preceq_{\tau} \mathcal{G}'$. By the antisymmetry of the partial order, we get that $\mathcal{G}' \equiv \mathcal{G}$.

3. The zero is the smallest element in \mathcal{S}_V , if it exists, that is

$$\mathbb{L}(\hat{0}) \subseteq \bigcap_{\mathcal{G} \in \mathcal{S}_V} \mathbb{L}(\mathcal{G}) = \emptyset, \quad E(\hat{0}) \subseteq \bigcap_{\mathcal{G} \in \mathcal{S}_V} E(\mathcal{G}) = \emptyset.$$

Moreover, the sets $\mathbb{L}(\hat{0}) \supseteq \emptyset$ and $E(\hat{0}) \supseteq \emptyset$, therefore, $\mathbb{L}(\hat{0}) = \emptyset$ and $E(\hat{0}) = \emptyset$, which also implies that $\mathbb{E}_L(\hat{0}) = \emptyset$.

4. The unit is the largest element in \mathcal{S}_V , if it exists, that is

$$\mathbb{L}(\hat{1}) \supseteq \bigcup_{\mathcal{G} \in \mathcal{S}_V} \mathbb{L}(\mathcal{G}) = L, \quad E(\hat{1}) \supseteq \bigcap_{\mathcal{G} \in \mathcal{S}_V} E(\mathcal{G}) = F_V$$

where $\mathbb{E}_L(\hat{1})$ is as large as possible, i.e.

$$\mathbb{E}_L(\hat{1}) \supseteq \bigcup_{\mathcal{G} \in \mathcal{S}_V} \mathbb{E}_L(\mathcal{G}) = F_V.$$

Since $\mathbb{L}(\mathcal{G}) \subseteq L$ for all $\mathcal{G} \in \mathcal{S}_V$ then $\mathbb{L}(\hat{1}) = L$. Similarly, $E(\hat{1}) = F_V$ with the number of asymmetric homologous edges as large as possible, i.e. $\mathbb{E}_L(\hat{1}) = F_L$.

Therefore, $\langle \mathcal{S}_V, \preceq_{\tau} \rangle$ forms a complete lattice where the distributivity law can be verified by the \wedge, \vee operations through the intersection and the union of sets \mathbb{L}, E and \mathbb{E}_L . \square

Chapter 3

Model search over the twin lattice

The selection of colored graphical models, given some data, is a major challenge because the number of different models grows super-exponentially with the number of variables. In this chapter, we focus on greedy search methods which perform local moves on the lattice structure of the model space, as efficiently as possible. We implement a backward elimination stepwise procedure on the twin lattice, and show that it is more efficient than an equivalent procedure on the model inclusion lattice. Furthermore, we show that the use of the twin lattice allows us to avoid an incoherent step in model search, which will be explained in Section 3.2. The performance of the proposed procedure is evaluated on simulated data and applied to the identification of a brain network from functional MRI data.

3.1 Dimensions of the search spaces and related works

In this section, we will compare the dimensions of the model spaces of uncolored graphical models with colored graphical models for paired data. The results of [Gehrmann \(2011\)](#) for the general family of colored graphs show the super-exponential growth of the dimension of the colored graph space. For example, when p is 4 or 6, the number of uncolored graphs is 64 and 32,768, respectively, whereas the number of colored graphs is 13,155 and 2.127469×10^{12} , instead. In our context, we consider colored graphs for paired data which is a subset of general colored graphs. A general formula for the dimension of such space is not available, however, with $p = 4$ or $p = 6$, the numbers of colored graphs for paired data are 400 and 1,000,216, respectively. An easier task is to compare the number of perfectly-paired uncolored graphs with that of perfectly-paired PDCGs. In this case, for a given number of vertices p , there are $p(p-2)/4$ possible pairs of twin edges where each pair has two possibilities of colorings. Therefore, the number

of perfectly-paired colored graphs for paired data is given by the product of the number of colorings on twin vertices, $2^{p/2}$, multiplied by the sum of the number of colorings on k pairs of twin edges times the number of uncolored graphs having exactly k pairs of twin edges on the edge set E , which is $2^k \binom{\frac{p(p-2)}{4}}{k}$. More formally, it is specified as

$$2^{p/2} \sum_{k=0}^{p(p-2)/4} 2^k \binom{\frac{p(p-2)}{4}}{k},$$

whereas the number of perfectly-paired uncolored graphs with p vertices can be computed as

$$\sum_{k=0}^{p(p-2)/4} \binom{\frac{p(p-2)}{4}}{k} = 2^{p(p-2)/4}.$$

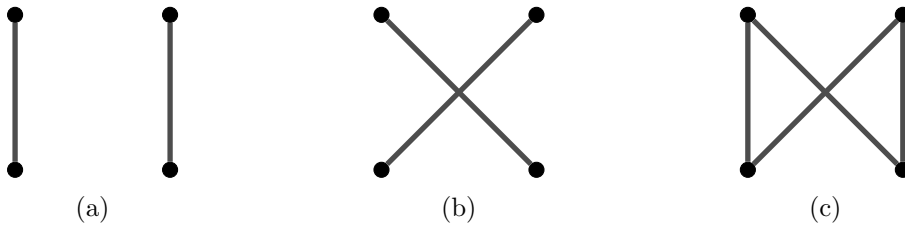


FIGURE 3.1: Perfectly-paired uncolored graphs with (a)-(b) one pair of homologous edges, and (c) two pairs of homologous edges on the vertex set of 4 vertices.

Figure 3.2 shows the comparison on the logarithmic scales of the number of “perfectly-paired” uncolored graphs in black, and the number of “perfectly-paired” colored graphs for paired data in blue when p varies from 2 to 20 vertices. Intuitively, we found that the number of graphs from both of such cases grows super-exponentially with the number of vertices p ; however, the number of “perfectly-paired” PDCGs grows faster than the number of “perfectly-paired” uncolored graphs.

It is also interesting to consider the complete graph. Indeed, there is only one complete graph on p vertices, but the number of possible colorings of the twin vertices and the twin edges on the complete graph is given by

$$2^{\frac{p}{2} + \frac{p(p-2)}{4}} = 2^{\frac{p^2}{4}},$$

and it is clear that also in this case the number of PDCGs on the complete graph grows super-exponentially with the number of vertices. Figure 3.3 gives the logarithmic scale of numbers of colored graphs for paired data where the case of the complete uncolored graphs is presented in red and dots and the case of the perfectly-paired uncolored graphs is presented in blue. The dimensions of PDCGs based on two such uncolored graphs are computed with different number of vertices p from 2 to 20.

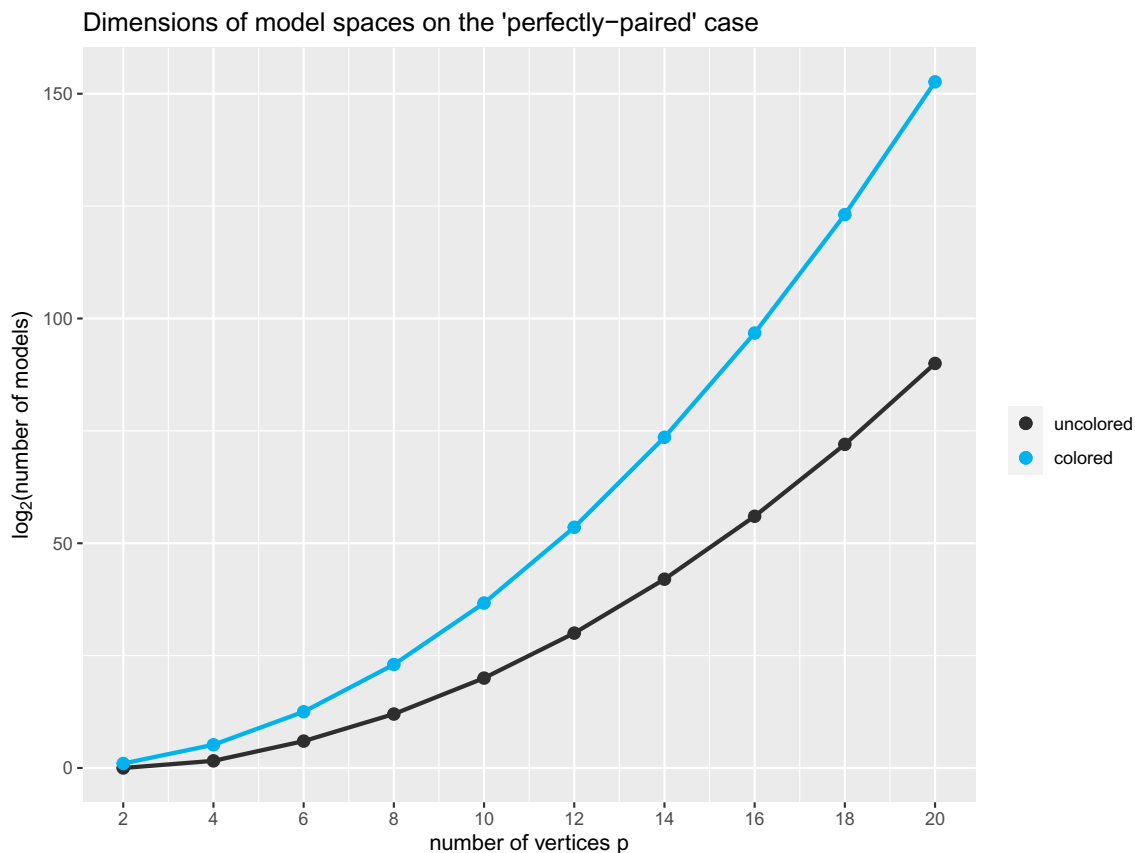


FIGURE 3.2: Logarithmic scale of the number of “perfectly-paired” (uncolored) graphical models (illustrated in black) and the number of “perfectly-paired” colored graphical models for paired data (illustrated in blue) with different even number of vertices p from 2 to 20.

The above discussion on the dimension of the model space shows that, for the family of colored graphical models, model search is much more challenging than for the family of (uncolored) undirected graphical models. Indeed, model search of colored graphical models is a relatively recent area of research, and the computational complexity of the existing procedures is such that they can be applied only to a much smaller number of variables with respect to the existing procedures for undirected graphical models.

In this thesis, we focus on greedy search procedures. Within this framework, [Gehrmann \(2011\)](#) investigated the properties of the model inclusion lattice of colored graphical models and implemented greedy search procedures for four relevant subfamilies of colored Gaussian graphical models to exploit the lattice structure with respect to model inclusion. The examples considered in their work involve at most models with 5 variables.

Although we focus on greedy search procedures, the structure of model space can be useful also under other approaches. For example, in a Bayesian framework, [Li et al. \(2020\)](#) explored the search space of RCON models by the local moves, which combines

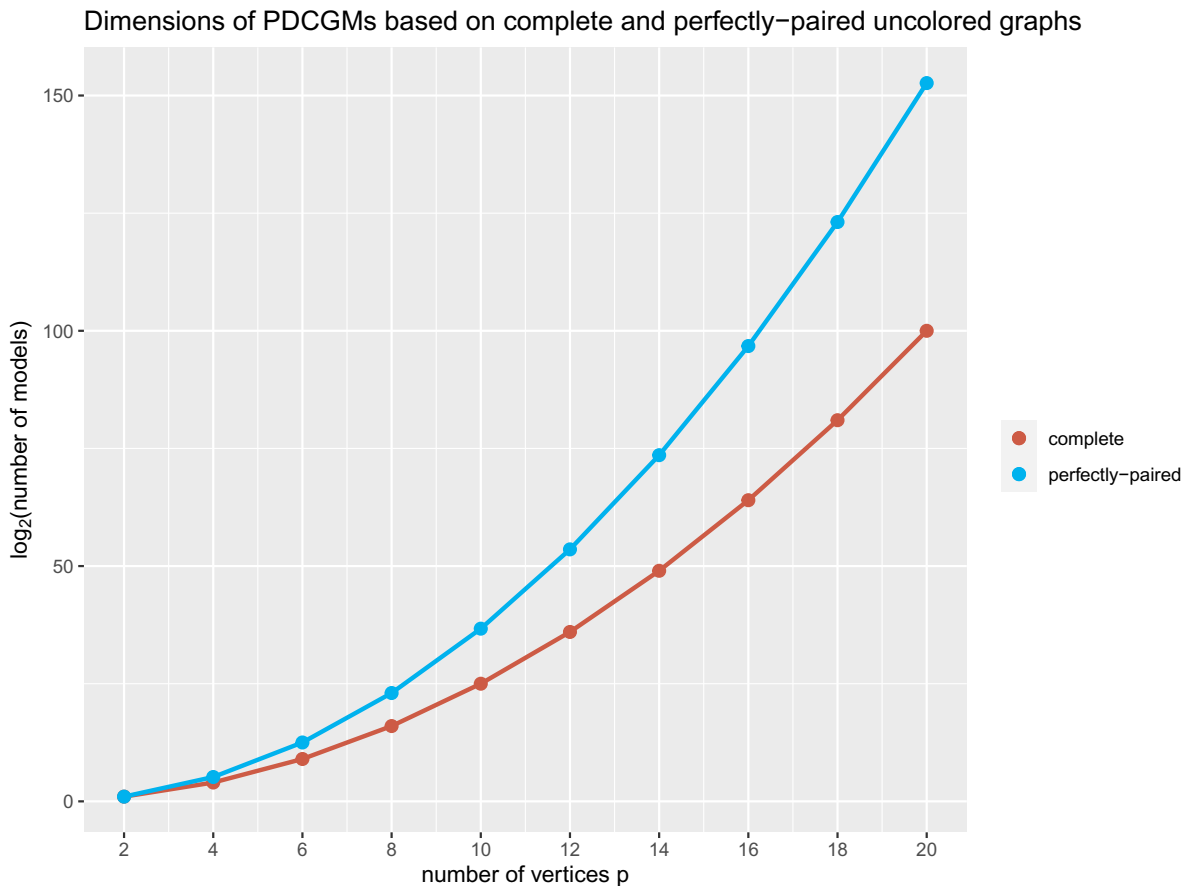


FIGURE 3.3: Logarithmic scale of numbers of colored graphical models for paired data represented by the complete uncolored graphs (illustrated in red), and the “perfectly-paired” uncolored graphs (illustrated in blue), with different even number of vertices p from 2 to 20.

the linear regression with a double reversible jump Markov chain Monte Carlo method. This approach is conducted for data set with up to 11 variables. However, in order to reduce the computational complexity, some simplifying assumptions are required.

Alternatively, [Ranciati *et al.* \(2021\)](#) proposed the penalized likelihood method with a fused type penalty function, introduced in [Tibshirani *et al.* \(2005\)](#), to learn multiple networks in a context of dependent samples. Specifically, they applied their method to a problem involving 70 variables. Nevertheless, this method cannot be directly compared with ours because we also consider symmetries involving edges across the two groups. For other works on the application of penalized likelihood methods on colored graphical models, we see also [Vinciotti *et al.* \(2016\)](#) and [Li *et al.* \(2021\)](#).

3.2 Principle of coherence

In this section, we will discuss about the coherence principle in the model testing and investigate the coherent steps through two relevant examples of the lattices of PDCGMs. The *principle of coherence* was introduced in [Gabriel \(1969\)](#) where it is stated that: “in any procedure involving multiple comparisons no hypothesis should be accepted if any hypothesis implied by it is rejected”. The implementation of the coherence principle in model search procedures is usually done by stating that “a procedure involving testing a set of model ought not accept a model while rejecting a more general model” (see [Edwards and Havránek 1987](#)). For simplification, we will say the “accepted” model instead of the “non-rejected” model. Hence, if a model \mathcal{M} is rejected then all its sub-models should be also rejected. For this reason, the coherence is naturally implemented in model inclusion lattices which make explicit the submodel relation. A great practical advantage on the usage of the coherence principle is to reduce the number of hypotheses needed to be tested. Thereby, on the construction of a feasible model selection procedure, it allows us to reduce considerably the number of tested models in the searching path. Although coherence is not the only one desideratum property (see e.g. [Izbicki and Esteves \(2015\)](#), [Bickel and Patriota \(2019\)](#) for other properties), it is by far the most emphasized one in the literature of multiple hypothesis testing (see e.g. [Gabriel \(1969\)](#), [Antoch and Hanousek \(2001\)](#), [Sonnemann \(2008\)](#), [Gehrmann \(2011\)](#), [Edwards \(2012, chapter 6\)](#), [Patriota \(2013\)](#), and references therein). It is therefore both theoretically important and practically convenient to use the coherence principle in model search procedures. Furthermore, the implementation of this principle strongly depends on the type of models considered and, in the case we are considering, on the lattice structure used to encode the model space. In the following example, we show that, for the family of PDCGMs, the implementation of the coherent step is subtler than in traditional (un-colored) undirected graphical models, and one cannot deal with it by only considering model inclusion relation.

Example 3.1. *We apply Edwards-Havránek method on the sublattice structure of RCON models represented by the colored graphs for paired data with the set of 4 vertices, under the model inclusion displayed in [Figure 3.4](#). The aim of this procedure is to identify the simplest models that are accepted. In this context, we will test four hypotheses for models represented respectively by graphs (1), (2), (3), and (4). We assume that at the significance level α , by the chi-square approximations for the distributions of the corresponding likelihood ratio test relative to the saturated model, we reject the hypothesis for (1) and accept simultaneously hypotheses for (2) and (3). Hence, this is not coherent*

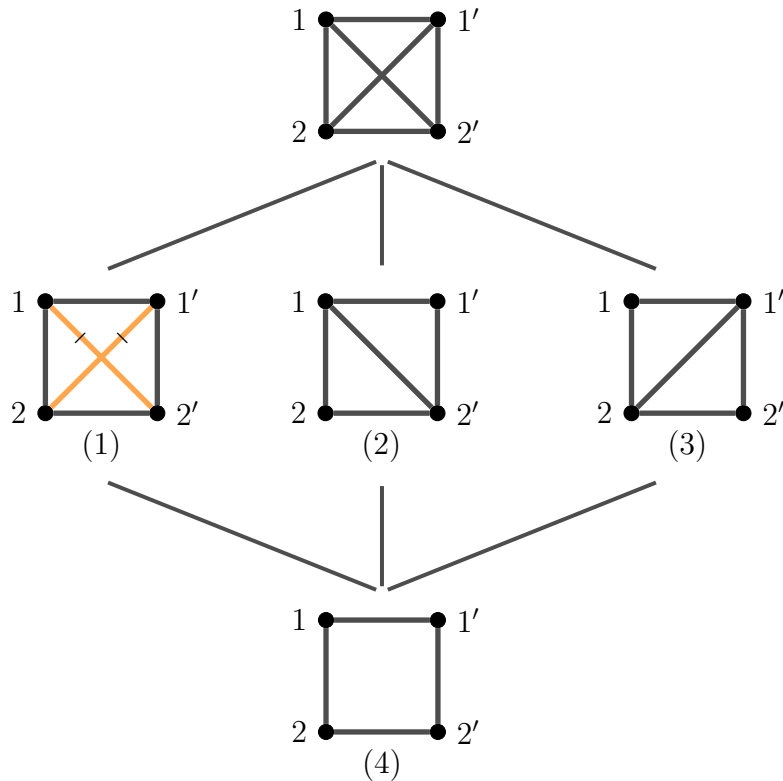


FIGURE 3.4: Sublattice structure of colored graphs for paired data with the set of 4 vertices under the model inclusion $\preceq_{\mathcal{C}}$.

in the sense that we can conclude that on the concentration matrix the corresponding parameters $\theta_{21'} = 0$ and $\theta_{12'} = 0$ but cannot accept that they have the same value.

Figure 3.5 gives the sublattice according to the twin order for 5 PDCGMs represented by the colored graphs of Figure 3.4. Notice that, models (1), (2), and (3) are now arranged in a two-layer structure. Thus, in the backward elimination stepwise procedure we will test model (1) first then apply the appropriate coherent step for models (2) and (3). In particular, in this procedure, at one-step movement, if the model testing for (1) is accepted then we will test the model represented by (4). It makes sense if the testing of the model represented by (4) is rejected then we can implicitly understand that models represented by (2) and (3) are both simultaneously rejected; otherwise models represented by (2) and (3) are both simultaneously accepted. On the other hand, if the testing model for (1) is rejected then we will further test (2) and (3) and reject (4). In this case, it also makes sense if the tests for (2) and (3) are both rejected or one of them is accepted.

The rest of this chapter is devoted to the implementation of a backward elimination stepwise procedure for the family of PDCGMs on the lattice under the twin order \preceq_{τ} .

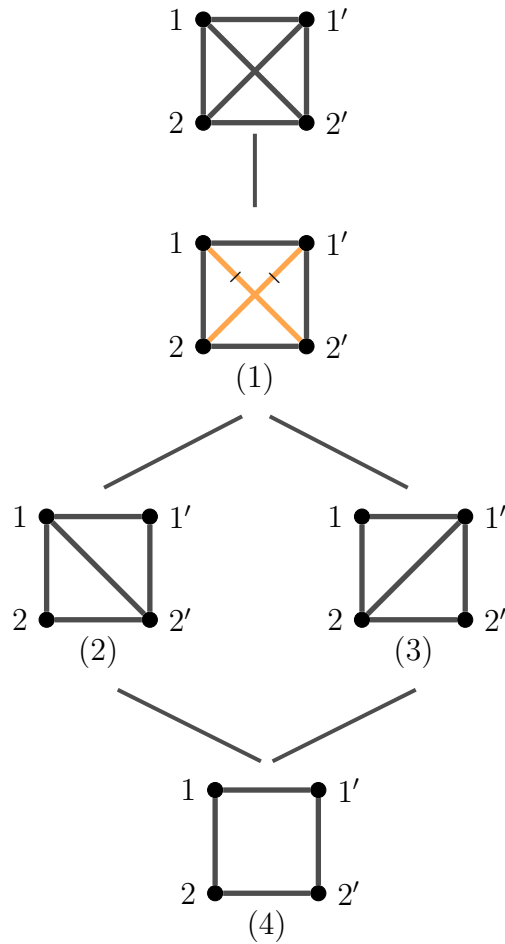


FIGURE 3.5: Sublattice structure of colored graphs for paired data with the set of 4 vertices under the twin order \preceq_{τ} .

3.3 Greedy search on the twin lattice

In this section, we will describe step-by-step the backward elimination stepwise procedure with the coherent moves on the lattice of PDCGMs under the twin order \preceq_{τ} , defined in Section 2.4.2. More specifically, we consider the family of colored graphical models for paired data restricting the equality constraints on the concentration matrix (see Højsgaard and Lauritzen 2008) when a statistical test at level α is applied to decide whether a model is rejected or non-rejected. Remark that, we will say the “accepted” model instead of the “non-rejected” model. We remark that backward elimination stepwise procedures on the model inclusion lattice have two main drawbacks:

1. they implicitly involve an incoherent step as shown in Section 3.2,
2. in order to explore the lattice structure, one has to obtain neighboring models computed by applying the meet operation which, inherited from the set partition lattice, is computationally demanding.

Therefore, we propose a method that adapts the idea from the backward elimination as well as respects the principle of coherence such that we can employ convenient properties from the twin order \preceq_τ in the model search. We can visually compare the twin lattice with the model inclusion lattice by noticing that, in the Hasse diagram representation of the model inclusion lattice, the model on the top is the saturated model and every model \mathcal{M} is directly linked to a set of nearest submodels, which are \preceq_C incomparable and all of them have the same dimension. We will refer to these models as to the *neighbor submodels of \mathcal{M}* . The twin lattice “refines” this structure by splitting each set of neighbor submodels into an upper layer and a lower layer, as in Figure 2.18. Accordingly, our procedure specifies coherent moves between these two layers.

3.3.1 The backward elimination stepwise procedure on the twin lattice

Particularly, we start from the saturated model, and set the rejection sets associated to the upper and lower layers, respectively \mathcal{R}_1 and \mathcal{R}_2 , are equal to the empty set. At each stage of the algorithm, for the upper layer, we only consider models that are \preceq_τ incomparable with models in \mathcal{R}_1 , then, by the likelihood ratio testing at level α , we update the rejection set \mathcal{R}_1 by adding rejected models to it. For the lower layer, we only consider models that are \preceq_τ incomparable with models in \mathcal{R}_2 and \preceq_τ smaller than models having any symmetric twin edges in \mathcal{R}_1 . Similarly, by the likelihood ratio testing at level α , we update the rejection set \mathcal{R}_2 by adding rejected models to it. Finally, the locally optimal model is chosen among models that do not belong to $\mathcal{R}_1 \cup \mathcal{R}_2$. We repeat the procedure until the searching of the algorithm in the iterative manner exceeds the maximum number of iterations. This number is optional. Furthermore, inside the loop of the procedure, the algorithm will stop when models in the set of neighbors at the current stage are all rejected. The procedure is described in detail below.

Procedure A: Stepwise backward elimination procedure for PDCGMs under the twin order \preceq_τ

1. Start from the saturated model: $\mathcal{M}^* \leftarrow$ saturated model;
2. Set $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ where $\mathcal{R}_1 = \mathcal{R}_2 = \emptyset$;
3. Split the set of neighbor submodels of \mathcal{M}^* into two layers according to \preceq_τ ;
4. **The upper layer:**

4.a The upper layer of the neighborhood of \mathcal{M}^* is denoted by $\mathcal{N}_1(\mathcal{M}^*)$, which can be obtained as follows:

- remove one vertex from \mathbb{L} at a time;
- remove one edge from \mathbb{E}_L at a time;
- remove one edge $e \in E$ such that $e \equiv \tau(e)$;
- remove both $e, \tau(e) \in E$ at a time if $e \notin \mathbb{E}_L$.

Formally, $\mathcal{N}_1(\mathcal{M}^*) = \{\mathcal{M} \in \mathcal{S}_V : \mathcal{M} \prec_\tau \mathcal{M}^*\}$, roughly speaking, $\mathcal{N}_1(\mathcal{M}^*)$ contains models \mathcal{M} in \mathcal{S}_V such that \mathcal{M} is covered by \mathcal{M}^* through the twin order \preceq_τ ;

4.b The weakly rejected models in this layer are models in $\mathcal{N}_1(\mathcal{M}^*)$ such that they are submodels of any model in \mathcal{R}_1 , i.e.

$$\{\mathcal{M} \in \mathcal{N}_1(\mathcal{M}^*) \text{ s.t. } \mathcal{M} \preceq_\tau \mathcal{M}_r \text{ for any } \mathcal{M}_r \in \mathcal{R}_1\};$$

4.c $\mathcal{N}'_1(\mathcal{M}^*) \leftarrow \mathcal{N}_1(\mathcal{M}^*) \setminus \{\text{weakly rejected models in step 4.b}\}$;

4.d Test models in $\mathcal{N}'_1(\mathcal{M}^*)$ and remove from $\mathcal{N}'_1(\mathcal{M}^*)$ the rejected models:

$$\mathcal{A}_1(\mathcal{M}^*) \leftarrow \mathcal{N}'_1(\mathcal{M}^*) \setminus \{\text{rejected models in } \mathcal{N}'_1(\mathcal{M}^*)\};$$

4.e Update

$$\mathcal{R}_1 \leftarrow \mathcal{R}_1 \cup \{\text{weakly rejected models in step 4.b}\} \cup \{\text{rejected models in } \mathcal{N}'_1(\mathcal{M}^*)\};$$

5. The lower layer:

5.a The lower layer of the neighborhood of \mathcal{M}^* , denoted by $\mathcal{N}_2(\mathcal{M}^*)$, is determined as: for any $\mathcal{M} \in \mathcal{N}_1(\mathcal{M}^*)$ which contains pairs of edges $e, \tau(e) \in E(\mathcal{M})$ such that $e \neq \tau(e)$ and $e \notin \mathbb{E}_L(\mathcal{M})$, we:

- remove e , keep $\tau(e)$,
- remove $\tau(e)$, keep e ,

for each such pair $e, \tau(e)$ in $E(\mathcal{M})$;

5.b The weakly rejected models in the lower layer are models in $\mathcal{N}_2(\mathcal{M}^*)$ such that they are:

- submodels of any model in \mathcal{R}_2 , i.e.

$$\{\mathcal{M} \in \mathcal{N}_2(\mathcal{M}^*) \text{ s.t. } \mathcal{M} \preceq_\tau \mathcal{M}_r \text{ for any } \mathcal{M}_r \in \mathcal{R}_2\};$$

- for each model in $\mathcal{A}_1(\mathcal{M}^*)$ which contains pairs of twin edges $e, \tau(e)$ such that $e \notin \mathbb{E}_L$, the corresponding models in $\mathcal{N}_2(\mathcal{M}^*)$ are weakly rejected.

5.c $\mathcal{N}'_2(\mathcal{M}^*) \leftarrow \mathcal{N}_2(\mathcal{M}^*) \setminus \{\text{weakly rejected models in step 5.b}\};$

5.d Test models in $\mathcal{N}'_2(\mathcal{M}^*)$ and remove from $\mathcal{N}'_2(\mathcal{M}^*)$ the rejected models:

$$\mathcal{A}_2(\mathcal{M}^*) \leftarrow \mathcal{N}'_2(\mathcal{M}^*) \setminus \{\text{rejected models in } \mathcal{N}'_2(\mathcal{M}^*)\};$$

5.e Update

$$\mathcal{R}_2 \leftarrow \mathcal{R}_2 \cup \{\text{weakly rejected models in step 5.b}\} \cup \{\text{rejected models in } \mathcal{N}'_1(\mathcal{M}^*)\};$$

6. Update $\mathcal{M}^* \leftarrow$ the model with the highest p -value among models in $\mathcal{A}_1(\mathcal{M}^*) \cup \mathcal{A}_2(\mathcal{M}^*)$;

Repeat steps 3-6 until one of the following stop conditions is satisfied:

- the iterative index exceeds the maximum number of iterations;
- models in the set of neighbors at the current stage are all rejected, i.e. $\mathcal{A}_1(\mathcal{M}^*) = \mathcal{A}_2(\mathcal{M}^*) = \emptyset$.

We notice that, the set of neighbor submodels of the model \mathcal{M}^* is specified fully in Steps 4.a and 5.a. Moreover, at each layer, we can summarize steps as follows:

- Steps 4.b and 5.b determine the weakly rejected models at each layer that are submodels of any models in \mathcal{R}_1 and \mathcal{R}_2 , respectively;
- Steps 4.c and 5.c update the set of neighbors of \mathcal{M}^* by removing weakly rejected models;
- Steps 4.d and 5.d aim to apply the model testing and remove rejected models;
- Steps 4.e and 5.e update the rejection sets \mathcal{R}_1 and \mathcal{R}_2 .

Finding the weakly rejected models at each stage at Steps 4.b and 5.b requires a costly expensive computation since the larger the rejected sets in the further stage, the larger the number of comparisons of each model in the neighborhood set with each rejected model in \mathcal{R} . To deal with this problem, it is sufficient to store the minimal models of \mathcal{R}_1 and \mathcal{R}_2 . However, seeking the minimal models in the increasingly large sets is equally costly. Therefore, in practise, at each layer, (upper and lower), we follow a

different approach for finding directly the set of neighbors $\mathcal{N}'_{(\cdot)}$ at Steps 4.c and 5.c: we apply the meet operation under the twin order, formulated in Theorem 2.6, between the chosen model \mathcal{M}^* with each model in the accepted set $\mathcal{A}_{(\cdot)}$. We then update $\mathcal{A}_{(\cdot)}$ by removing rejected models in $\mathcal{N}'_{(\cdot)}$ after testing models via a certain goodness-of-fit test, and continue the procedure until the stop conditions are satisfied. This idea is described more clearly in the next part of the section by the R implementation below.

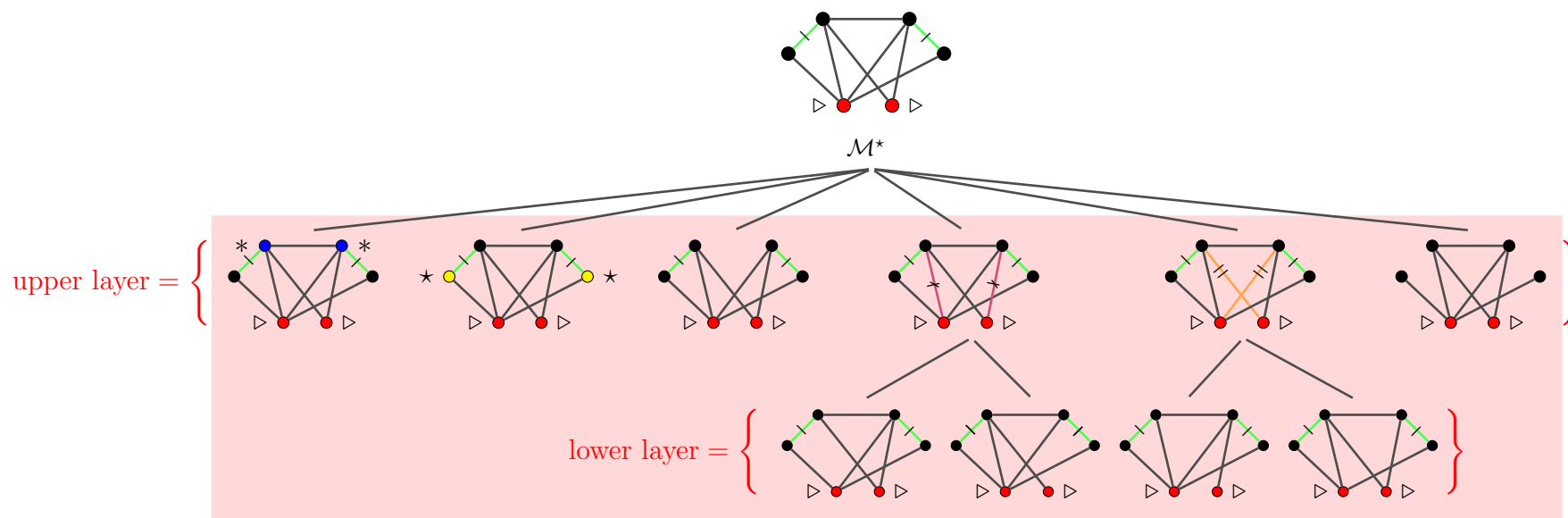


FIGURE 3.6: Example of how to split the set of neighbors of a certain colored graphical model for paired data \mathcal{M}^* in $\mathcal{S}_{[6]}$ into two layers where the nearest neighbors of \mathcal{M}^* are models belonging to the red shaded area.

3.3.2 R implementation

In the programming language R developed in [R Core Team \(2021\)](#), we express a colored graphical model for paired data by a set including p and a list of $(\mathbb{L}, E, \mathbb{E}_L)$ where p is an integer indicating the number of vertices, \mathbb{L} is a vector of (sorted) integers from 1 to p , and (E, \mathbb{E}_L) are two-column matrices with each row corresponding to one edge in the graph.

For the computation of the maximum likelihood estimates, we use the `rcox()` function in the `gRc` package developed by [Højsgaard *et al.* \(2007\)](#) in the statistical environment R. With the setting of `type = "rcon"` as an argument in `rcox()`, the class of RCON models is fitted, in other words, the concentration matrices are estimated with the additional equality constraints implied by the input colored graphs.

For the model selection procedure in [Section 3.3.1](#), we have written the function `backwardCGMpd()` which adapts the idea of backward method in the stepwise approach for the family of PDCGMs under the partial order \preceq_τ . Then, all the operations we applied here are induced by the twin order \preceq_τ . Furthermore, to specify more efficiently the sets of neighbors \mathcal{N}'_1 and \mathcal{N}'_2 at [Steps 4.c](#) and [5.c](#), respectively, we wrote the function `meet.operation()`, which is formally expressed as

$$\mathcal{N}'_k(\mathcal{M}^*) = \{\mathcal{M}^* \wedge \mathcal{M}_a \text{ for all } \mathcal{M}_a \in \mathcal{A}_k\}, \quad k = 1, 2, \quad (3.1)$$

where \wedge is the meet operation formulated in [Theorem 2.6](#), \mathcal{M}^* is the locally chosen model, and \mathcal{A}_k , $k = 1, 2$ are sets of models updated by [Steps 4.d](#) and [5.d](#).

To find \mathcal{A}_k , $k = 1, 2$ conveniently such that we can control the the number of removed models from a list in \mathbf{R} , we start from the sets of neighbor submodels of the saturated models, see Steps 4.a and 5.a.

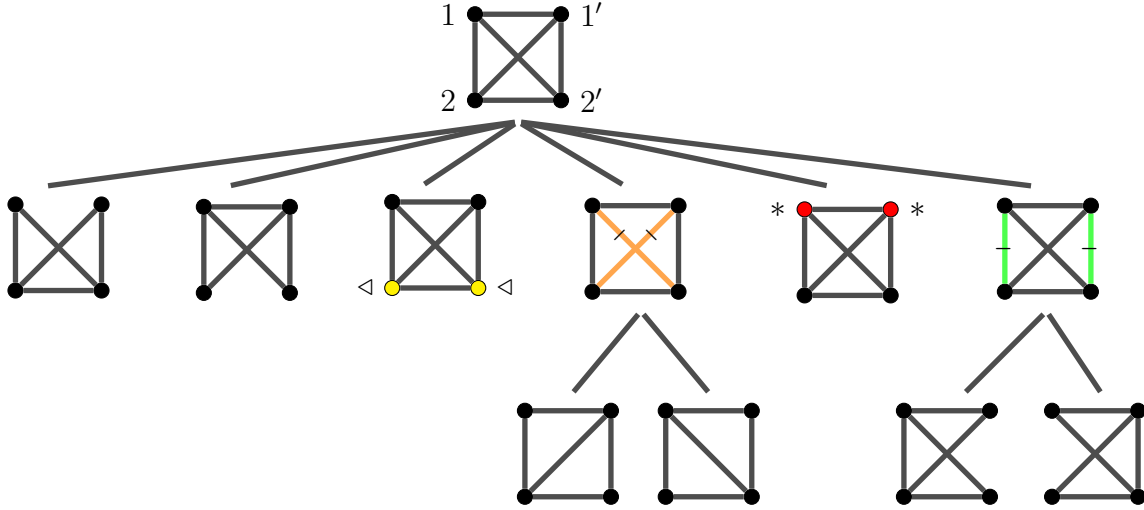


FIGURE 3.7: The set of neighbor submodels of the saturated model with 4 vertices is split into two layers by the twin order \preceq_τ where models in the upper layer and the lower layer can be found at Steps 4.a and 5.a.

Then, for the upper layer, after testing, we assign each model to an index 0 if it is rejected, and 1 otherwise. Hence, \mathcal{A}_1 contains models having the index 1. Next, for each model having symmetric twin edges in the upper layer that is accepted, the corresponding models in the lower layer are assigned to the index 0 and 1 for others. Then, we only test models in the lower layer having the index 1 and update the vector of indices by 0 for rejected models. Thus, \mathcal{A}_2 contains models having the index 1 in the lower layer. In this way, going to further stage, we can save the computation by working only Steps 4.c, 4.d, and 5.c, 5.d for the upper and lower layers, respectively, where

- Steps 4.c and 5.c are specified by Equation 3.1,
- Steps 4.d and 5.d are specified by updating the vectors of indices for each layer.

We note that, in the stepwise backward procedure for PDCGMs based on the twin order \preceq_τ , the index 0 assigned to each model can be implicitly understood by saying that this model either is rejected (i.e. the model found at Step 4.b or in the first point of Step 5.b) or need not to be tested in coherent step (i.e. the model in the second point of Step 5.b).

3.4 Greedy search on the model inclusion lattice

To compare the performance of the procedure described above, applied on the twin lattice of Section 3.3, in this section, we also carry out a similar backward elimination stepwise approach on the lattice of PDCGMs under the model inclusion \preceq_C .

3.4.1 The backward elimination stepwise procedure on the model inclusion lattice

On the model inclusion lattice of PDCGMs, the set of neighbors is remained in the same layer instead of being split as in the approach for \preceq_τ . In this method, we also determine the weakly rejected models and remove them from the set of neighbor submodels at each stage. Every operation on the model inclusion lattice of PDCGMs is induced by the order \preceq_C , found in Section 1.2.4. A sketch of the procedure is provided below.

Procedure B: Stepwise backward elimination procedure for PDCGMs under the model inclusion \preceq_C

1. Start from the saturated model: $\mathcal{M}^* \leftarrow$ saturated model;
2. Set the rejected set $\mathcal{R} = \emptyset$;
3. **Repeat**
 - 3.a The neighborhood of \mathcal{M}^* , denoted by $\mathcal{N}(\mathcal{M}^*)$, is obtained as follows:
 - remove one vertex from \mathbb{L} at a time;
 - remove one edge from \mathbb{E}_L at a time;
 - remove one edge $e \in E$ at a time;
 - remove both $e, \tau(e) \in E$ at a time if $e \notin \mathbb{E}_L$.

Formally, $\mathcal{N}(\mathcal{M}^*) = \{\mathcal{M} \in \mathcal{S}_V : \mathcal{M} \prec_C \mathcal{M}^*\}$, roughly speaking, $\mathcal{N}(\mathcal{M}^*)$ contains models \mathcal{M} in \mathcal{S}_V such that \mathcal{M} is covered by \mathcal{M}^* through the model inclusion \preceq_C ;

- 3.b The weakly rejected models are models in $\mathcal{N}(\mathcal{M}^*)$ such that they are submodels of any model in \mathcal{R} , i.e.

$$\{\mathcal{M} \in \mathcal{N}(\mathcal{M}^*) \text{ s.t. } \mathcal{M} \preceq_C \mathcal{M}_r \text{ for any } \mathcal{M}_r \in \mathcal{R}\};$$

3.c $\mathcal{N}'(\mathcal{M}^*) \leftarrow \mathcal{N}(\mathcal{M}^*) \setminus \{\text{weakly rejected models in step 3.b}\};$

3.d Test models in $\mathcal{N}'(\mathcal{M}^*)$ and remove from $\mathcal{N}'(\mathcal{M}^*)$ the rejected models:

$$\mathcal{A}(\mathcal{M}^*) \leftarrow \mathcal{N}'(\mathcal{M}^*) \setminus \{\text{rejected models in } \mathcal{N}'(\mathcal{M}^*)\};$$

3.e Update

$$\mathcal{R} \leftarrow \mathcal{R} \cup \{\text{weakly rejected models in step 3.b}\} \cup \{\text{rejected models in } \mathcal{N}'(\mathcal{M}^*)\};$$

4. Update $\mathcal{M}^* \leftarrow$ the model with the highest p -value among models in $\mathcal{A}(\mathcal{M}^*)$;

Until one of the following stop conditions is satisfied:

- the iterative manner exceeds the maximum number of iterations;
- models in the neighbor set at the current stage are all rejected, i.e. $\mathcal{A}(\mathcal{M}^*) = \emptyset$.

3.4.2 R implementation

We have written the function `backwardsubmodel()` for the backward elimination stepwise on the model inclusion lattice of PDCGMs. The use of the alternative representation $(V, E, \mathbb{L}, \mathbb{E}_L)$, introduced in Section 2.4, is applicable conveniently to express PDCGMs on the model inclusion lattice, even though operations on such lattice are determined by the set of $(\mathcal{V}, \mathcal{E})$ as in Section 1.2.4. In this method, the terms “acceptance” and “rejection” of models are decided by the likelihood ratio testing relative to the saturated model at significance level α . Moreover, identifying weakly rejected models at Step 3.b requires much more computational effort, therefore, as for the twin lattice, the function `meet.operation.submodel()` gets the set of neighbors \mathcal{N}' based on components $(V, E, \mathbb{L}, \mathbb{E}_L)$, obtained formally by

$$\mathcal{N}'(\mathcal{M}^*) = \{\mathcal{M}^* \wedge \mathcal{M}_a \text{ for all } \mathcal{M}_a \in \mathcal{A}\}, \quad (3.2)$$

where \wedge is the meet operation defined in Section 1.2.4, and \mathcal{A} is the set of models updated in Step 3.d. Identification of the set \mathcal{A} is carried out similarly by starting the set of neighbors of the saturated models determined in Step 3.a.

Then, after testing, we assign each model to 0 if it is rejected, 1 otherwise. Hence, the set \mathcal{A} contains models having the index 1. In this way, we can work only on Steps 3.c and 3.d where

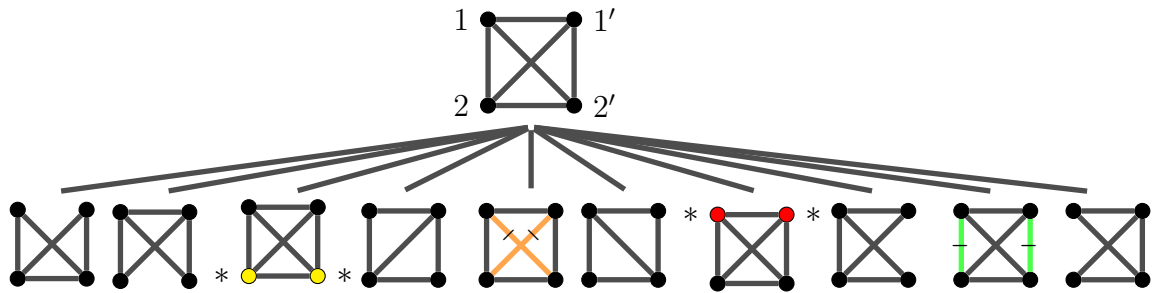


FIGURE 3.8: The set of neighbor submodels of the saturated model with 4 vertices by the model inclusion order $\preceq_{\mathcal{C}}$.

- Step 3.c is obtained by Equation 3.2,
- Step 3.d is specified by update the vector of indices.

Although we perform the PDCGMs by the representation $(V, E, \mathbb{L}, \mathbb{E}_L)$ for easier manipulation, the computing of the meet operation between models is still computationally expensive.

3.5 Comparison and implementation details

In this section, we conduct multiple numerical simulations to illustrate the performance of the greedy search on the twin lattice compared to the model inclusion lattice of PDCGMs. Furthermore, we apply the model selection procedures on fMRI data to investigate the dynamic activity of brain regions between two hemispheres of the human brain.

3.5.1 Numerical simulations

To illustrate the performance of the proposed methods on the lattices of PDCGMs based on the twin order \preceq_{τ} , developed in Section 3.3, and the model inclusion $\preceq_{\mathcal{C}}$, described in Section 3.4, we present the explored simulation experiments as follows. We consider two scenarios, called A and B, where the density of edges in A is sparser with $d_A = 18\%$ whereas the density of edges in B is denser with $d_B = 35\%$. On present edges, the numbers of pairs of symmetric twin vertices and twin edges for each scenario are listed in the below table.

We now have 8 cases of PDCGs presenting various densities of edges and symmetries and corresponding to different values of p , visualized by figures in Table 3.2. For each case, first we generate an uncolored graph with the number of vertices depending on p and the number of edges depending on the density given at each scenario. Then, the

p		Scenario A (number of pairs)	Scenario B (number of pairs)
8	edges	1 symmetric, 1 asymmetric	3 symmetric, 1 asymmetric
	vertices	1 symmetric, 3 asymmetric	3 symmetric, 1 asymmetric
12	edges	1 symmetric, 4 asymmetric	6 symmetric, 3 asymmetric
	vertices	2 symmetric, 4 asymmetric	4 symmetric, 2 asymmetric
16	edges	2 symmetric, 7 asymmetric	10 symmetric, 5 asymmetric
	vertices	2 symmetric, 6 asymmetric	6 symmetric, 2 asymmetric
20	edges	3 symmetric, 11 asymmetric	16 symmetric, 8 asymmetric
	vertices	2 symmetric, 8 asymmetric	8 symmetric, 2 asymmetric

TABLE 3.1: The number of symmetric twin vertices and edges for scenaria A and B for different numbers of vertices p from 8 to 20.

symmetries on twin vertices and twin edges are depicted by the same colors, which are listed in Table 3.1.

Next, for each generated PDCG used as an input graph, the values of the concentration matrix, called $\tilde{\Theta}$, are estimated by the `rcox()` function, where the input empirical covariance matrix is the $p \times p$ *equicorrelation matrix*, with the diagonal entries equaling 1 and all off-diagonal entries equaling 0.5. Thus, $\tilde{\Theta}$ is considered the “true” precision matrix in the joint normal distribution of the simulated data.

We then sample the normal distributed random vector $\mathbf{Y}_{[p]}$ with zero vector mean and covariance matrix obtained by taking the inverse of $\tilde{\Theta}$. The dimension of $\mathbf{Y}_{[p]}$ varies in $\{8, 12, 16, 20\}$, for fixed $n = 100$ independent random samples. These values are used across 20 simulated datasets.

To express the computational efficiency computation of the methods investigated, we calculate the amount of computer time taking to run algorithms by the `proc.time()` function in the statistical environment R. This is a commonly used way to compute the real elapsed time since the process was started. Then, for each case of the PDCG, we make the elapsed time on average in seconds over 20 simulated datasets. The growth of the computational cost is visualized by the graphic tool `ggplot`, where the horizontal axis presents values of p , and the vertical axis presents the averaged elapsed time in seconds, as shown in Figure 3.9. Here, we perform two plottings corresponding to scenarios A and B: the elapsed time of the procedure for the twin order \preceq_{τ} is displayed in red, and the elapsed time of the procedure for submodel relation $\preceq_{\mathcal{C}}$ is displayed in blue.

Moreover, in order to evaluate how “good” the obtained models are, we compute some performance scores, which measure the identification of the zero and symmetric structures. In particular, for the structure, we use *the edge positive-predicted value* (ePPV)

TABLE 3.2: Generated colored graphs for paired data for scenarios A and B (from left to right) with different number of vertices p (from top to bottom).

p	Scenario A	Scenario B
8		
12		
16		
20		

and *the edge true-positive rate* (eTPR), which are the proportions between the number of true edges (eTP) and either the number of edges (#edges) in the selected graph, or the number of edges eP in the true graph. Moreover, we also compute *the edge true-negative rate* (eTNR) as the ratio between the number of true missing edges (eTN) and the number of missing edges (eN) in the true graph. In other words, these quantities

$$ePPV = \frac{eTP}{\#edges}, \quad eTPR = \frac{eTP}{eP}, \quad eTNR = \frac{eTN}{eN},$$

measure the accuracy of the graph structure of models recovered from the model selection procedure. For the identification of the symmetries, we define similarly *the symmetry positive-predicted value* (sPPV) and *the symmetry true-negative rate* (sTNR) which are computed as

$$sPPV = \frac{sTP}{\#sym}, \quad sTPR = \frac{sTP}{sP}, \quad sTNR = \frac{sTN}{sN},$$

where sTP, sTN, and #sym are the number of pairs of true symmetric edges, true symmetric missing edges and symmetric edges in the estimated model. Moreover, sP and sN are the number of pairs of symmetric edges and symmetric missing edges in the true model.

According to the running time we recorded over 20 datasets, shown by the last column in Table 3.9 and displayed by graphics in Figure 3.9, we can see that, in both scenarios, the model selection procedure on the twin lattice, proposed in Sections 3.3, is considerably faster than the similar approach on the model inclusion lattice, described in Section 3.4. The efficient computation of the greedy search on the twin lattice is evident as p is larger. Although the averaged elapsed time of the procedure for \preceq_τ also grows, it is still considerably smaller than the running time of the method for \preceq_C , which grows super-exponentially with the number of vertices. This is due to the simplification in the implementation of the operations in the lattice structure ordered by \preceq_τ , inherited from the set inclusion lattice. Moreover, together with the elapsed time of the algorithm, we also report the number of fitted models per algorithm by graphics in Figure 3.10. From the obtained results of tested models in the local moves, we demonstrate the practical advantage of the coherent step on the twin lattice of PDCGMs: this encourages us to apply our search procedure to PDCGMs with a larger number of variables.

Furthermore, the recorded results in Table 3.3 are also taken on average over 20 simulated data set. As far as the structure of the uncolored graphs is concerned, the models obtained from the approach on the twin lattice have larger values of ePPV, eTPR, eTNR, sTPR, sTNR almost everywhere, whereas sPPV is smaller than models from

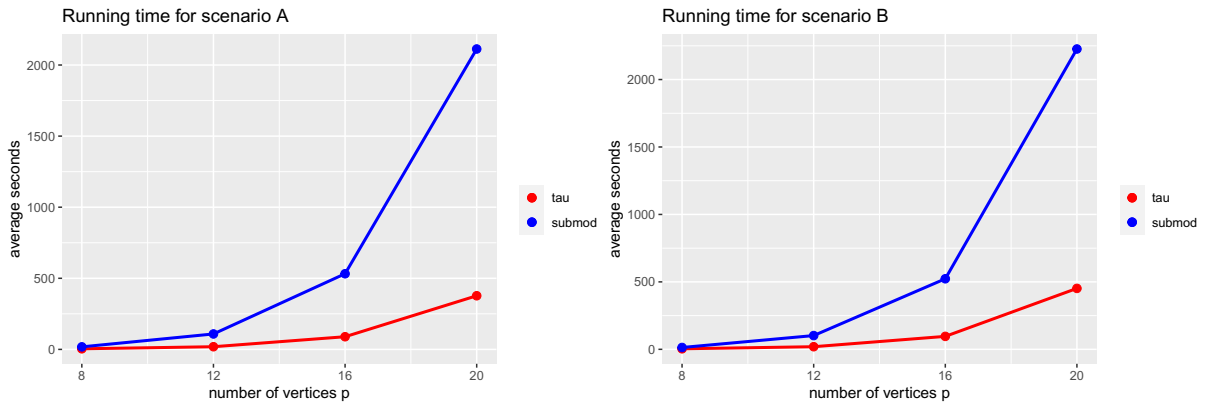


FIGURE 3.9: Averaged seconds on elapsed time from the backward elimination procedures based on the twin order \preceq_{τ} (illustrated in red) and the model inclusion $\preceq_{\mathcal{C}}$ (illustrated in blue) of two scenarios A (on the left) and B (on the right).

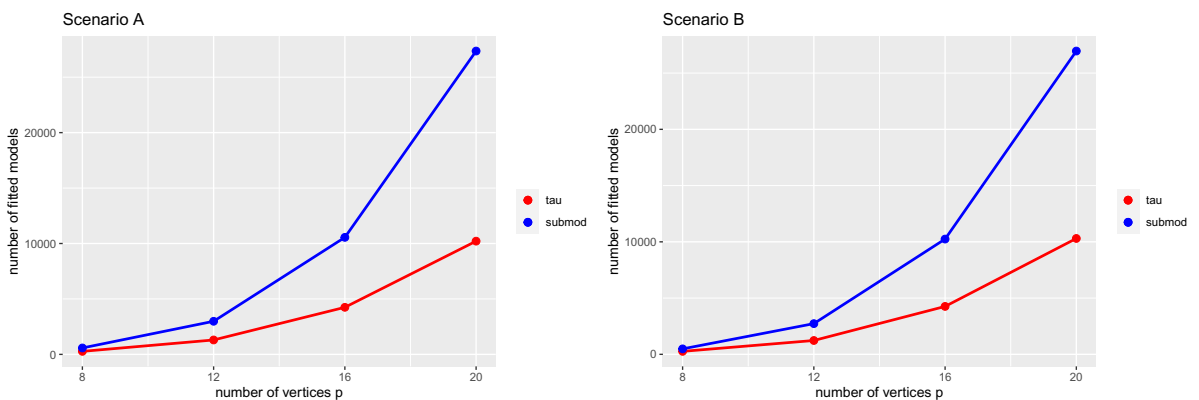


FIGURE 3.10: Averaged number of fitted models from the backward elimination procedures based on the twin order \preceq_{τ} (illustrated in red) and the model inclusion $\preceq_{\mathcal{C}}$ (illustrated in blue) of two scenarios A (on the left) and B (on the right).

the procedure on the model inclusion lattice. In two scenarios, both procedures gain a similar recovery structure of the uncolored graphs. In particular, we get graphs that are denser in the scenario A and graphs that are sparser in the scenario B . However, the procedure based on the twin order \preceq_{τ} tends to recover models with “perfectly” symmetric structure, which is evident in the scenario A . Nonetheless, this method has a good performance behavior in terms of recovery of symmetries for the case of scenario B , with higher proportions in sTPR relative to the number of edges eP in the simulated graphs.

Readers can find the scripts written in the R language that implement the procedures described in Sections 3.3 and 3.4 at <https://github.com/NgocDung-NGUYEN/backwardCGM-PD>.

TABLE 3.3: Performance measures of the model selection procedure for the lattice structure equipped by the partial orders \preceq_τ and \preceq_C . Results are recorded as mean (and standard deviation) computed across the 20 replicated datasets for each scenario. Here, #edges and #sym are computed as the average number of edges, average number of pairs symmetric edges, and their standard deviations are rounded, while **Time** is on the average of seconds.

Scenario	p	Order	Graph structure				Symmetries				Time _(s)	
			#edges	ePPV%	eTPR%	eTNR%	#sym	sPPV%	sTPR%	sTNR%		
A	8	\preceq_τ	7(2)	76.68	100.00	91.52	2(1)	41.67	95.00	89.44	4.02	
		\preceq_C	7(2)	75.41	100.00	91.30	2(1)	46.67	95.00	85.56	17.20	
	12	\preceq_τ	17(3)	71.22	97.92	90.37	6(1)	15.99	90.00	87.61	18.77	
		\preceq_C	17(3)	70.23	98.75	90.00	5(1)	17.34	90.00	83.91	108.55	
	16	\preceq_τ	27(4)	74.83	88.64	92.70	9(1)	18.53	85.00	89.43	89.10	
		\preceq_C	28(4)	70.98	87.05	91.48	8(1)	19.32	77.50	84.77	532.02	
	20	\preceq_τ	44(8)	64.24	82.21	89.49	16(3)	13.47	70.00	86.18	378.74	
		\preceq_C	46(7)	60.11	78.97	88.04	13(3)	11.97	51.67	80.00	2102.20	
	B	8	\preceq_τ	11(2)	84.54	89.50	89.72	5(1)	64.08	93.33	92.50	3.78
			\preceq_C	11(2)	83.59	89.00	89.44	4(1)	64.83	85.00	85.83	14.56
12		\preceq_τ	23(4)	81.78	80.00	89.65	9(2)	56.28	79.17	87.35	19.43	
		\preceq_C	23(4)	81.25	78.48	89.53	7(2)	63.26	73.33	83.53	101.94	
16		\preceq_τ	34(5)	72.49	57.86	87.63	12(2)	52.38	64.00	86.09	96.02	
		\preceq_C	31(4)	74.50	55.24	89.49	9(2)	63.36	54.00	82.97	522.97	
20		\preceq_τ	51(9)	69.74	53.41	87.02	18(2)	48.17	54.38	84.07	451.71	
		\preceq_C	48(7)	67.81	48.64	87.22	12(2)	52.97	39.38	78.98	2226.35	

3.5.2 Application to fMRI data

In this section, the backward elimination stepwise procedures for PDCGMs based on the twin order \preceq_τ and the model inclusion \preceq_c are applied to investigate the dynamic activity of brain regions between two hemispheres via functional MRI data. This is a multimodal imaging dataset coming from a pilot study of the Enhanced Nathan Kline Institute-Rockland Sample project. A detailed description of the project, scopes, and technical aspects can be found at http://fcon_1000.projects.nitrc.org/indi/enhanced/. For the subjects in this study, data on personal covariates, such as anxiety diagnosis, age, gender, and handedness, can be found at http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html. The fMRI time series are recorded on 70 spatial Region of Interests (ROIs) based on the Desikan atlas (see Desikan *et al.* 2006) at 404 equally spaced time. This is provided by Greg Kiar and Eric Bridgeford from NeuroData at Johns Hopkins University, who graciously pre-processed the raw DTI and R-fMRI imaging data available at http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html, using the pipelines ndmg and C-PAC. Particularly, the R-fMRI monitors brain functional activity at different regions via dynamic changes in the blood oxygenation level dependent (BOLD) signal, when, in this study, the subjects are simply asked to stay awake with eyes open. The data that we apply our methods on are residuals estimated from the vector autogression models, carried out to remove the temporal dependence (see Ranciati *et al.* 2021).

According to the available information on subjects, we focus on two participants, indexed as subject 14 and subject 15, who have the same psychological traits with no

neuropsychiatric diseases and right-handedness; however, the first person is only 19 years old while the second person is 57 years old. For each subject, we will study the following 22 cortical regions in the frontal lobe which is the largest of our brain and the center for planning, speaking, body movement control, as well as problem-solving. We are also concerned with other 14 brain regions; some of them are in the anterior temporal lobe, which plays a key role in semantic memory—our knowledge of objects, people, words, and facts; and others are the superior, middle and inferior temporal gyrus, parahippocampal and entorhinal areas, fusiform gyrus and temporal pole (see [Bonner and Price 2013](#), [Raslau et al. 2015](#)). The numbers of regions for each analytic lobe are counted from both hemispheres. The estimated symmetric structure of the brain networks allows us to explore the different brain activity on the cortical regions of interests between two hemispheres, for each individual considered.

The obtained models from the model selection procedures on the twin lattice and on the model inclusion lattice are summarized as colored graphs in Tables 3.4 and 3.5, respectively. The anterior temporal and frontal lobes are represented by colored graphs from the left to the right. From the top to the bottom, results relate to subjects 14 and 15. In these figures, each pair of homologous vertices and edges of the same colors represents identical values of the associated entries on the diagonal and off-diagonal in Θ , respectively. The recovered structures and symmetries from models learned by each procedure are shown in Table 3.6, where each column expresses (in order): the numbers of edges (**#edges**), densities of the uncolored graphs (**den.(%)**), number of asymmetric edges “on the left” and “on the right” (respectively **#asymmetric E_L** and **#asymmetric E_R**); moreover, the last three columns of the table specify the number of pairs of symmetric twin vertices (**#pairs of symmetric vertices**), the overall number of pairs of twin edges (**#pairs of edges**), and the number of pairs of symmetric twin edges (**#pairs of symmetric edges**).

From Table 3.6, we gather that PDCGMs obtained on the twin lattice and the model inclusion lattice have the same characteristics. In particular, for both subjects, we find that there are in fact only few symmetric vertices in the graphs. Namely, for subject 14, all of twin vertices in the the anterior temporal lobe are asymmetric, and there are 5 pairs of symmetric vertices among 11 pairs of twin vertices in the frontal lobe; for subject 15, there are only 3 pairs of symmetric vertices from both of anterior temporal and frontal lobes. Furthermore, on both of anterior temporal and frontal lobes, for subject 14, the number of asymmetric edges “on the left” is larger than “on the right”, while, for the subject 15, the number of asymmetric edges “on the right” is much larger than “on the left”. However, compared to the models selected on the model inclusion

lattice, the models from the twin lattice are represented by colored graphs that are denser with more “perfectly-paired” and “perfectly-symmetric” edges, shown in the last two columns of Table 3.6.

The model selection procedure on the twin lattice, introduced in Section 3.3, is more efficient than the equivalent approach on the model inclusion lattice. More concretely, on the anterior temporal lobe, the backward elimination process on the twin lattice requires 30.24 elapsed seconds for fitting 1, 286 models for subject 14, and 25.25 elapsed seconds for fitting 1, 138 models for subject 15; whereas, the process on the model inclusion lattice requires 146.68 elapsed seconds with 3, 147 fitted models, and 134.02 elapsed seconds with 2, 054 fitted models, respectively. Furthermore, on the frontal lobe, for subjects 14 and 15, the process on the twin lattice requires 577.72 elapsed seconds for fitting 8, 196 models, and 511.78 elapsed seconds to fit 7, 511 models, respectively, whereas, the process on the model inclusion lattice requires 1767.05 elapsed seconds with 15, 311 fitted models, and 1684.14 elapsed seconds with 14, 211 fitted models, respectively. Based on the recorded results, the backward elimination stepwise procedure on the twin lattice has the advantage of being computationally efficient on searching models with a large number of variables, whereas the computation on the model inclusion lattice is too demanding.

Accordingly, we extend the application of the backward elimination stepwise method on the twin lattice, for the data sets for subjects 14 and 15, including $p = 36$ cortical regions in both the anterior temporal and frontal lobes. For case of visualization, we analyze the optimal models obtained into colored subgraphs, as shown in Figures 3.11 and 3.12 for subjects 14 and 15, respectively, where the graphs in the first row have asymmetric edges “to the left” and “to the right”; and the graphs in the second row have symmetric twin edges within and across the two hemispheres. The computational time of the algorithm for the subjects 14 and 15 are 7.95 hours and 7.4 hours, respectively. Moreover, it is useful to detect symmetries on the graphical models, in particular, as we can see, for both subjects, there are some symmetric vertices. In addition, the numbers of asymmetric edges “on the left” and “on the right” behave similarly to previous analyses for the anterior temporal and frontal lobes in the sense that, for the subject 14, the number of asymmetric edges “on the left” is greater than “on the right”, which is the opposite for subject 15.

TABLE 3.4: Graphical representation of the colored graphical models for paired data obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_τ , proposed in Section 3.3, with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. From left to right: anterior temporal lobe and the frontal lobe. From top to bottom: subject 14 and subject 15.

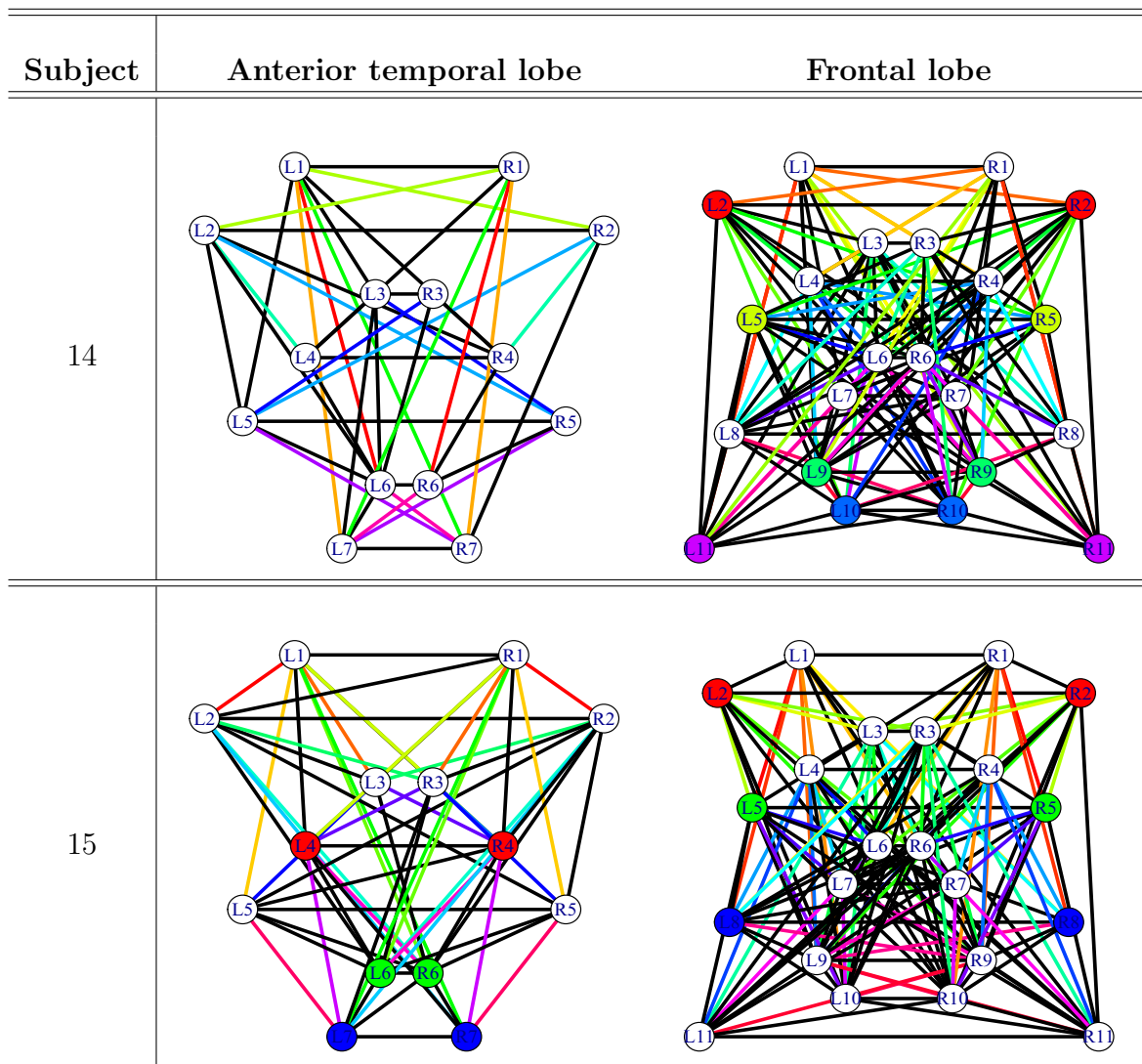


TABLE 3.5: Graphical representation of the colored graphical models for paired data obtained from the application of the stepwise backward elimination procedure based on the model inclusion \preceq_c , described in Section 3.4, with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. From left to right: anterior temporal lobe and the frontal lobe. From top to bottom: subject 14 and subject 15.

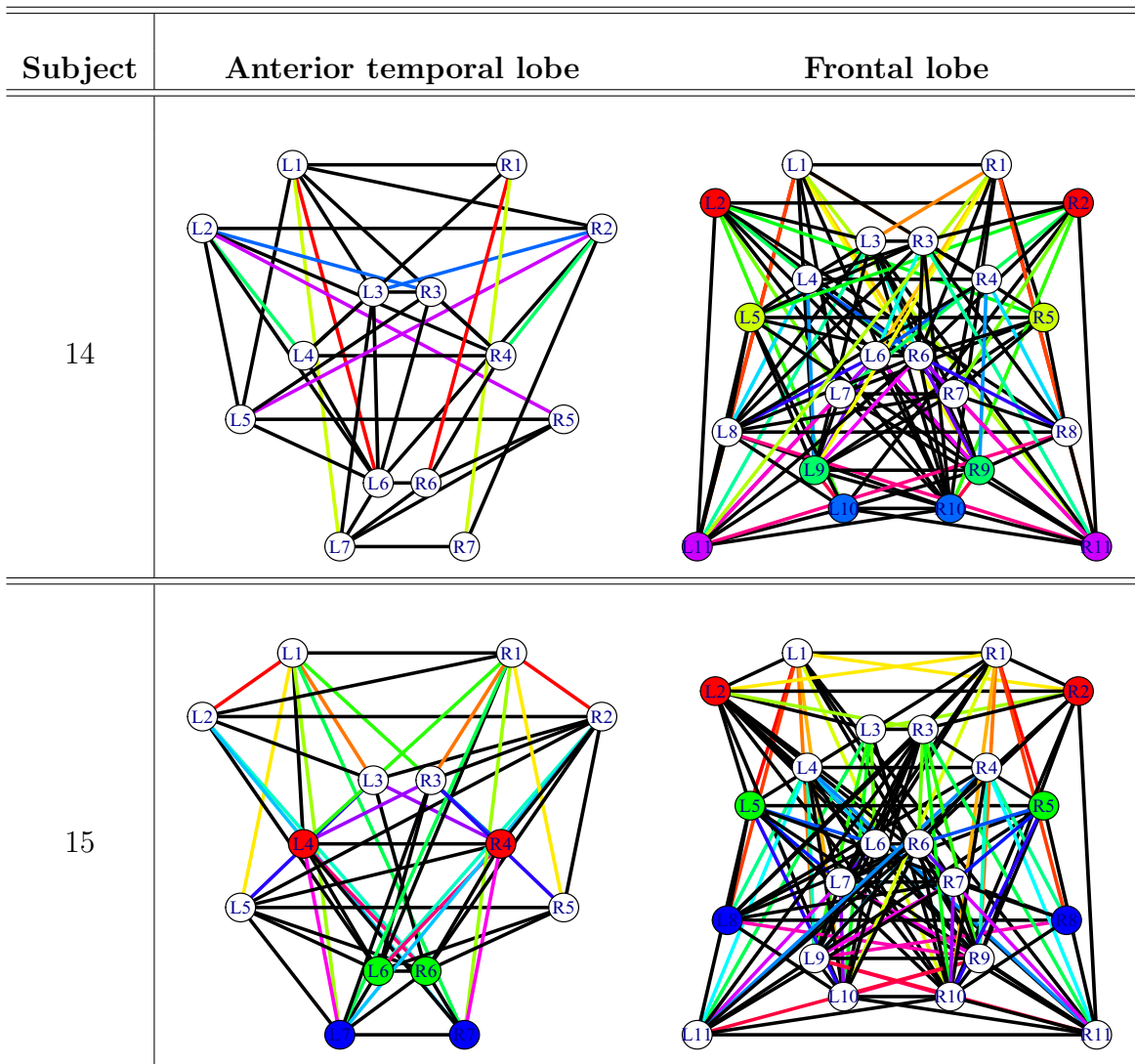


TABLE 3.6: Recorded results from the application to fMRI data of the backward elimination stepwise procedure on the twin lattice, proposed in Section 3.3, and similar approach on the model inclusion lattice, described in Section 3.4.

Subject	Lobe	Order	# edges	den.(%)	# asymmetric		# pairs of		
					E_L	E_R	symmetric vertices	edges	symmetric edges
14	Anterior temporal	\setminus_{τ}	43	47.25	12	6	0	13	9
		$\setminus_{\mathcal{C}}$	40	43.96	13	10	0	9	5
	Frontal	\setminus_{τ}	138	59.74	35	33	5	50	30
		$\setminus_{\mathcal{C}}$	127	54.98	37	34	5	36	23
15	Anterior temporal	\setminus_{τ}	63	69.23	11	16	3	25	15
		$\setminus_{\mathcal{C}}$	58	63.74	10	16	3	20	13
	Frontal	\setminus_{τ}	143	61.9	29	39	3	53	32
		$\setminus_{\mathcal{C}}$	136	58.87	31	42	3	44	26

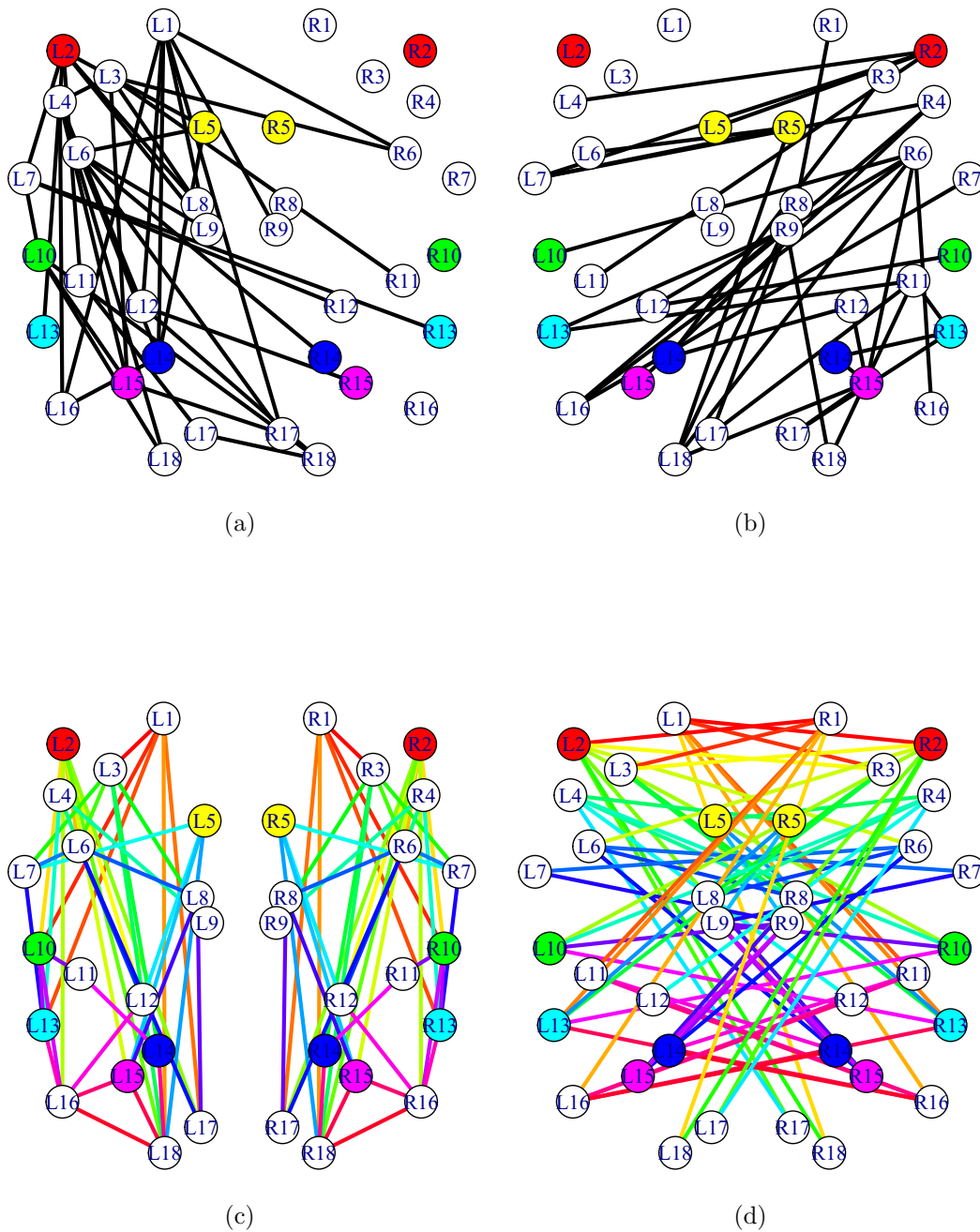


FIGURE 3.11: The analyzed colored graphs of the PDCGM for the subject 14 obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_{τ} , with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. Figures (a) and (b) present asymmetric edges “on the left” and “on the right”, respectively. Figures (c) and (d) present symmetric twin edges between and across two hemispheres, respectively.

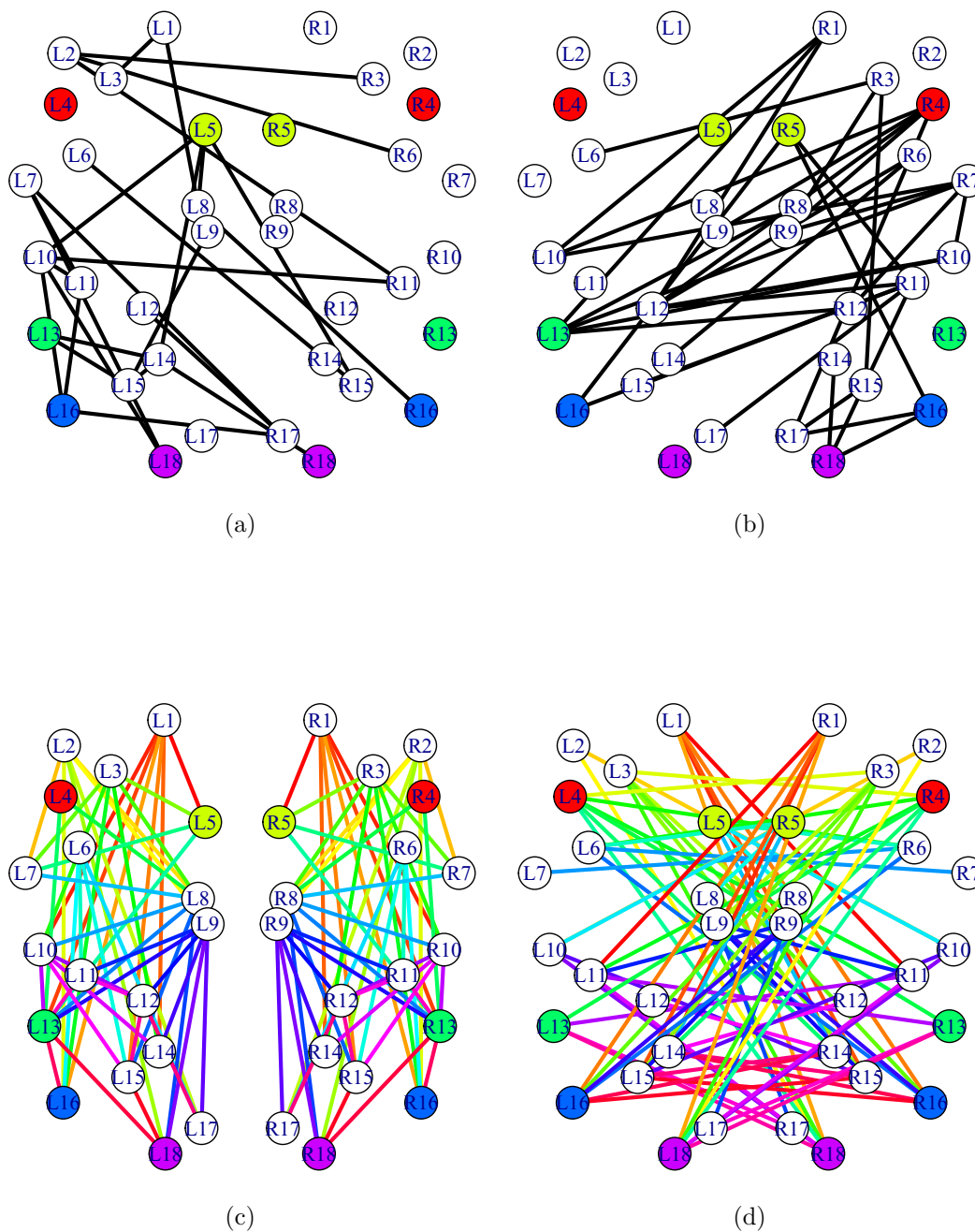


FIGURE 3.12: The analyzed colored graphs of the PDCGM for the subject 15 obtained from the application of the stepwise backward elimination procedure based on the twin order \preceq_{τ} , with the likelihood ratio test at the significance level $\alpha = 0.05$. The colored shaded vertices and colored edges represent symmetric diagonal and off-diagonal concentrations, respectively. Figures (a) and (b) present asymmetric edges “on the left” and “on the right”, respectively. Figures (c) and (d) present symmetric twin edges between and across two hemispheres, respectively.

3.6 Other approaches

In this section, we will investigate other approaches in the model selection for the family of colored graphical models for paired data.

3.6.1 Forward inclusion stepwise procedure on the twin lattice

In the previous section, we have shown that the backward elimination stepwise procedure of the twin lattice for PDCGMs can be applied for a larger number of variables, in particular, in the application to fMRI data, the model selection for PDCGMs on the twin lattice with 36 variables has been carried out. However, with the desire to be able to proceed in the higher dimensional setting, it is a good point to consider the *forward inclusion stepwise procedure*. In this section, we will investigate the coherent step in the forward selection on the lattice of PDCGMs under the twin order \preceq_τ .

Unlike the backward elimination, the forward inclusion method starts with a null model (i.e. the model with full conditional independence restrictions) and proceeds to add parameters one at a time until the stop condition is satisfied. Therefore, the forward selection is used when the number of variables under consideration is very large, even larger than the sample size. However, in the following example, we show that the forward stepwise procedure on the twin lattice for PDCGMs cannot implement with the coherent step.

Example 3.2. *We apply the forward stepwise procedure on the sublattice of RCON models for paired data represented by colored graphs with 4 vertices, displayed in Figure 3.13. We start from the null model which is represented by (1) and assume that, by the likelihood ratio test relative to the null model at the significance level α , we accept simultaneously the hypotheses (2) and (3). Hence, this is not coherent in the sense that we can conclude that on the concentration matrix, the corresponding parameters $\theta_{12'} = 0$ and $\theta_{21'} = 0$ but cannot accept that they have the same value and are equal to zero, i.e., we cannot accept (1).*

In the case we start from the null model and assume that it is accepted, then, by the equivalent form of the principle of coherence in the model search procedures (see [Edwards and Havránek 1987](#)), all models that include the null model are considered to be accepted. With a tendency to choose the simplest model among accepted models in the model selection, in this case, the null model is always selected.

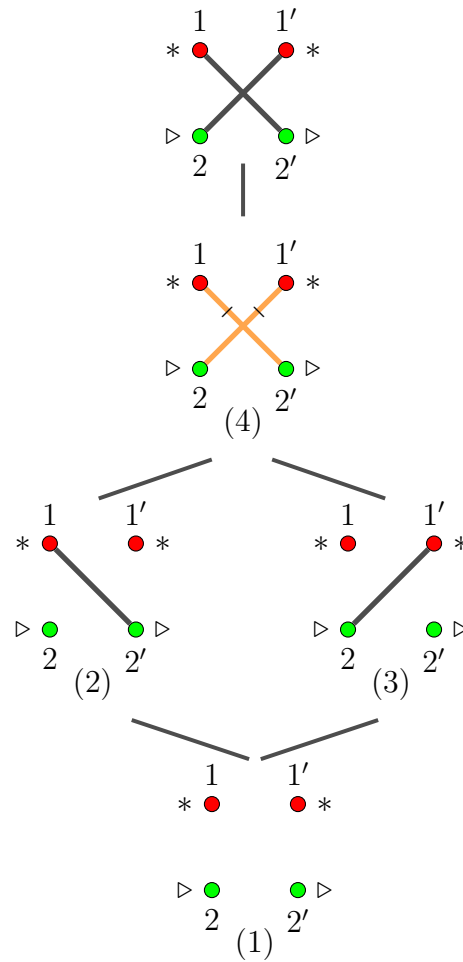


FIGURE 3.13: Sublattice structure of colored graphs for paired data with the set of 4 vertices under the twin order \preceq_{τ} .

3.6.2 Penalized likelihood approach

Another procedure that would be interesting to investigate in the model selection is the *penalized likelihood approach*. This is an effective method attempting to find a sparse and shrinkage estimator of the parameters, and thus model selection and estimation can be conducted simultaneously.

More specifically, [Yuan and Lin \(2007\)](#) proposed the *graphical lasso* that employs an l_1 penalty on the elements of the concentration matrix to encourage the sparsity and at the same time give shrinkage estimates. However, this method cannot be implemented on the PDCGMs since it encourages the number of missing edges without the symmetry structure in the associated graph. Table 3.6.2 shows recorded results of selected models obtained from the graphical lasso implemented in R to fMRI data. The optimal penalty parameter in the glasso regularization is chosen by the *extended Bayesian information criterion* (see [Chen and Chen 2008](#)). As we can see, the number of symmetric twin

vertices and twin edges cannot be measured.

Moreover, as we mentioned in Section 3.1, the *symmetric graphical lasso* proposed in [Ranciati et al. \(2021\)](#) cannot be applied to the case of our consideration of paired data because variables \mathbf{Y}_L and \mathbf{Y}_R are not independent.

TABLE 3.7: Recorded results from the application to fMRI data of the graphical glasso (glasso) approach.

Subject	Lobe	# edges	den.(%)	E_L	E_R	# pairs of		
						symmetric vertices	edges	symmetric edges
14	Anterior temporal	28	31.32	9	14	-	7	-
	Frontal	114	49.57	53	50	-	34	-
15	Anterior temporal	21	23.08	8	9	-	5	-
	Frontal	106	45.89	54	41	-	28	-

Concluding remarks

In this thesis, we mainly focus on the lattice structure and how to explore this lattice in a greedy search procedure on the model space for the family of colored graphical models for paired data.

The implementation of a greedy search procedure requires the specification of a number of different aspects and the selected models of the procedure will depend on all the choices made. One such choice concerns the specification of the structure of the model space and the type of moves used to explore the model space. This thesis has concentrated on this part of the procedure. Other aspects that may be of interest concern the comparison of alternative statistical tests used to compute the p -value as well as the application of the multiple testing corrections to the p -value.

Although we have focused on a backward elimination stepwise procedure starting from the saturated model, in the high dimensional setting, it may be more efficient to use a forward inclusion stepwise procedure starting from the null model. One difficulty for such a procedure is that the coherent step cannot be implemented on the twin lattice for PDCGMs, as shown in Section 3.6. Moreover, to the best of our knowledge, the recent penalized likelihood methods cannot be directly compared to our approach. In particular, we consider the graphical lasso proposed in [Yuan and Lin \(2007\)](#) which encourages the sparsity without the symmetry of the associated graph, and the symmetric graphical lasso proposed in [Ranciati *et al.* \(2021\)](#) which encourages the sparsity and the similarity between two independent subgraphs. These problems are also discussed in Section 3.6. These are all objects of future works.

Bibliography

- Antoch, J. and Hanousek, J. (2001) Model selection and simplification using lattices. *SSRN Electronic Journal* . (Cited on page 61.)
- Bickel, D. R. and Patriota, A. G. (2019) Self-consistent confidence sets and tests of composite hypotheses applicable to restricted parameters. *Bernoulli* **25**(1), 47–74. (Cited on page 61.)
- Birkhoff, G. (1940) *Lattice theory*. Volume 25. American Mathematical Society. (Cited on pages 7 and 10.)
- Bondy, J. and Murty, U. (2008) *Graph Theory, 6 Springer*. Volume 244. (Cited on page 14.)
- Bonner, M. F. and Price, A. R. (2013) Where is the anterior temporal lobe and what does it do? *Journal of Neuroscience* **33**(10), 4213–4215. (Cited on page 81.)
- Brualdi, R. A. (1977) *Introductory combinatorics*. Pearson Education India. (Cited on page 13.)
- Canfield, E. R. (2001) Meet and join within the lattice of set partitions. *The Electronic Journal of Combinatorics* pp. R15–R15. (Cited on page 13.)
- Chatfield, C. and Collins, A. J. (2018) *Introduction to multivariate analysis*. Routledge. (Cited on pages 21 and 25.)
- Chen, J. and Chen, Z. (2008) Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**(3), 759–771. (Cited on page 89.)
- Cox, D. R. and Wermuth, N. (2014) *Multivariate dependencies: Models, analysis and interpretation*. Chapman and Hall/CRC. (Cited on page 26.)
- Davey, B. A. and Priestley, H. A. (2002) *Introduction to lattices and order*. Cambridge University Press. (Cited on pages 10 and 11.)
- Dempster, A. P. (1972) Covariance selection. *Biometrics* pp. 157–175. (Cited on page 24.)

- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T. *et al.* (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**(3), 968–980. (Cited on page 80.)
- Edwards, D. (2012) *Introduction to graphical modelling*. Springer Science & Business Media. (Cited on page 61.)
- Edwards, D. and Havránek, T. (1987) A fast model selection procedure for large families of models. *Journal of the American Statistical Association* **82**(397), 205–213. (Cited on pages ii, vi, 61, and 88.)
- Gabriel, K. R. (1969) Simultaneous test procedures—Some theory of multiple comparisons. *The Annals of Mathematical Statistics* pp. 224–250. (Cited on pages ii, vi, and 61.)
- Gehrmann, H. (2011) Lattices of graphical Gaussian models with symmetries. *Symmetry* **3**(3), 653–679. (Cited on pages 3, 18, 19, 33, 57, 59, and 61.)
- Grätzer, G. (2002) *General lattice theory*. Springer Science & Business Media. (Cited on pages 7 and 10.)
- Gross, J. L. and Yellen, J. (2003) *Handbook of graph theory*. CRC Press. (Cited on pages 14 and 17.)
- Habib, M., Medina, R., Nourine, L. and Steiner, G. (2001) Efficient algorithms on distributive lattices. *Discrete Applied Mathematics* **110**(2-3), 169–187. (Cited on page 11.)
- Højsgaard, S. and Lauritzen, S. L. (2008) Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 1005–1027. (Cited on pages i, v, 24, 28, 29, and 63.)
- Højsgaard, S., Lauritzen, S. L. *et al.* (2007) Inference in graphical Gaussian models with edge and vertex symmetries with the `gRc` package for R. *Journal of Statistical Software* **23**(6), 1–26. (Cited on page 69.)
- Izbicki, R. and Esteves, L. G. (2015) Logical consistency in simultaneous statistical test procedures. *Logic Journal of the IGPL* **23**(5), 732–758. (Cited on page 61.)
- Krzanowski, W. (2000) *Principles of multivariate analysis*. Volume 23. OUP Oxford. (Cited on page 21.)
- Lauritzen, S. L. (1996) *Graphical models*. Volume 17. Clarendon Press. (Cited on pages i, v, 22, 25, and 26.)

- Li, Q., Gao, X. and Massam, H. (2020) Bayesian model selection approach for coloured graphical Gaussian models. *Journal of Statistical Computation and Simulation* **90**(14), 2631–2654. (Cited on page 59.)
- Li, Q., Sun, X., Wang, N. and Gao, X. (2021) Penalized composite likelihood for colored graphical Gaussian models. *Statistical Analysis and Data Mining: The ASA Data Science Journal* . (Cited on page 60.)
- Maathuis, M., Drton, M., Lauritzen, S. and Wainwright, M. (2018) *Handbook of Graphical Models*. CRC Press. (Cited on page 26.)
- Munro, J. I. and Sinnamon, C. (2018) Time and space efficient representations of distributive lattices. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 550–567. (Cited on page 11.)
- Patriota, A. G. (2013) A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems* **233**, 74–88. (Cited on page 61.)
- Pittel, B. (2000) Where the typical set partitions meet and join. *The Electronic Journal of Combinatorics* **7**(1), R5. (Cited on page 13.)
- R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (Cited on pages ii, vi, and 69.)
- Ranciati, S., Roverato, A. and Luati, A. (2021) Fused graphical lasso for brain networks with symmetries. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **70**(5), 1299–1322. (Cited on pages i, v, 60, 80, 90, and 93.)
- Raslau, F., Mark, I., Klein, A., Ulmer, J., Mathews, V. and Mark, L. (2015) Memory part 2: The role of the medial temporal lobe. *American Journal of Neuroradiology* **36**(5), 846–849. (Cited on page 81.)
- Rencher, A. and Christensen, W. (2012) *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics. Wiley. ISBN 9781118391679. (Cited on page 25.)
- Schechter, E. (1996) *Handbook of Analysis and its Foundations*. Academic Press. (Cited on page 7.)
- Sonnemann, E. (2008) General solutions to multiple testing problems. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **50**(5), 641–656. (Cited on page 61.)

- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005) Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(1), 91–108. (Cited on page 60.)
- Tong, Y. L. (2012) *The multivariate normal distribution*. Springer Science & Business Media. (Cited on page 25.)
- Tsai, K., Koyejo, O. and Kolar, M. (2021) Joint Gaussian graphical model estimation: A survey. *arXiv preprint arXiv:2110.10281* . (Cited on pages i and v.)
- Vinciotti, V., Augugliaro, L., Abbruzzo, A. and Wit, E. C. (2016) Model selection for factorial Gaussian graphical models with an application to dynamic regulatory networks. *Statistical Applications in Genetics and Molecular Biology* **15**(3), 193–212. (Cited on page 60.)
- Whitakker, J. (1990) Graphical models in applied multivariate analysis. (Cited on pages 21 and 24.)
- Yuan, M. and Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**(1), 19–35. (Cited on pages 89 and 93.)

Dung Ngoc NGUYEN

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 338 770 5721
e-mail: ngocdung.nguyen@studenti.unipd.it

Current Position

Since October 2018; (expected completion: December 2021)

PhD Student in Statistical Sciences, University of Padova.

Thesis title: Model selection for colored graphical models for paired data

Supervisor: Prof. Alberto Roverato.

Research interests

- **Statistical learning:** supervised and unsupervised learning, model inference, e.g. bootstrap methods, Bayesian methods, high-dimensional problems, e.g. regularization, principal component analysis (PCA), etc.
- **Applied statistics:** statistical inference, statistical modelling, regression analysis and diagnostics, model selection on search spaces, e.g. greedy search.
- **Graphical models:** structure learning and model search for (undirected) graphical models, e.g. graphical Lasso, model selection of colored graphical models for paired data, inferring the distributions of quantities of interests given the observations.
- **Biostatistics:** statistical methods and models for large and complex biological data sets, inference for biological/genetic networks.

Education

September 2017 – August 2018

Master (*laurea specialistica/magistrale*) degree in Science, Technology and Health, Mathematics and Applications Specialty.

Université de Tours

Title of dissertation: “Limits for large random graphs”

Supervisor: Prof. Romain Abraham

Final mark: 16.796/20.

September 2013 – August 2017

Bachelor degree (*laurea triennale*) in Sciences.

University of Sciences, Vietnam National University, Faculty of Mathematics and Computer Science

Title of dissertation: “Multiplicative censoring models”

Supervisor: Prof. Dang Duc Trong

Final mark: 8.21/10.

Further education

October 2021

Course name: Statistical aspects of Deep Neural Networks.

Organizing Institution: the Department of Economics, Quantitative Methods and Business Strategies of the University of Milan-Bicocca, Italy.

Organizer: Ph.D. in Economics, Statistics and Data Science.

Instructor: Professor Omiros Papaspiliopoulos (Bocconi University, Milano, Italy).

September 2020

Course name: Regression Modelling with Large Data Sets.

Organizing Institution: the Department of Statistical Sciences of the University of Padova.

Organizer: Department of Statistical Sciences.

Instructor: Professor Ioannis Kosmidis (University of Warwick and Turing Fellow of The Alan Turing Institute, UK).

April – June 2020

Course name: Statistical Consulting with the group's title "Predicting the economic effects of COVID-19 and of the lockdown measures".

Organizing Institution: the Department of Statistical Sciences of the University of Padova.

Organizer: Department of Statistical Sciences.

Instructors: Professors Antonio Canale and Bruno Scarpa, (University of Padova, Italy).

Work experience

August – October 2021

Employer: Department of Statistical Sciences Paolo Fortunati, University of Bologna, Italy.

Awards and Scholarship

Date: December 20, 2021

Research grant (Type A) entitled "Structure learning of multiple graphical models for dependent samples"- Research Supervisor Prof. Alberto Roverato.

Computer skills

- Programming languages: Advanced R, Python, MATLAB.
- Operating systems: Ubuntu, macOS.

Language skills

- Vietnamese: mother tongue.
- English: fluent.
- Italian: basic.

Publications

Working papers (to be in progress)

Title: "Model selection for colored graphical models for paired data".

References

Prof. Alberto Roverato

Department of Statistical Sciences, University of Padova
Address: via Cesare Battisti, 241-243, 35121 Padova. Italy
Phone: +39 049 827 4183
e-mail: alberto.roverato@unipd.it

