



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 XXXII

Heterogeneous Graphical Models with Applications to Omics Data

Coordinatore del Corso: Prof. Massimiliano Caporin

Supervisore: Prof. Monica Chiogna

Dottorando/a: Ilaria Bussoli

02 December 2019

Abstract

Thanks to the advances in bioinformatics and high-throughput methodologies of the last decades, a large unprecedented amount of biological data coming from various experiments in metabolomics, genomics and proteomics is available. This has led the researchers to conduct more and more comprehensive molecular profiling of biological samples through different multiple aspects of genomic activities, thus introducing new challenges in the developments of statistical tools to integrate and model multi-omics data. The main research objective of this thesis is to develop a statistical framework for modelling the interactions between genes when their activity is measured on different domains; to do so, our approach relies on the concept of multilayer network, and how structures of this type can be combined with graphical models for mixed data, i.e., data comprising variables of different nature (e.g., continuous, categorical, skewed, to name a few). We further develop an algorithm for learning the structure of the undirected multilayer networks underlying the proposed models, showing its promising results through empirical analyses on cancer data, which was downloaded from the public TCGA consortium.

Sommario

Grazie ai progressi negli ultimi decenni raggiunti dalla bioinformatica e dalle metodologie high-throughput, una quantità senza precedenti di dati biologici provenienti dai vari esperimenti di metabolomica, genomica e proteomica è disponibile alla comunità scientifica. Ciò ha permesso ai ricercatori di condurre una profilazione molecolare sempre più completa di vari campioni biologici, poiché è diventato possibile misurare i diversi aspetti che compongono le attività genomiche. Questo ha fatto sì che nascessero nuove sfide negli sviluppi di strumenti statistici per integrare e modellare dati multi-omici. Il principale obiettivo di ricerca di questa tesi è quello di sviluppare un quadro statistico per inferire le interazioni tra geni quando la loro attività è misurata su domini diversi; per fare ciò, il nostro approccio si basa sul concetto di multilayer network e su come strutture di questo tipo possano essere combinate con modelli grafici per dati misti, vale a dire dati che comprendono variabili di diversa natura (ad esempio, variabili continue, categoriali, skewed, per nominarne alcune). Nella tesi viene inoltre sviluppato un algoritmo per l'apprendimento della struttura di multilayer network non orientate che sono alla base dei modelli proposti, mostrandone i promettenti risultati attraverso analisi empiriche su dati oncologici, scaricabili dal sito pubblico del consorzio TCGA.

Individuals are not stable things, they are fleeting. Chromosomes too are shuffled into oblivion, like hands of cards soon after they are dealt. But the cards themselves survive the shuffling. The cards are the genes. The genes are not destroyed by crossing-over, they merely change partners and march on. Of course they march on. That is their business. They are the replicators and we are their survival machines. When we have served our purpose we are cast aside. But genes are denizens of geological time: genes are forever.

Richard Dawkins, *The Selfish Gene*

Acknowledgements

I would like to thank my supervisor, Professor Monica Chiogna, for her full support until the end, her faith in my abilities and her skilled guidance throughout this PhD project. She has been a clear lighthouse when the “research sea” was rough and stormy. She was a patient mentor, a diligent teacher and a passionate researcher for all these past three years, and I cannot think of anyone else doing a better job of growing me up at an academic level (although I still have a lot to learn). I would like to thank also Sofia Massa, who introduced me to the unique world of Oxford University, whose being full of ideas provided me with the bases on which my thesis is founded, and whose curiosity made me passionate about this branch of research. I also thank Chiara Romualdi and her lab for providing the data and valuable insights from the biological perspective, making this work possible.

I thank all my colleagues from the 32nd PhD cycle, Ana and Huiting foremost. Your kindness and love will be forever in my heart, as you, and your friendship is a rare treasure I will cherish for the rest of my life. Thank to Meme, Ale, Fede, and Andrea, for all the laughs and joyful and crazy moments we passed together. I thank Moin and Mohammed, who opened my eyes on other realities and, sometimes, on what is really important in life. Without all of you, guys, I couldn't be the woman and researcher I am today, and I thank you so much for being this amazing group of people.

I thank Silvio, my sweet “raggio di sole”, for believing in who I am and who we are, for supporting me every time and in every case, for never losing his patience, and for all the philosophical and very long discussions we like to have more or less every weekend. I thank Anna and Giulia, for their everlasting friendship without boundaries, their unconditional love for life and for all the journeys we walked together. Although we have not seen each other as often as during the master, I thank Ilaria, Genny and Luisa, because they never fail to make me feel cuddled every time we meet. I thank Paola, who remains a pillar of laughs, comics and support in all my life decisions. I thank my parents for being a source of strength, always ready to help and listen to me. Finally, I thank myself, because I pursued the goals I had set myself even when things seemed sour.

Contents

List of Figures	vi
List of Tables	viii
Introduction	1
Overview	1
Main contributions of the thesis	3
1 Graphs, conditional independences and Markov properties	5
1.1 Review on Graph theory	6
1.2 Multilayer networks	8
1.2.1 Notation and terminology	8
1.2.2 Some new definitions: multi-, intra-, inter-layer separators	11
1.3 Conditional independence	14
1.4 Markov properties	17
1.4.1 Standard definitions	18
1.4.2 Extension to multilayer networks	19
2 Background in probabilistic graphical models for heterogeneous data	21
2.1 Literature review	21
2.2 The main instigator: conditional Gaussian GMs	23
3 Multilayer Conditional Gaussian Graphical Models	27
3.1 Rationale	27
3.2 Model specification	29
3.2.1 Data format	29
3.2.2 The model	31
4 Structure learning of undirected multilayer graphs	37
4.1 Literature review	37
4.2 The PC-CGRM algorithm	39
4.3 Simulation studies	45
5 Application to Omics data	55
5.1 Ovarian cancer data	55
5.1.1 Why ovarian cancer	55
5.1.2 Preprocessing	56

5.2 Results	59
6 Conclusions	63
Appendix A Properties of conditional independence	65
Appendix B Conditioning and marginalizing on CG distributions	68
Appendix C Output of PC-CGRM on real data	82

List of Figures

1.1	(A) A simple undirected and connected graph $G = (V, E)$ and (B) the subgraph of G induced by $A = \{1, 2, 3, 4\}$	7
1.2	A multilayer network $M = (V_M, E_M, V, \mathbf{L})$ (panel (A)) and the graph $G_M = (V_M, E_M)$ underlying M (panel (B)).	9
1.3	A marked graph (panel A) pictured as a multilayer network (panel B). . .	11
1.4	Examples of multilayer separators for the (connected) multilayer network M of panel (a): multi-layer separator (panel (b)); intra-layer separator (panel (c)); inter-layer separator (panel (d)). In panels (b), (c), and (d), the edges directly involved in the path precluded by the separator are coloured, purple, light green and light pink, respectively.	15
3.1	Interpretation of distinct measurements on the same p genes for a specific activity; panel (A) represents our approach, while panel (B) shows what it is assumed in literature.	28
4.1	Multilayer network of scenario 1. Panel (B) shows the network in its entirety; black dots represent categorical variables and white dots represent continuous variables. Panel (B) contains the representation of the same network through its layers.	46
4.2	Compacted visualization of the multilayer network for $p = 20$ layers. . . .	48
4.3	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 1$	49
4.4	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 2$	49
4.5	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 5$	50
4.6	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 1$	50
4.7	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 2$	51
4.8	TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 5$	51
5.1	Boxplots of the absolute frequencies of each CNVs category computed on all 15587 genes.	58
5.2	Inferred undirected multilayer graph for <i>Ca2+</i> pathway with cardinality $m = 1$ (panel (A)) in comparison with cardinality $m = 5$ (panel (B)); the nominal level in both cases is $\alpha = 0.001$	61

5.3	Inferred undirected multilayer graph for <i>Binding and Uptake of Ligands by Scavenger Receptors</i> pathway, with cardinality $m = 1$ (panel (A)) in comparison with cardinality $m = 5$ (panel (B)); the nominal level in both cases is $\alpha = 0.001$	62
C.1	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.001$	82
C.2	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.001$	83
C.3	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.001$	83
C.4	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.001$	84
C.5	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.001$	84
C.6	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.01$	85
C.7	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.01$	85
C.8	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.01$	86
C.9	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.01$	86
C.10	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.01$	87
C.11	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.05$	87
C.12	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.05$	88
C.13	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.05$	88
C.14	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.05$	89
C.15	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.05$	89
C.16	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.1$	90
C.17	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.1$	90
C.18	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.1$	91
C.19	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.1$	91
C.20	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.1$	92
C.21	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.001$	93
C.22	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.001$	94
C.23	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.001$	94
C.24	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.001$	95
C.25	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.001$	95
C.26	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.01$	96
C.27	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.01$	96
C.28	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.01$	97
C.29	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.01$	97
C.30	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.01$	98
C.31	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.05$	98
C.32	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.05$	99
C.33	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.05$	99
C.34	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.05$	100
C.35	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.05$	100
C.36	Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.1$	101
C.37	Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.1$	101
C.38	Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.1$	102
C.39	Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.1$	102
C.40	Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.1$	103

List of Tables

1.1	Examples of basic cases leading to disconnectedness in a multilayer network M with $d = 1$ aspects.	13
3.1	Illustration of equivalent mathematical structures discussed in the data format section: the random matrix \mathbb{Y} , its representation by multilayer network if it is read by rows (nodes in $\{1, \dots, K\}$ and layers in $\{1, \dots, p\}$) and its counterpart by columns (nodes in $\{1, \dots, p\}$ and layers in $\{1, \dots, K\}$).	31
4.1	Runtime of PC-CGRM for the first scenario with sample size $n = 100$	52
4.2	Runtime of PC-CGRM for the first scenario with sample size $n = 200$	52
4.3	Runtime of PC-CGRM for the first scenario with sample size $n = 500$	52
4.4	Runtime of PC-CGRM for the first scenario with sample size $n = 1000$	53
4.5	Runtime of PC-CGRM for the second scenario with sample size $n = 200$	53
4.6	Runtime of PC-CGRM for the second scenario with sample size $n = 500$	53
4.7	Runtime of PC-CGRM for the second scenario with sample size $n = 1000$	54
4.8	Runtime of PC-CGRM for the second scenario with sample size $n = 2000$	54
5.1	Frequency distribution of CNVs in the 267 available patients for 9 out of 15587 genes	57
5.2	Summary indices for each category of CNV values.	57
5.3	Number of annotated and available genes for each REACTOME pathway resulting from MOSClip.	59

Introduction

Overview

Thanks to the advances in technology, bioinformatics and high-throughput methodologies of the last decades, a large unprecedented amount of biological data coming from various experiments in metabolomics, genomics and proteomics is available. This has led the researchers to conduct more and more comprehensive molecular profiling to study biological samples through different multiple aspects of genomic activities, – as DNA methylation, gene expression and co-expression, copy number variations and microRNA to name a few – thus introducing new challenges in the developments of statistical tools to integrate and model multi-omics data.

Normally, data analysis of genes and their activity is conducted at a single-level manner, in the sense that joint analysis of omics data recorded from different sources or of different types is not considered. But, generally speaking, investigating only single parts of a more complex and interconnected reality does not deliver fair, accurate or efficient results. This is why examining a (not only) biological phenomenon from multiple angles and using multiple types of data can provide important additional mechanistic insight, both in understanding its real processes and in comprehending how other events, dependent on it, can develop: some key examples are the spreading of a disease or how a cancer cell can gone haywire.

All those biological phenomena that take place inside the cell involve complex interaction between genes. When the analysis of genomic activity is conducted at a single-level manner, these dependencies are commonly represented in the form of a graph, or network, and have to be inferred from experimental data: a widely used tool in this case is the family of graphical models, whose strength relies in representing complex multivariate distributions by the adjacency structure of a graph. In this context, learning a graphical model from the data is equivalent, statistically speaking, to infer on the parameters of the multivariate distribution defined on the set of variables – the genes – at hand. Notice that, since we are in a single-level type of analysis, the genes must have all

the same domain: if we collected gene expression levels, we are assuming a multivariate Gaussian distributions, while, if we collected gene counts in a micro-array format, we must assume a multivariate distribution able to respect this counting nature.

But when the analysis of genomic activity operates in a multi-level manner, the situation becomes much more complex.

A first problem we meet is that trying to integrate all the different aspects in a single graph is not visually, and conceptually, the best of solutions, because the graph risks becoming too cramped and packed. A graph-like object capable of keeping track of multiple levels at once, usually seen in social and psychological studies, is a multilayer network. As far as we know, multilayer networks are up to now used as the algebraic structures underlying another type of statistical models called random graph models, whose key feature is to determine the probability that a particular property of the underlying network is likely to arise: for example, the most famous model is the Erdős-Rényi one, where all graph with a fixed number of edges have all the same probability to occur.

A second problem is the nature itself of the data collected: when we are dealing with data that detects different characteristics of the same genes, it is easy for these features to have different domains: DNA methylation is usually found in the form of count data, gene expression levels are usually continuous, copy number variations are interpreted as categorical variables, and so on. Up to now in a parametric framework, there are only two types of models that take into account the heterogeneous nature of the variables: the conditional Gaussian graphical models of Lauritzen (1996) and the mixed exponential Markov random fields of Yang *et al.* (2014b). The first model consists of Gaussian and categorical variables and it is defined in an exact manner, in the sense that it is not derived as an approximation of another distribution and it does not involve any type of penalization in its structural learning, but its number of parameters grows exponentially with the number of variables. On the other side, mixed exponential Markov random fields are such that each variable, with respect to the others, can be derived from different densities belonging to the exponential family of distributions by using a conditional modelling approach, so they are not restricted to Gaussian and categorical data, but they are usually approximations, since the computation of the normalising constant for these distributions is often not exact, and structural learning involves lasso penalization.

In this thesis, under some modelling constraints, we combine the above-mentioned approaches in defining our graphical model for heterogeneous data, and we tackle its structure learning on undirected multilayer networks. To the best of our knowledge,

this is the first attempt at combining these different objects in an unique framework.

The outline of the thesis is as follows. In Chapter 1, we briefly review some key concepts of graph theory and multilayer networks, we resume the notion and properties of conditional independence between random variables and Markov properties of a graph and we adapt them to random vectors and multilayer networks. Chapter 2 covers the state-of-the-art of graphical models for heterogeneous data, with a particular attention to conditional Gaussian models; Chapter 3 presents our proposal in the context of undirected heterogeneous graphical models, connecting it to the models discussed in Chapter 2. In Chapter 4 we address the structural learning techniques present in literature for heterogeneous graphical models and we present our approach, whose aim is to learn the structure of complex and heterogeneous data on multilayer networks. In Chapter 5 we apply our approach to a real data set of different genomic activities recorded in ovarian cancer cells. Chapter 6 contains main conclusions drawn from this project up to date and possible directions for future research.

Main contributions of the thesis

Main contributions of the thesis can be summarized as follows.

1. Definition of intra-layer and inter-layer separators as generalization of the concept of separator for multilayer networks.
2. Definition of Markov properties for random vectors on multilayer networks.
3. Definition of an heterogeneous graphical model, where heterogeneous is related to the mixed nature of the data, whose underlying structure corresponds to a multilayer network and able to represent the key feature of the models proposed by Lauritzen (1996) and Cheng *et al.* (2017).
4. Definition of a supervised learning algorithm of undirected heterogeneous multilayer network. The core of the algorithm resumes the approach of Yang *et al.* (2014b) for a single-level graph, where the neighbourhood of each node is estimated in turn by solving a lasso penalized regression problem and the resulting local structures stitched together to form the global graph. To face possible inaccurate inferences when dealing with models of high dimensions, we substitute penalized estimation with a likelihood ratio testing procedure on the parameters of the local regressions following the lines of the PC algorithm (Spirtes *et al.*, 2000) adapted to a multilayer structure.

5. Application of the novel approach on real data from ovarian cancer cells collections.

Chapter 1

Graphs, conditional independences and Markov properties

In this Chapter, we review and extend some of the mathematical backbones and probabilistic concepts essential for the understanding and the implementation of the models that will be exploited later on in this work. Since our main goal is the integration of different sources of information, we need a formal mathematical structure able to accommodate the problem. Graphs are exactly the ingredient that we necessitate; in particular, graphs such as multilayer or multilevel networks can offer a way to represent clearly and efficiently the complexity of the integration of different sources of data.

Probabilistic graphical models are an elegant framework linking the logical structure pictured by graphs to probability distributions. The two elements playing a key role in allowing such link to be established are conditional independences and Markov properties. This is the reason why graphs, conditional independences and Markov properties are the focus of this chapter.

It is worth noting that, in what follows, we do not make any distinction between a graph and a network. Albert-László Barabási, in one of his latest books (Barabási, 2016), affirms that, although in the scientific literature the terms network and graph are used interchangeably, there is a subtle distinction between the two terminologies. Usually, the term *network* is associated to real systems, like the WWW (a network of web documents linked by URLs) or society (a network of individuals linked by family, friendship or professional ties), whilst the term *graph* is associated to the mathematical representation of networks. Yet, this distinction is not structural in nature, because in their essence the two objects are the same: both are built by objects (called nodes and vertices) connected by other objects (links, edges). As the distinction is more at an interpretation level, we use the two terms indistinctly.

Let us go in detail on all the notions we spoke of so far.

1.1 Review on Graph theory

A *graph*, or network, is a mathematical object defined as an ordered pair $G = (V, E)$, where V is a finite set of *vertices* or nodes, and E is the set of *edges*, or links, that connect unordered pair of not necessarily distinct vertices of G , so that $E \subseteq V \times V$. The pair of vertices defining an edge are called *endpoints* of the edge, while the edge connecting a pair of vertices is said to *join* the vertices. The *order* of G is the number of vertices that compose it, and it is denoted as $v(G) = |V|$, while the *size* of G is the number of its edges, and it is denoted as $e(G) = |E|$, where $|\cdot|$ is the mathematical operator for cardinality.

If the pair of vertices defining an edge is not ordered, that edge is called *undirected*, and it is usually denoted as $\{u, v\} \in E$, or $u \sim v$, $u, v \in V$, whereas it is graphically represented by a line; the vertices u and v are said to be *adjacent* or *neighbours*, and the set of neighbours of a vertex v in a graph G is written as $\text{ne}(v)$. If all edges in E are undirected, then G is an *undirected graph*. Moreover, if the graph is made in such a way that every vertex is never connected with itself and, for each edge, there exists only one edge connecting the same pair of vertices, then the graph is called *simple* (i.e., no loops and no parallel edges). From now on, we focus on simple undirected graphs, since they will be the building blocks of more complex structures.

If we restrict the attention to $A \subseteq V$, a subset of the vertex set, this subset will *induce* a *subgraph* $G_A = (A, E_A)$, where the edge set $E_A = E \cap (A \times A)$ is obtained from G by keeping edges in E that have both endpoints in A . In this case, $\text{ne}(A)$ can be defined as the set of neighbours of vertices in A that are not themselves elements of A ,

$$\text{ne}(A) = \cup_{v \in A} \text{ne}(v) \setminus A.$$

Related to the concept of neighbourhood is the property of *closure* of A , defined as $\text{cl}(A) = A \cup \text{ne}(A)$, that is, the set of all vertices that are either in A or near A .

Given $u, v \in V$, a *walk* is a sequence $W := w_0 e_1 w_1 \dots w_{n-1} e_n w_n$, with $w_0 = u$ and $w_n = v$, whose terms are alternately vertices and edges of G not necessarily distinct, such that w_{i-1} and w_i are the endpoints of e_i , $i = 1, \dots, n$: in other words, a sequence of edges and vertices where each edge's endpoints are the two vertices adjacent to it; for simplicity, in general a walk is expressed only by the vertices it passes through, i.e., $W := w_0, w_1, \dots, w_{n-1}, w_n$. A *path* from u to v is a walk where all vertices are distinct, except possibly the first and last. The shortest path from u to v is the path from u

to v having minimal length. The subscript n represents the length of the walk or path (equivalently, the number of edges along it); furthermore, we say that u and v *connect* and denote it with $u \rightleftharpoons v$. A graph is *connected* if, for every partition of its vertex set into two non-empty sets A and B , there is at least one edge with one endpoint in A and the other endpoint in B (or, there is at least a path or walk between vertices in A and vertices in B); if that is not the case, the graph is *disconnected*. A subset $S \subseteq V$ is called *uv-separator* if all paths from u to v intersect S . Given $A, B, S \subseteq V$, the subset S separates A from B if it is an *uv-separator* for all $u \in A$ and $v \in B$.

It is possible to find the notions explained so far in any book on graph theory, although we have relied on recent literature (Bondy and Murty, 2008). These concepts are also illustrated through an example in Figure 1.1. Panel (a) shows a simple undirected and connected graph $G = (V, E)$, with $V = \{1, 2, 3, 4, 5, 6, 7\}$ and edge set

$$E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{4, 6\}, \{6, 7\}, \{5, 7\}\}.$$

The order and size of G are $v(G) = 7$ and $e(G) = 8$, respectively. The set of neighbours of, for example, vertex 3 is $ne(3) = \{1, 2, 4\}$. Panel (b) shows the simple undirected and connected subgraph of G , $G_A = (A, E_A)$, induced by $A = \{1, 2, 3, 4\} \subset V$, where $E_A = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}\}$. The neighbourhood of A is $ne(A) = \{5, 6\}$, and its closure is $cl(A) = \{1, 2, 3, 4, 5, 6\} = V \setminus \{7\}$. Possible walks from, for example, 3 to 6 are $W_1 : 3, 4, 5, 7, 6$, $W_2 : 3, 4, 6$ or $W_3 : 3, 4, 5, 4, 6$; W_2 is the shortest path from 3 to 6. A possible separating set S for G could be $S = \{3, 4\}$, since its deletion entails a partition of G into two disconnected sets.

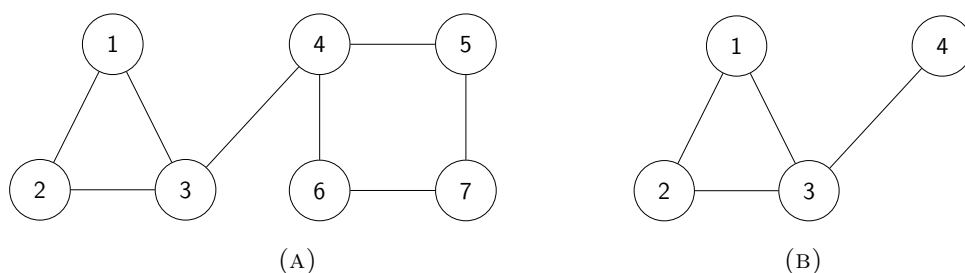


FIGURE 1.1: (A) A simple undirected and connected graph $G = (V, E)$ and (B) the subgraph of G induced by $A = \{1, 2, 3, 4\}$.

It will be useful to introduce the definition of *marked graphs* proposed in Lauritzen (1996): their main characteristic is that the vertices are partitioned into groups, differentiating nodes representing qualitative variables from the ones corresponding to quantitative variables. If we use Δ to represent the former set of vertices and Γ to represent the latter, then we can decompose V as $V = \Delta \cup \Gamma$, with $\Delta \cap \Gamma = \emptyset$. They

are correspondingly pictured by *dots* (black coloured nodes) if the vertices belong to Δ (*dots* for *discrete* variables), and by *circles* (white coloured nodes) if the vertices belong to Γ (*circles* for *continuous* variables). These types of graphs we will be helpful to visually reproduce the models we will infer, since we will model qualitative and quantitative variables at the same time.

1.2 Multilayer networks

1.2.1 Notation and terminology

The key ingredients in these structures are the presence of *layers* and *aspects* in addition to vertices and edges. By layers, we intend a combinations of features of the nodes, e.g., a task, or activity, or category: for example, they can represent different types of connections involved in social networks, or various air companies connecting airports around the globe. Aspects instead help coordinate relationships between different layers, e.g., time or locations. In the most general definition of multilayer networks given by Kivela *et al.* (2014), each node can belong to any subset of the layers, in turn included in aspects, and edges can involve any pairwise connection between the possible combinations of nodes and layers: hence, a node u in layer α of aspect \mathbf{a} can be connected to node v in layer β of aspect \mathbf{b} .

Formally speaking, let V be the set of vertices and d the number of aspects; define a sequence $\mathbf{L} = \{L_a\}_{a=1}^d$ of sets of elementary layers (layers whose combination give rise to layers containing information from different aspects) such that a set of elementary layers L_a exists for each aspect a . As explained in Kivela *et al.* (2014), it is possible to construct a sequence of layers by assembling a set of all of the combinations of elementary layers using the Cartesian product of $L_1 \times \dots \times L_d$. Moreover, for each choice of node and layer, define $V_M \subseteq V \times L_1 \times \dots \times L_d$ as the subset containing only the node-layer combinations in which a node is present in the corresponding layer. For example, the node-layer $(u, \alpha_1, \dots, \alpha_d)$ represents node u on layer $\alpha = (\alpha_1, \dots, \alpha_d)$ (it is abbreviated sometimes as (u, α)). A pair of node-layers directly connected by an edge are said to be adjacent. Since the definition of multilayer network so far used is the most general one, it requires the occurrence of all possible edges between any pair of node-layers, so the edge set E_M is defined as a set of pairs of possible combinations of nodes and elementary layers, i.e., $E_M \subseteq V_M \times V_M$. Putting everything together, a multilayer network is then a quadruplet $M = (V_M, E_M, V, \mathbf{L})$, where V_M, E_M, V, \mathbf{L} are the quantities defined above.

Note that, if the number of aspects is 0 (i.e., $d = 0$), then the multilayer network M is a single graph, where $V_M = V$ and $E = E_M \subseteq V \times V$.

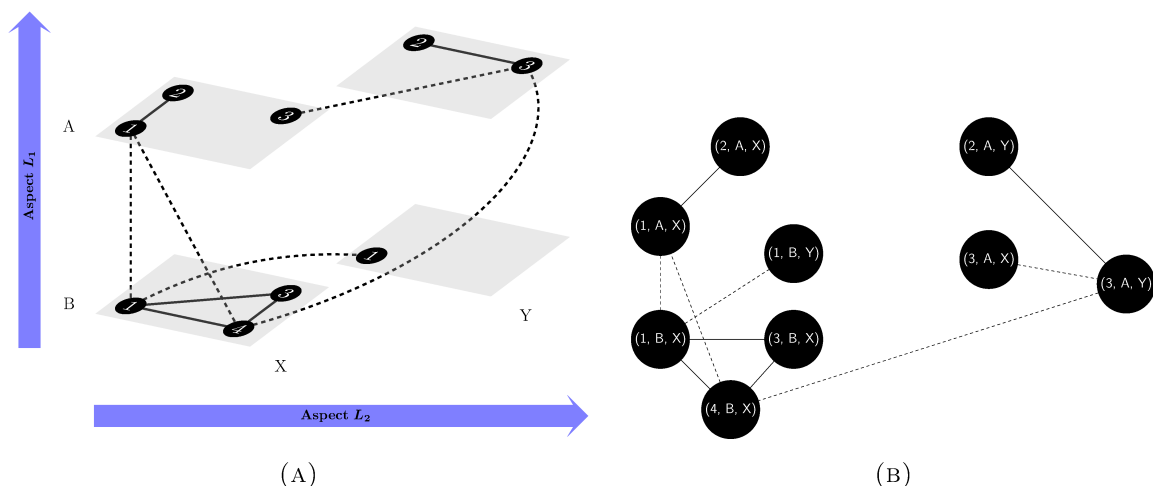


FIGURE 1.2: A multilayer network $M = (V_M, E_M, V, \mathbf{L})$ (panel (A)) and the graph $G_M = (V_M, E_M)$ underlying M (panel (B)).

Figure 1.2 illustrates, with the help of an example, some of the concepts discussed so far. Panel (a) shows a possible example of multilayer network $M = (V_M, E_M, V, \mathbf{L})$. It has a total of four nodes, i.e., $V = \{1, 2, 3, 4\}$, and two aspects, which have corresponding elementary-layer sets $L_1 = \{A, B\}$ and $L_2 = \{X, Y\}$, giving rise to a total of four different layers: (A, X) , (B, X) , (A, Y) and (B, Y) . Each layer contains some subset of V . The set of node-layer tuples is

$$V_M = \{(1, A, X), (2, A, X), (3, A, X), (2, A, Y), (3, A, Y), (1, B, X), (3, B, X), (4, B, X), (1, B, Y)\} \subset V \times L_1 \times L_2.$$

The nodes connect to each other in a pairwise manner both within and between layers. The edges inside each layer are drawn as solid lines, while edges connecting node-layer tuples in different layers are marked with dashed lines. Panel (b) shows the same multilayer network M , but interpreted as an undirected graph $G_M = (V_M, E_M)$: its edges are showed in the same manner as if they would be drawn in a multilayer network structure, but the labels of the nodes now are changed, reflecting the fact that the set V_M is now the set of nodes of G_M .

The transformation shown in Figure 1.2, panel (b), can be applied to any type of multilayer network, in the sense that the first two elements of an arbitrary multilayer network M can always generate a graph $G_M = (V_M, E_M)$. This double perspective allows to interpret a multilayer network as a graph with specifically labelled nodes. It appears natural, therefore, define a multilayer network M *undirected* if the underlying

graph G_M is undirected. Similarly, if G_M is a simple graph, then M is a *simple* multilayer network. From now on, we will consider only simple undirected multilayer networks, because they will be the basic structures on which we will build our models.

Furthermore, it is typical to distinguish between edges that cross layers and edges inside a single layer: the latter are called *intra-layer edges* and are defined as the set $E_A = \{((u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta})) \in E_M : \boldsymbol{\alpha} = \boldsymbol{\beta}\}$, while the former are called *inter-layer edges*, and are defined as the complementary set of E_A in E_M , i.e., $E_C = E_M \setminus E_A = \{((u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta})) \in E_M : \boldsymbol{\alpha} \neq \boldsymbol{\beta}\}$. Moreover, define as *coupling edges* the subset $E_{\tilde{C}} \subseteq E_C$ containing those edges for which the two nodes represent the same entity in different layers, that is $E_{\tilde{C}} = \{((u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta})) \in E_C : u = v\}$. To these different types of edges correspond different types of graphs (or subgraphs of G_M): we obtain an *intra-layer graph* $G_A = (V_M, E_A)$, an *inter-layer graph* $G_C = (V_M, E_C)$ and a *coupling graph* $G_{\tilde{C}} = (V_M, E_{\tilde{C}})$.

Since a generic multilayer network M can be interpreted as a graph $G_M = (V_M, E_M)$ whose node labels are the tags of the node-layer tuples of M , the concepts of *neighbourhood*, *walk* and *path* for M are the same as the ones we saw in Section 1.1. Indeed, for the particular use we make of it, different copies of the same node in different layers are considered to be a set of distinct objects, because they will represent different measurements of the same quantity. This means that passing from a layer to another is considered a step inside a walk, and each step is defined as occurring between a pair of node-layer tuples of M . Also the concept of marked graph is easily transferred to the case of multilayer networks: indeed, marked graphs are a particular class of node-coloured graphs, i.e., graphs whose nodes are labelled by colors. In particular, a legal colouring means no adjacent nodes have the same color, while an illegal colouring does not impose such a restriction. Since there is no such constraint on the distributions of colours on the nodes of a marked graph (two adjacent vertices can be both discrete, hence having the same colour), the latter can therefore be thought of as node-coloured graph where the colouring is illegal and the colour used are two (black for qualitative variables and white for quantitative ones). In turn, (illegal) node-coloured graphs can be interpreted as multilayer network, having $d = 1$ aspects and considering each layer as one of possible colours provided by the vertices of the graph: for marked graphs, then the corresponding multilayer networks have two layers, one for qualitative variables (the “black” layer) and one for quantitative variables (the “white” layer).

Figure 1.3 shows an example of a marked graph interpreted as a multilayer network. Panel (a) shows a simple undirected marked graph G_{marked} . The multilayer network M corresponding to G_{marked} is shown in panel (b). The color of nodes of G_{marked} are

transferred to the layers of M , while the grey colour for the nodes in M is only for aesthetic reasons.

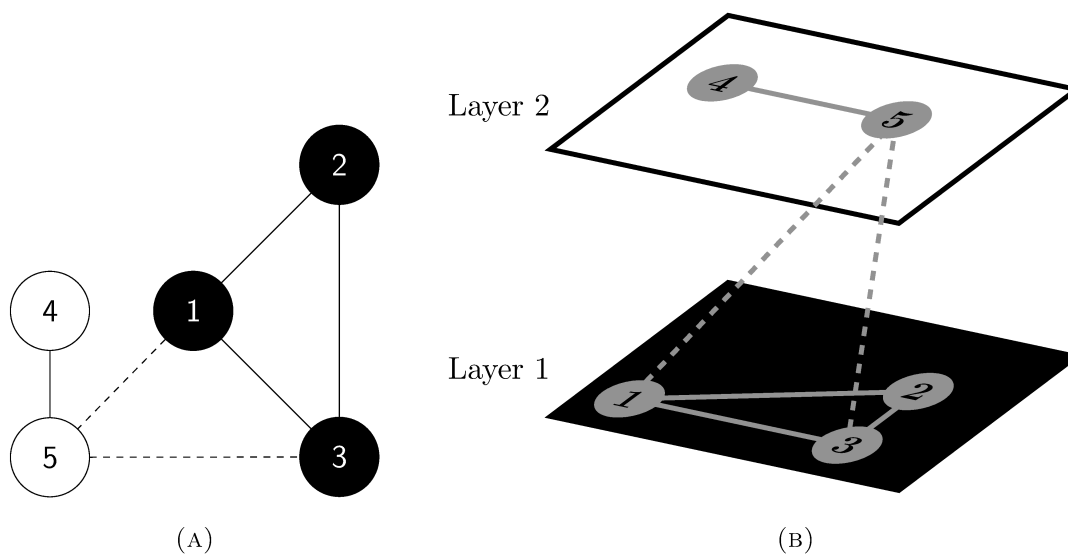


FIGURE 1.3: A marked graph (panel A) pictured as a multilayer network (panel B).

1.2.2 Some new definitions: multi-, intra-, inter-layer separators

More care deserves the concept of separation in connected multilayer networks. Although the representation through G_M could allow to retrieve the definitions of connectivity and separation seen in Graph Theory, the characterization of separation for multilayer networks is complicated by the differentiation between intra-layer and inter-layer edges. In this Section, we will introduce some new definitions that allow to specialize the concept of separation in multilayer networks.

Definition 1.1. A multilayer network $M = (V_M, E_M, V, \mathbf{L})$ is *connected* if there is at least a path from any node-layer $(u, \alpha) \in V_M$ to any other node-layer $(v, \beta) \in V_M$ in M .

This definition is made possible because paths, so as walks, are assumed to be composed by different type of edges. Clearly, if a multilayer network is not connected, it is said to be *disconnected*. Disconnectedness is easy to identify, because there are few basic cases in which we can find it. Other cases are only a product of the combination of the conditions found in the following proposition.

Proposition 1. A multilayer network $M = (V_M, E_M, V, \mathbf{L})$ is *disconnected* if and only if at least one of the following conditions is true:

- (a) there exists at least one layer α whose intra-layer graph $G_\alpha = (V_\alpha, E_\alpha)$ is disconnected and such that there is no inter-layer edge that connects the different partitions of G_α ;
- (b) there exists at least one isolated layer α in M , i.e., a layer whose vertices are not endpoints of any inter-layer edge in E_M .

Proof. If M is disconnected, then also its representation as a graph, $G_M = (V_M, E_M)$, is disconnected. But disconnectedness happens if and only if there exist two nodes of V_M , let us call them (u, α) and (v, β) , such that no path in G_M has those nodes as endpoints. Hence, we consider two cases: (i) (u, α) and (v, β) belong to the same layer (i.e., $\alpha = \beta$), and (ii) (u, α) and (v, β) belong to two different layers. In situation (i), the two nodes can be written as (u, α) and (v, α) ; but then no path exists between them if there is no sequence of intra-layer edges in $E_\alpha \subseteq E_A$ and inter-layer edges in E_C that connects them, implying that the intra-layer graph $G_\alpha = (V_\alpha, E_\alpha)$, $V_\alpha \subseteq V_M$, is disconnected and no inter-layer edge connects its different partitions; this is indeed the condition (a). For situation (ii), let us make the further assumption that all intra-layer graphs are connected: then, no path exists between (u, α) and (v, β) if there are no inter-layer edges that touch the layer α , β or any layer in between; this means that there is a layer whose vertices are endpoints of only intra-layer edges and not inter-layer edges, which is the situation described in point (b). \square

The cases covered in the above given proposition can be visualized, for a multilayer network M with $d = 1$ aspects, in Table 1.1. In particular, it is visualized also a special case of condition (a), which is when at least one of the endpoints of the path of interest is an isolated node-layer tuple, i.e., a node-layer tuple with no edges incident on it.

Such characterization of disconnection opens to a definition of separator set slightly more articulated than the one we saw before. Given a multilayer network $M = (V_M, E_M, V, \mathbf{L})$, the most general definition of separator for two node-layer tuples is obtained by avoiding any type of discrimination between the types of edges in E_M .

Definition 1.2. Let $M = (V_M, E_M, V, \mathbf{L})$ be a multilayer network and $S \subseteq V_M$. Then S is a $u^\alpha v^\beta$ -separator if all paths from the node-layer $(u, \alpha) \in V_M$ to the node-layer $(v, \beta) \in V_M$ intersect S .

Definition 1.3. Let $M = (V_M, E_M, V, \mathbf{L})$ be a multilayer network and $A, B, S \subseteq V_M$. Then S is a *multi-layer separator* of A from B if S is a $u^\alpha v^\beta$ -separator from all $(u, \alpha) \in A$ to all $(v, \beta) \in B$.

If we take into account the fact that paths can run along intra-layer as inter-layer edges, we can furthermore specify the definition of S as a multilayer separator

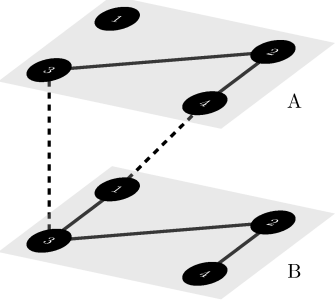
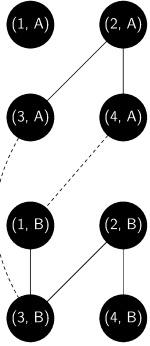
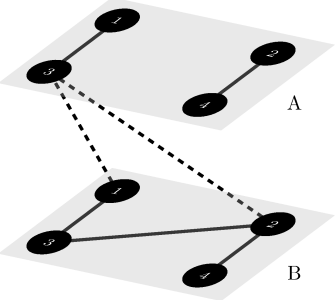
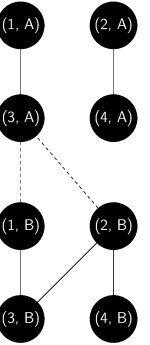
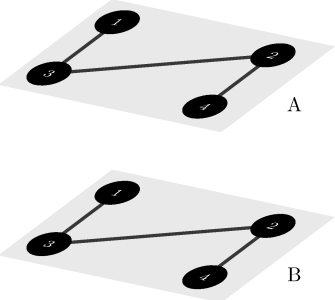
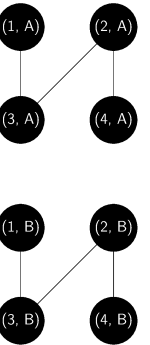
Multilayer network M	Graph G_M	Source of disconnectedness
		Isolated node-layer $(1, A)$. It is a special case of condition (a) of Proposition 1.
		Disconnected intra-layer graph $G_A = (V_A, E_A)$, with vertex set $V_A = \{(u, A) : u \in \{1, 2, 3, 4\}\}$, edge set $E_A = \{((1, A), (3, A)), ((2, A), (4, A))\}$, and such that the partition $\{(2, A), (4, A)\}$ is never touched by any inter-layer edge.
		Isolated layers A and B .

TABLE 1.1: Examples of basic cases leading to disconnectedness in a multilayer network M with $d = 1$ aspects.

Definition 1.4. Let $M = (V_M, E_M, V, \mathbf{L})$ be a multilayer network and $S \subseteq V_M$ a multilayer separator in M . If S is composed only by *some* node-layer tuples of a single layer, then it is said to be an *intra-layer* separator, because these tuples are connected only through intra-layer edges.

Equivalently, if S is an intra-layer separator on layer α and it is cut out from M , then the related intra-layer graph G_α is disconnected, such that there is no inter-layer edge that connects the different partitions of G_α , thus the condition (a) of Proposition 1 is satisfied.

Definition 1.5. Let $M = (V_M, E_M, V, \mathbf{L})$ be a multilayer network and $S \subseteq V_M$ a multilayer separator in M . If S is an *entire* layer, then it is said to be an *inter-layer* separator,

because the only way not to interrupt any path passing through that layer is to arrive and exit from it with inter-layer edges.

Equivalently, if S is an inter-layer separator and it is cut out from M , condition (b) of Proposition 1 comes to be, disconnecting the multilayer network in disjoint partitions.

Examples of different separators in multilayer networks can be visualized in Figure 1.4. Panel (a) contains the original multilayer network. In panel (b), a generic multilayer-separator S is highlighted in a purple colour, composed by seven node-layer couples, of which three lying on layer A $((2, A), (3, A), (4, A))$, and the other four on layer B $((1, B), (2, B), (3, B), (4, B))$. Panel (c) shows an intra-layer separator, highlighted in light green, defined as the set $S = \{(2, A), (3, A)\}$. Panel (d) shows the inter-layer separator $S = \{(1, B), (2, B), (3, B), (4, B)\}$, highlighted in light pink. In panels (b), (c), and (d), the edges directly involved in the path precluded by the separator are also coloured, purple, light green and light pink, respectively.

1.3 Conditional independence

Let X, Y, Z be random variables with a joint distribution P . Then, X is said to be *conditionally independent* of Y given Z under P , written $X \perp\!\!\!\perp Y \mid Z$, if, for any measurable set A in the sample space of X there exists a version of the conditional probability $P(A \mid Y, Z)$ which is a function of Z alone (Lauritzen, 1996). In particular, if X, Y and Z are discrete random variables, $X \perp\!\!\!\perp Y \mid Z$ implies

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z), \quad (1.1)$$

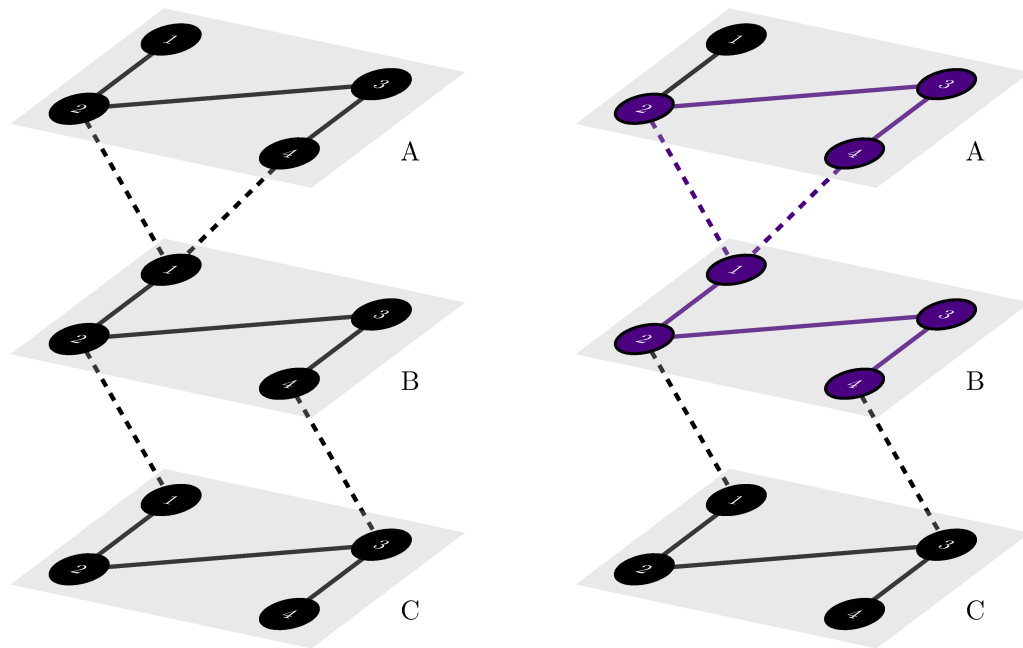
if and only if $P(Z = z) > 0$ for all possible values z . When X, Y and Z admit a joint density with respect to a product measure μ , then $X \perp\!\!\!\perp Y \mid Z$ implies

$$f_{XY|Z}(x, y \mid z) = f_{X|Z}(x \mid z)f_{Y|Z}(y \mid z), \quad (1.2)$$

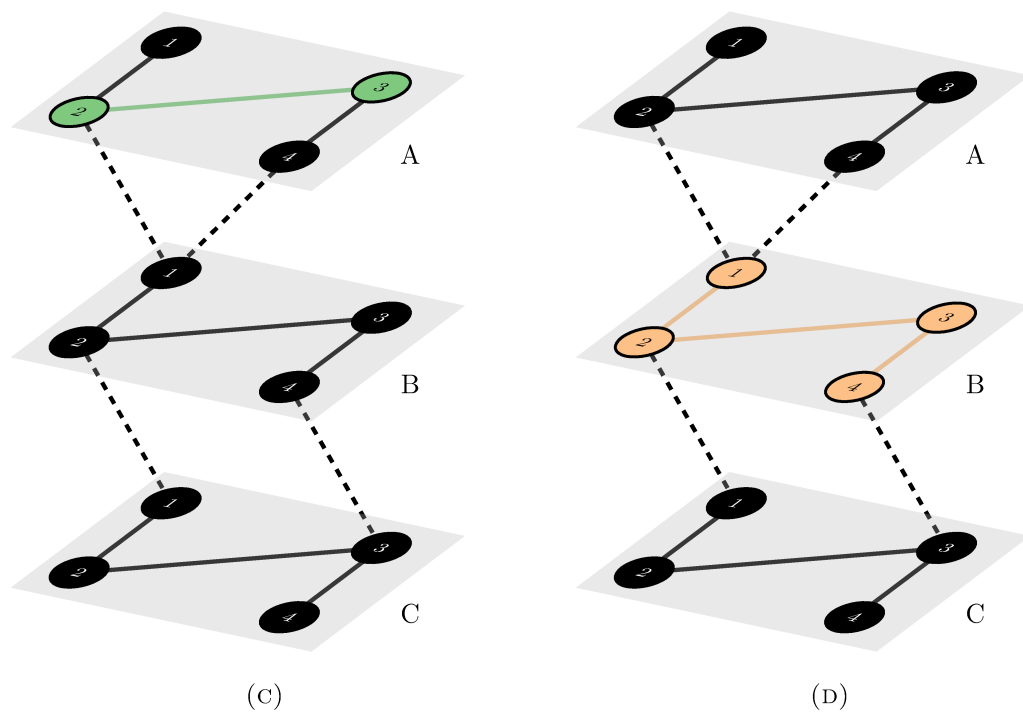
holding almost surely with respect to P . If all densities are continuous, then (1.2) must hold for all z such that $f_Z(z) > 0$. Notice also that, if all densities are continuous, and for all possible values of z ,

$$f_{XY|Z}(x, y \mid z) = f_{X|Z}(x \mid z)f_{Y|Z}(y \mid z) \Leftrightarrow f_{XYZ}(x, y, z)f_Z(z) = f_{XZ}(x, z)f_{YZ}(y, z).$$

The definition of conditional independence between *random vectors*, i.e., vectors of

(A) Example of multilayer network M .

(B)



(c)

(d)

FIGURE 1.4: Examples of multilayer separators for the (connected) multilayer network M of panel (a): multi-layer separator (panel (b)); intra-layer separator (panel (c)); inter-layer separator (panel (d)). In panels (b), (c), and (d), the edges directly involved in the path precluded by the separator are coloured, purple, light green and light pink, respectively.

random variables, is easily derived using the conditional cumulative distribution function: let $\mathbf{X} = (X_1, \dots, X_n)^T$, $\mathbf{Y} = (Y_1, \dots, Y_m)^T$ and $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ be three random vectors with a joint distribution (density) P . Then \mathbf{X} is conditionally independent of \mathbf{Y} given \mathbf{Z} under P , written $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$, if and only if they are independent in their conditional cumulative distribution given \mathbf{Z} for all their possible values $\mathbf{x} = (x_1, \dots, x_n)^T$, $\mathbf{y} = (y_1, \dots, y_m)^T$ and $\mathbf{z} = (z_1, \dots, z_p)^T$,

$$F_{\mathbf{XY}|\mathbf{Z}}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = F_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} \mid \mathbf{z})F_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} \mid \mathbf{z}),$$

with

$$F_{\mathbf{XY}|\mathbf{Z}}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = P(X_1 \leq x_1, \dots, X_n \leq x_n, Y_1 \leq y_1, \dots, Y_m \leq y_m \mid Z_1 = z_1, \dots, Z_p = z_p),$$

$$F_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} \mid \mathbf{z}) = P(X_1 \leq x_1, \dots, X_n \leq x_n \mid Z_1 = z_1, \dots, Z_p = z_p),$$

$$F_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} \mid \mathbf{z}) = P(Y_1 \leq y_1, \dots, Y_m \leq y_m \mid Z_1 = z_1, \dots, Z_p = z_p).$$

In what follows, for the purpose of simplicity, we will limit ourselves to the consideration of random variables, and not of random vectors.

It is useful to recap some elementary properties of conditional independence for random variables. In particular, the relation $X \perp\!\!\!\perp Y \mid Z$ enjoys the following properties, whose proofs are given in Appendix A. Let h be an arbitrary measurable function on the sample space of X ; then,

(P1) if $X \perp\!\!\!\perp Y \mid Z$, also $Y \perp\!\!\!\perp X \mid Z$ (symmetry);

(P2) if $X \perp\!\!\!\perp Y \mid Z$ and $U = h(X)$, also $U \perp\!\!\!\perp Y \mid Z$;

(P3) if $X \perp\!\!\!\perp Y \mid Z$ and $U = h(X)$, also $X \perp\!\!\!\perp Y \mid (Z, U)$;

(P4) if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, also $X \perp\!\!\!\perp (W, Y) \mid Z$.

Another property of the conditional independence relation worth considering is the following:

(P5) If $X \perp\!\!\!\perp Y \mid Z$, $X \perp\!\!\!\perp Z \mid Y$, and there is no non-trivial logical relationship between Y and Z , then $X \perp\!\!\!\perp (Y, Z)$.

This condition is always true when the joint distribution (density) of all variables with respect to a product measure μ is positive and continuous (see Lauritzen (1996) for further proof).

Core ideas expressed by properties (P1) to (P4) can be used to define a new mathematical structure able to capture conditional independence or, in a more general way

of saying things, the irrelevance of certain events in interpreting others. This structure is called *semi-graphoid* and was introduced by Pearl in 1988 (Pearl, 1988). In this algebraic framework, X, Y and Z are not random variables, but arbitrary disjoint subsets of a finite set, and $U = h(X)$ is replaced by the notation $U \subseteq X$. If we suppose that the corresponding algebraic version of the condition (P5) holds also in this more abstract context, then the structure is called *graphoid*. A remarkable example for our application of this latter structure is given by the concept of a separator in a graph: indeed, let A, B, S be disjoint subsets of the vertex set V of a simple undirected graph $G = (V, E)$, with E defined as in the previous section, and define

$$A \perp^G B \mid S \Leftrightarrow S \text{ separates } A \text{ from } B \text{ in } G.$$

Then, the corresponding properties (P1) – (P5) in terms of Graph Theory become the following:

$$(P1) \text{ If } A \perp^G B \mid S, \text{ then } B \perp^G A \mid S.$$

$$(P2) \text{ If } A \perp^G B \mid S \text{ and } U \subseteq A, U \perp^G B \mid S.$$

$$(P3) \text{ If } A \perp^G B \mid S \text{ and } U \subseteq A, A \perp^G B \mid S \cup U.$$

$$(P4) \text{ If } A \perp^G B \mid S \text{ and } A \perp^G D \mid S \cup B, \text{ whatever } D \subseteq V, \text{ then } A \perp^G B \cup D \mid S.$$

$$(P5) \text{ If } A \perp^G B \mid S \text{ and } A \perp^G S \mid B, \text{ then } A \perp^G B \cup S.$$

It is easy to see that (P1) – (P5) are valid also for a multilayer network M because of the correspondence between M itself and its representation as graph $G_M = (V_M, E_M)$, so that the concepts of multi-, intra- and inter-layer separator are nothing else than particular cases of a general separator $S \subseteq V_M$ in G_M .

1.4 Markov properties

In this last section, we combine the probabilistic notion of conditional independence with the notion of separation in a (multilayer) graph, this being the last necessary step to do before switching to the inferential models. From now on, the nodes of (multilayer) graphs always represent random variables (vectors) and the edges the type of stochastic dependence between them.

1.4.1 Standard definitions

Let V be the vertex set of a simple undirected graph $G = (V, E)$, with $V = \{1, \dots, p\}$. Suppose to have a collections of random variables X_1, \dots, X_p , taking values in the set of probability spaces $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_p\}$, which can be real-finite dimensional vector spaces or finite and discrete sets. Here, X_v represents the random variable associated to node v , $v \in V$. For $A \subseteq V$, let $\mathcal{X}_A = \times_{j \in A} \mathcal{X}_j$ denote the probability spaces of the random variables with indexes in A , and $\mathbf{x}_A = \{x_j : j \in A\}$ and $\mathbf{X}_A = \{X_j : j \in A\}$ the elements of \mathcal{X}_A and the random variables, respectively. Then, a probability measure P on \mathcal{X} is said to obey

(P) the *pairwise Markov property* (*pairwise MPs*), relative to G , if, for any pair (u, v) of non-adjacent vertices,

$$X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{V \setminus \{u, v\}};$$

(L) the *local MPs*, relative to G , if, for any vertex $u \in V$,

$$X_u \perp\!\!\!\perp \mathbf{X}_{V \setminus \text{cl}(u)} \mid \mathbf{X}_{\text{ne}(u)};$$

(G) the *global MPs*, relative to G , if, for any triple (A, B, S) of disjoint subsets of V such that $A \stackrel{G}{\perp} B \mid S$,

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S.$$

As Lauritzen (1996) shows in his book, all these MPs are related to each other: in the simplest situation, i.e., in case of undirected graph G and for any probability distribution (density) on \mathcal{X} , it holds that $(G) \Rightarrow (L) \Rightarrow (P)$, being a direct consequence of properties (P1) – (P4) for conditional independence. So, $(G) \Rightarrow (L) \Rightarrow (P)$ holds also for any semi-graphoid in G . Suppose that also (P5) is valid, that in this context can be written as

(P5) if A, B, C, D are all disjoint subsets of V , $A \perp\!\!\!\perp B \mid C \cup D$ and $A \perp\!\!\!\perp C \mid B \cup D$, then $A \perp\!\!\!\perp B \cup C \mid D$.

Then, as it was proved by Pearl and Paz (1987), also the reverse implication is true, meaning that $(P) \Rightarrow (L) \Rightarrow (G)$. In particular, this is true if P has a positive and continuous density with respect to a product measure μ , and therefore to any graphoid in G .

As conditional independence is strictly linked to the factorization of P , so are the Markov properties. A probability measure P on \mathcal{X} is said to *factorize according to G*

if, for all complete subsets $A \subseteq V$, there exists non-negative functions ψ_A that depend on \mathbf{x} only through \mathbf{x}_A and there exists a product measure $\mu = \times_{j \in V} \mu_j$ on \mathcal{X} such that P has density $f_{\mathbf{X}}(\mathbf{x})$ with respect to μ . If P factorises according to G , then P has property (F) and the set of such probability measures is written as $M_F(G)$. Then, for any undirected graph G and for any probability distribution (density) on \mathcal{X} , it is true that (F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P). To obtain the opposite direction of the implication, we have to invoke the Hammersley-Clifford theorem:

Theorem 1.6. *A probability distribution P with positive and continuous density $f_{\mathbf{X}}(\mathbf{x})$ with respect to a product measure μ satisfies the pairwise MP with respect to an undirected graph G if and only if it factorizes according to G .*

The proof can be found in Lauritzen (1996). In other words, this theorem presents the necessary and sufficient conditions under which a strictly positive distribution (density) function can be represented as undirected graph, linking the concepts of adjacency and connectivity of Graph Theory with the conditional independence property on Probability Theory.

1.4.2 Extension to multilayer networks

But how do MPs translate into the world of multilayer networks? Once again, it is very useful to go back to the representation of $M = (V_M, E_M, V, \mathbf{L})$ via graph $G_M = (V_M, E_M)$, because all the properties discussed so far can be applied – with related consequences and advantages – as for a generic undirected graph G . Although it is a trivial and intuitive passage, for our knowledge it was never attempted in literature.

For this conversion to be done, it is necessary to work with node-layer tuples (which are, actually, the nodes of G_M), instead of nodes. Indeed, we can restate the previous results starting from a collection of random vectors $\mathbf{X} = \{\mathbf{X}_{\alpha}, \dots, \mathbf{X}_{\beta}\}$, with $\mathbf{X}_{\alpha} = \{X_{(u, \alpha)}\}_{(u, \alpha) \in V_M}$, arbitrary α . Each of these vectors takes values in the set of probability spaces \mathcal{X} , which can be a Cartesian product of real-finite dimensional vector spaces or finite and discrete sets. If $A \subseteq V$, we use the notation \mathcal{X}_A to indicate the probability spaces of the random variables vectors with indexes in A , while we denote as \mathbf{x}_A and \mathbf{X}_A the elements of \mathcal{X}_A and the random vectors, respectively. Moreover, let V be the vertex set of a simple undirected multilayer network $M = (V_M, E_M, V, \mathbf{L})$, with $V_M = \{(u, \alpha) : u \in V \text{ and } \alpha \in \mathbf{L}\}$. Then, a probability measure P on \mathcal{X} is said to obey

(P) the *pairwise Markov property* (*pairwise MPs*), relative to M , if, for any pair $((u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta}))$ of non-adjacent node-layer tuples,

$$X_{(u, \boldsymbol{\alpha})} \perp\!\!\!\perp X_{(v, \boldsymbol{\beta})} \mid \mathbf{X}_{V_M \setminus \{(u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta})\}};$$

(L) the *local MPs*, relative to M , if, for any node-layer tuple $(u, \boldsymbol{\alpha}) \in V_M$,

$$X_{(u, \boldsymbol{\alpha})} \perp\!\!\!\perp \mathbf{X}_{V_M \setminus \text{cl}((u, \boldsymbol{\alpha}))} \mid \mathbf{X}_{\text{ne}((u, \boldsymbol{\alpha}))};$$

(G) the *global MPs*, relative to M , if, for any triple (A, B, S) of disjoint subsets of V_M such that S is a multilayer-separator of A and B ,

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S.$$

Markov properties for the case of multilayer networks follow immediately by what was already proved in Lauritzen (1996), simply by replacing G with G_M and updating the node labels of G_M with the labels of the node-layer tuples of M . Consequently, the sequence of implications (P) \Leftrightarrow (L) \Leftrightarrow (G) combined with the Hammersley-Clifford theorem hold also supposing an underlying undirected multilayer network structure, meaning that the latter can be used to represent a strictly positive distribution (density).

Chapter 2

Background in probabilistic graphical models for heterogeneous data

In this chapter, we will explore the major statistical and graphical models proposed in the literature to incorporate heterogeneous data, i.e., data with different domains (e.g., variables coming from different sources or with different behaviours). After a general and concise review on the literature up to date, we will direct our attention on the conditional Gaussian models of Lauritzen and Wermuth (1989), being them the backbone structure of our model, which will be presented in the next chapter.

Since we will be working with undirected and simple multilayer networks, we will focus only on probabilistic graphical models on undirected (multilayer) graphs, called simply *undirected graphical models* (GMs) or *Markov random fields* (MRFs).

2.1 Literature review

Despite the increasingly massive presence of data of different nature, in reality there are few probabilistic graphical models able to manage and represent them.

As stated in the extensive review on graphical models for heterogeneous data of Dimitrova and Kocarev (2018), a first approach to simultaneously model variables with various domains was to interpret them as derivatives of latent Gaussian MRFs (MRFs whose underlying distribution is Gaussian). Various approaches were proposed along this line. Rue *et al.* (2009) proposed the use of integrated nested Laplace approximation (INLA) techniques to infer the conditional independence between the converted variables; Dobra and Lenkoski (2011) employed the Gaussian copula. From a Bayesian perspective,

Bhadra *et al.* (2018) made use of Gaussian scale mixtures to model not-normal continuous data, or mixed continuous and discrete variable as well. Other approaches rely on non-parametric statistics, like penalty algorithms, or rank-based correlation estimators (Liu *et al.* 2009; Liu *et al.* 2012; Xue and Zou 2012; Fan *et al.* 2016).

All the above mentioned approaches show two major drawbacks. Generally speaking, acting on the original domain of the variables almost always causes loss of information. More specifically, all these works have clearly extended the use of Gaussian graphical models to continuous data, but they are not explicitly appropriate for treating binary variables and, therefore, cannot handle variables with heterogeneous domains in a proper way.

A pioneering class of parametric models able to handle not homogeneous data was presented thirty years ago by Steffen Lauritzen and Nanny Wermuth (Lauritzen and Wermuth, 1989) and later deeply discussed in Lauritzen (1996), with the name of *conditional Gaussian* graphical models. Such models, that will be discussed in the next section, can deal with Gaussian and categorical variables. One of their major drawbacks, is that the number of parameters grows in an exponential manner when increasing the number of variables. Hence, this family is not particularly suitable in high-dimensional contexts, such as, for example, the multi-omics ones. Further developments and generalizations were proposed to solve this issue: at first, the conditional Gaussian model was expanded in a hierarchical interaction model in Edwards (1990) and Edwards *et al.* (2010) – higher order interactions between variables are only possible if all lower order sub-interactions are included in the model –; more recently, a series of works expanded the range of algorithms to estimate the structure of the underlying graph of these models, as presented in the papers of Lee and Hastie (2015), Fellinghauer *et al.* (2013) and Cheng *et al.* (2017).

Another approach is offered by *mixed exponential* MRFs introduced by Yang *et al.* (2014b). This approach looks more general with respect to conditional Gaussian graphical models approach. Indeed, such models are based on the theory introduced by Wainwright and Jordan (2008) and built on deriving an approximation of the distribution of the graph structure through variational approximation techniques for which exact calculations are possible. The principal advantage of these mixed exponential MRFs is the idea that each node can be (conditionally) distributed as a member from a different exponential family distribution. Other works in this direction were presented in Chen *et al.* (2015), where the focus is on pairwise graphical models (i.e., graphical models containing at most interaction effects of order 2), Tansey *et al.* (2015), that expand the work of Yang *et al.* (2014b) in a vector-space framework, and Haslbeck and

Waldorp (2015), contributing with a structure estimation approach based on generalized covariance matrices able to maintain all variables on their proper domain.

2.2 The main instigator: conditional Gaussian GMs

In this section, we review the main results on conditional Gaussian models. See Lauritzen and Wermuth (1989) and Lauritzen (1996) for an extensive treatment.

Let $G = (V, E)$ be a marked graph, with $V = \Delta \cup \Gamma$ as defined in Chapter 1, Section 1.1. Let the order of G be $v(G) = |V| = p + q$, where $p = |\Gamma|$ and $q = |\Delta|$. A typical element of the joint state space is denoted as $\mathbf{x} = (\mathbf{z}_\Delta, \mathbf{y}_\Gamma)$, where $\mathbf{z}_\Delta = \{z_i\}_{i \in \Delta}$ are categorical values and $\mathbf{y}_\Gamma = \{y_\gamma\}_{\gamma \in \Gamma}$ are real values. Let $\mathbf{X} = (\mathbf{Z}_\Delta, \mathbf{Y}_\Gamma)$ be the corresponding vector of qualitative and quantitative random variables, respectively. Denote the complete state space as $\mathcal{X} = \mathcal{Z}_\Delta \times \mathcal{Y}_\Gamma \subseteq \mathcal{Z}_\Delta \times \mathbb{R}^p$. Then \mathbf{X} follows a *conditional Gaussian* (CG) distribution (Lauritzen and Wermuth, 1989), with strictly positive joint density (w.r.t. product of counting measures on \mathcal{Z}_Δ and Lebesgue measure on \mathcal{Y}_Γ), if

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega) &= f_{\mathbf{Z}_\Delta, \mathbf{Y}_\Gamma}(\mathbf{z}_\Delta, \mathbf{y}_\Gamma; g, h, \Omega) \\ &= \exp \left\{ g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)^T \mathbf{y}_\Gamma - \frac{1}{2} \mathbf{y}_\Gamma^T \Omega(\mathbf{z}_\Delta) \mathbf{y}_\Gamma \right\}, \end{aligned} \quad (2.1)$$

where g is a real-valued function of \mathbf{z}_Δ , h is a p -vector function depending on \mathbf{z}_Δ taking values in \mathbb{R}^p , and $\Omega(\mathbf{z}_\Delta) = [\omega^{\xi\gamma}(\mathbf{z}_\Delta)]$ is a $p \times p$ matrix function depending on \mathbf{z}_Δ taking values in the set of positive definite symmetric matrices. The name *conditional Gaussian* emphasizes the fact that the conditional random vector $\mathbf{Y}_\Gamma \mid \mathbf{Z}_\Delta = \mathbf{z}_\Delta$ is distributed as a p -variate Gaussian random variable with (conditional) expected value equal to $\Omega(\mathbf{z}_\Delta)^{-1} h(\mathbf{z}_\Delta)$, and (conditional) covariance matrix $\Omega(\mathbf{z}_\Delta)^{-1}$. Indeed, on writing

1. $\Omega(\mathbf{z}_\Delta)^{-1} = \Sigma(\mathbf{z}_\Delta)$,
2. $h(\mathbf{z}_\Delta)^T = \mu(\mathbf{z}_\Delta)^T \Omega(\mathbf{z}_\Delta)$,
3. $g(\mathbf{z}_\Delta) = \log p(\mathbf{z}_\Delta) + \frac{p}{2\pi} \log(2\pi) - \frac{1}{2} \log(|\Omega(\mathbf{z}_\Delta)|) - \frac{1}{2} \mu(\mathbf{z}_\Delta)^T \Omega(\mathbf{z}_\Delta) \mu(\mathbf{z}_\Delta)$,

where (and if and only if) $p(\mathbf{z}_\Delta) = \Pr(Z_\Delta = \mathbf{z}_\Delta) > 0$, we have

$$\mathbf{Y}_\Gamma \mid \mathbf{Z}_\Delta = \mathbf{z}_\Delta \sim N_p(\mu(\mathbf{z}_\Delta), \Sigma(\mathbf{z}_\Delta)).$$

Expanding 2.1, yields

$$\log f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega) = g(\mathbf{z}_{\Delta}) + \sum_{\gamma \in \Gamma} \sum_{\xi \in \Gamma} \mu_{\xi}(\mathbf{z}_{\Delta}) \omega^{\xi\gamma}(\mathbf{z}_{\Delta}) y_{\gamma} - \frac{1}{2} \sum_{\gamma \in \Gamma} \sum_{\xi \in \Gamma} y_{\gamma} y_{\xi} \omega^{\xi\gamma}(\mathbf{z}_{\Delta}) \quad (2.2)$$

It is possible (Lauritzen and Wermuth, 1989) to reparameterize the above given expansion in terms of interaction effects, making interactions terms relative to an arbitrary but fixed value \mathbf{z}_{Δ}^* . One obtains

1. $g(\mathbf{z}_{\Delta}) = \sum_{d \subseteq \Delta} \lambda_d(\mathbf{z}_{\Delta});$
2. $h(\mathbf{z}_{\Delta}) = \sum_{d \subseteq \Delta} \boldsymbol{\eta}_d(\mathbf{z}_{\Delta});$
3. $\Omega(\mathbf{z}_{\Delta}) = \sum_{d \subseteq \Delta} \Phi_d(\mathbf{z}_{\Delta});$

where functions indexed by d only depend on \mathbf{z}_{Δ} through \mathbf{z}_d , so that

$$\log f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega) = \sum_{d \subseteq \Delta} \lambda_d(\mathbf{z}_d) + \sum_{d \subseteq \Delta} \boldsymbol{\eta}_d(\mathbf{z}_d)^T \mathbf{y}_{\Gamma} - \frac{1}{2} \sum_{d \subseteq \Delta} \mathbf{y}_{\Gamma}^T \Phi_d(\mathbf{z}_d) \mathbf{y}_{\Gamma} \quad (2.3)$$

This parameterization is related to the Markov properties of a CG distribution. Indeed Lauritzen and Wermuth (1989) give a theorem providing conditions for the distribution (2.2) to be Markov with respect to a graph $G = (V, E)$ with $V = \Delta \cup \Gamma$. We report here the theorem, taken from the original article.

Theorem 2.1. *Let the $q + p$ random vector $\mathbf{X} = (\mathbf{Z}_{\Delta}, \mathbf{Y}_{\Gamma})$ follow a CG distribution as in (2.3). Then \mathbf{X} is a MRF (equivalently, it satisfies Markov properties with respect to a graph $G = (V, E)$, with $V = \Delta \cup \Gamma$), if and only if*

- $\lambda_d(\mathbf{z}_d) = 0$ if d is not complete in G , and
- $\eta_d^{\gamma}(\mathbf{z}_d) = 0$ (the γ -th element of mixed linear interactions $\boldsymbol{\eta}_d(\mathbf{z}_d)$ is zero) if $d \cup \{\gamma\}$ is not complete in G , and
- $\phi_d^{\gamma\xi}(\mathbf{z}_d) = 0$ (the (γ, ξ) -th element of mixed quadratic interactions $\Phi_d(\mathbf{z}_d)$ is zero) if $d \cup \{\gamma, \xi\}$ is not complete in G .

The previous theorem imposes the factorization condition on the interaction functions λ_d , $\boldsymbol{\eta}_d$, and Φ_d , in that interaction terms only involve variables that are neighbours.

To complete this overview, it is helpful to consider what happens to a CG density when we marginalise or condition upon it. Let rewrite Δ and Γ as a union of disjoint sets. In detail, let $B \subseteq V$ and $A = V \setminus B$, such that $\Delta = \Delta_A \cup \Delta_B$, $\Delta_A \cap \Delta_B = \emptyset$, and, consequently, $\Gamma = \Gamma_A \cup \Gamma_B$, $\Gamma_A \cap \Gamma_B = \emptyset$. Denote as $\mathbf{Z}_{\Delta} = (\mathbf{Z}_{\Delta_A}, \mathbf{Z}_{\Delta_B})$ and

$\mathbf{Y}_\Gamma = (\mathbf{Y}_{\Gamma_A}, \mathbf{Y}_{\Gamma_B})$ the corresponding random variables, and $\mathbf{z}_\Delta = (\mathbf{z}_{\Delta_A}, \mathbf{z}_{\Delta_B})$ and $\mathbf{y}_\Gamma = (\mathbf{y}_{\Gamma_A}, \mathbf{y}_{\Gamma_B})$ their observed values. Lauritzen (1996) provides in a sequence of propositions some restrictions on marginalization and conditioning on CG distributions. We report them, adapting the original notation to our.

Proposition 6.6 (Lauritzen (1996)). *Let $f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega)$ be a CG distribution and $\mathbf{X} = (\mathbf{Z}_{\Delta_A}, \mathbf{Z}_{\Delta_B}, \mathbf{Y}_{\Gamma_A}, \mathbf{Y}_{\Gamma_B})$ as previously specified. Then, for all values of $(\mathbf{z}_{\Delta_A}, \mathbf{y}_{\Gamma_A})$, the conditional density of $(\mathbf{Z}_{\Delta_B}, \mathbf{Y}_{\Gamma_B})$ given $(\mathbf{Z}_{\Delta_A}, \mathbf{Y}_{\Gamma_A}) = (\mathbf{z}_{\Delta_A}, \mathbf{y}_{\Gamma_A})$ is a CG distribution, or **CG regression**. Its canonical characteristics $(g_{B|A}, h_{B|A}, \Omega_{B|A})$ are*

$$\begin{aligned} g_{B|A}(\mathbf{z}_{\Delta_B}) &= g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)_{A\Gamma_A}^T \mathbf{y}_{\Gamma_A} - \frac{1}{2} \mathbf{y}_{\Gamma_A}^T \Omega(\mathbf{z}_\Delta)_{A\Gamma_A} \mathbf{y}_{\Gamma_A} - \ln \kappa(\mathbf{z}_{\Delta_A}, \mathbf{y}_{\Gamma_A}) \\ h_{B|A}(\mathbf{z}_{\Delta_B}) &= h(\mathbf{z}_\Delta)_B - \Omega(\mathbf{z}_\Delta)_{BA\Gamma_A} \mathbf{y}_{\Gamma_A} \\ \Omega_{B|A}(\mathbf{z}_{\Delta_B}) &= \Omega(\mathbf{z}_\Delta)_B, \end{aligned}$$

where $\kappa(\mathbf{z}_{\Delta_A}, \mathbf{y}_{\Gamma_A})$ is a normalizing constant

$$\begin{aligned} \kappa(\mathbf{z}_{\Delta_A}, \mathbf{y}_{\Gamma_A}) &= (2\pi)^{\frac{|B \cap \Gamma|}{2}} \sum_{\mathbf{z}_{\Delta_B} \in \{0,1\}^{|B \cap \Delta|}} \left\{ \exp \left\{ g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)_{A\Gamma_A}^T \mathbf{y}_{\Gamma_A} - \frac{1}{2} \mathbf{y}_{\Gamma_A}^T \Omega(\mathbf{z}_\Delta)_{A\Gamma_A} \mathbf{y}_{\Gamma_A} \right\} \times \right. \\ &\quad \left. \times [\det(\Omega(\mathbf{z}_\Delta)_B)]^{-\frac{1}{2}} \exp \left\{ \frac{1}{2} h_{B|A}(\mathbf{z}_{\Delta_B})^T \Omega(\mathbf{z}_\Delta)_B^{-1} h_{B|A}(\mathbf{z}_{\Delta_B}) \right\} \right\} \end{aligned}$$

Generally speaking, Proposition 6.6 affirms that the act of conditioning on any subset of variables preserves the CG distribution. This is not true if the operation is a marginalisation on any subsets of variables in \mathbf{X} . Indeed, there are only two cases in which the marginal of a CG distribution remains a CG distribution: when B is either a subset of Γ (Proposition 6.1), or a subset of Δ which has not influence on the continuous counterpart (Proposition 6.2).

Proposition 6.1 (Lauritzen (1996)). *If $f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega)$ is a CG distribution and $B \subseteq \Gamma$, implying $A = V \setminus B$ (i.e., $A \subseteq \Delta$ or $A \subseteq \Delta \cup \Gamma \setminus B$), then the marginal of $\mathbf{X}_A = (\mathbf{Z}_\Delta, \mathbf{Y}_{\Gamma_A})$ is a CG density with canonical characteristics (g_A, h_A, Ω_A) given as*

$$\begin{aligned} g_A(\mathbf{z}_\Delta) &= g(\mathbf{z}_\Delta) + \frac{1}{2} \left\{ |B| \ln(2\pi) - \ln \det(\Omega(\mathbf{z}_\Delta)_B) + h(\mathbf{z}_\Delta)_B^T [\Omega(\mathbf{z}_\Delta)_B]^{-1} h(\mathbf{z}_\Delta)_B \right\} \\ h_A(\mathbf{z}_\Delta) &= h(\mathbf{z}_\Delta)_A - \Omega(\mathbf{z}_\Delta)_{AB} [\Omega(\mathbf{z}_\Delta)_B]^{-1} h(\mathbf{z}_\Delta)_B \\ \Omega_A(\mathbf{z}_\Delta) &= \Omega(\mathbf{z}_\Delta)_A - \Omega(\mathbf{z}_\Delta)_{AB} [\Omega(\mathbf{z}_\Delta)_B]^{-1} \Omega(\mathbf{z}_\Delta)_{BA}. \end{aligned}$$

Proposition 6.2 (Lauritzen (1996)). *If $f_{\mathbf{X}}(\mathbf{x}; g, h, \Omega)$ is a CG distribution and B is such that $B \subseteq \Delta$ and $B \perp\!\!\!\perp \Gamma \mid \Delta \setminus B$ (i.e., $A \subseteq \Gamma$ or $A \subseteq \Gamma \cup \Delta \setminus B$), then the marginal of $\mathbf{X}_A = (\mathbf{Z}_{\Delta_A}, \mathbf{Y}_\Gamma)$ is a CG density with canonical characteristics $(g_A, \mathbf{h}_A, \Omega_A)$ not*

depending on the levels of \mathbf{z}_{Δ_B} :

$$g_A(\mathbf{z}_{\Delta_A}) = \ln \sum_{j \in \Delta_A} \exp g(z_j), \quad h_A(\mathbf{z}_{\Delta_A}) = h(\mathbf{z}_{\Delta}), \quad \Omega_A(\mathbf{z}_{\Delta_A}) = \Omega(\mathbf{z}_{\Delta}).$$

For any other types of subsets A and B of V that do not satisfy the conditions imposed in Proposition 6.1 and Proposition 6.2, the CG distribution structure for \mathbf{X} is not preserved. These assumptions will be particularly useful in the next chapters, especially in the determination of the distributions participating in the conditional independence tests, which are necessary to infer the multilayer graph-like structure of the data.

Chapter 3

Multilayer Conditional Gaussian Graphical Models

In this chapter, we present the model specification at the core of our analyses and the rationale behind it. We start by introducing our point of view in Section 3.1, then we formally specify the notation and the models in Section 3.2. We tackle the problem of exponential growth of parameters by putting some constraints on the interaction effects, following the work of Cheng *et al.* (2017), and, by exploiting the Markov properties on multilayer networks, we will derive the relevant information for performing structure learning.

3.1 Rationale

In the wide world of integration of different sources of omics data, one of the main up-to-date questions is reconstructing the relations of highly connected systems in presence of distinct measurements of the same genetic activity for a given set of genes. Indeed, for the same p genes associated to a specific biological task, nowadays we can record their expression levels, i.e. continuous variables, their type and number of mutations, i.e., categorical variables, their RNA transcripts, i.e., counts, and many other types of biometric information, and miscellaneous aspects of the biological tasks that they perform can potentially be explored.

All models so far used in the vast literature of modelling interactions among entities have a major drawback: they do not model dependency among measurements of different nature. In the infant literature on models for heterogeneous data, variables with the same domain are usually modelled together, but different domains are not necessarily connected. This implies that, for example, the RNA-transcripts and the expression levels

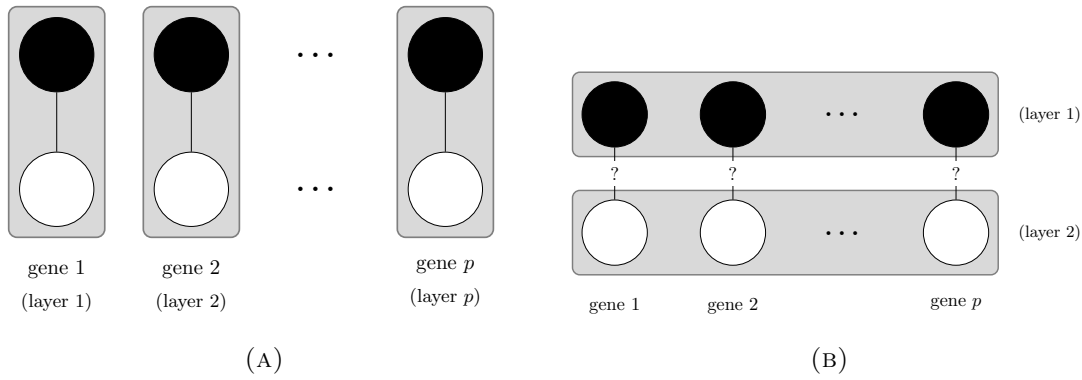


FIGURE 3.1: Interpretation of distinct measurements on the same p genes for a specific activity; panel (A) represents our approach, while panel (B) shows what it is assumed in literature.

of the same gene may not necessarily interact. In the multilayer network language, this translates into a multilayer network with $d = 1$ aspects, p vertices and K layers. Here, K is the number of different domains and each layer contains p homogeneous variables (e.g., a layer for RNA-transcripts, a layer for mutations, etc.) A 2-D simplified representation of a multilayer network of this kind, with $K = 2$, can be seen in panel (B) of Figure 3.1 (the black and white colours of the nodes symbolise their different nature).

From a biological point of view, as all features are recorded during the same genetic activity, they are likely to be implicitly connected and should not be analysed as being disconnected. This is why we reverse the paradigm usually found in the literature and treat each gene as a layer containing all measurements, of various nature, related to it. Within each layer, all measurements are allowed to be connected. In the multilayer network language, this translates into a multilayer network of $d = 1$ aspects, p layers (one for each gene), and K vertices (the features). We assume that all layers share the same set of features, somehow connected among them. Such connections might change from layer to layer, but, in the first instance, we assume that the intra-layer graph remains the same in each layer. In the applications that we tackled, this implied to impose a complete intra-layer graph, but by no means completeness of the intra-layer graph is a necessary condition for what follows. See panel (A) of Figure 3.1 for a 2-D simplified representation of the basic structure of our model (again, the black and white colours of the nodes symbolise their different nature).

By itself, this change of perspective does not entail big differences in the joint distribution of the random variables represented by the node-layer couples in the multilayer network, but it will bring dissimilar results when operations such as conditioning or marginalization will be carried out, as explained in a more formal manner in the next section.

3.2 Model specification

3.2.1 Data format

We assume to have a p -dimensional random vector whose elements are, themselves, K -dimensional random vectors of features or attributes. Such object can be portrayed as a $p \times K$ random matrix \mathbb{Y} of the form

$$\mathbb{Y} = \begin{bmatrix} Y_1^{(1)} & \dots & Y_1^{(k)} & \dots & Y_1^{(K)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_j^{(1)} & \dots & Y_j^{(k)} & \dots & Y_j^{(K)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_p^{(1)} & \dots & Y_p^{(k)} & \dots & Y_p^{(K)} \end{bmatrix}$$

in which each entry corresponds to a random variable $Y_j^{(k)}$ with values in $\mathcal{Y}_j^{(k)}$, $j = 1, \dots, p$; $k = 1, \dots, K$.

There are in fact two ways of looking at \mathbb{Y} : row by row or column by column (see (3.1)). Each row j , $j = 1, \dots, p$, is a random vector $\mathbf{Y}_j^T = (Y_j^{(1)}, \dots, Y_j^{(K)})^T$ with values in $\mathcal{Y}_j = \times_{k=1}^K \mathcal{Y}_j^{(k)}$. We do not impose any constraint on the family of distributions for variables in \mathbf{Y}_j^T , so that it could incorporate variables belonging to different families: continuous, discrete, categorical, and so on. It is worth noting that the sample space \mathcal{Y}_j contains the same ordered tuples for each j , $j = 1, \dots, p$. Similarly, each column k , $k = 1, \dots, K$, is a vector $\mathbf{Y}^{(k)} = (Y_1^{(k)}, \dots, Y_p^{(k)})^T$ with values in $\mathcal{Y}^{(k)} = \times_{j=1}^p \mathcal{Y}_j^{(k)}$. Now $\mathbf{Y}^{(k)}$ contains variables belonging to the same family of distributions, thus having the same domain. On the other side, the sample space $\mathcal{Y}^{(k)}$ potentially changes with k , according to the family to which variables $Y_j^{(k)}$ belong.

$$\mathbb{Y} = \begin{bmatrix} \mathbf{Y}_1^T \\ \vdots \\ \mathbf{Y}_j^T \\ \vdots \\ \mathbf{Y}_p^T \end{bmatrix} = \begin{bmatrix} \mathbf{Y}^{(1)} & \dots & \mathbf{Y}^{(k)} & \dots & \mathbf{Y}^{(K)} \end{bmatrix}. \quad (3.1)$$

A simplified representation of the random matrix \mathbb{Y} is the form of “stacked” vector, i.e., \mathbb{Y} is re-arranged as a $p \times K$ -variate random vector by unfolding by row or by column

the matrix itself:

$$\mathbb{Y} = \left[Y_1^{(1)}, \dots, Y_1^{(K)}, \dots, Y_p^{(1)}, \dots, Y_p^{(K)} \right]^T = \left[Y_1^{(1)}, \dots, Y_p^{(1)}, \dots, Y_1^{(K)}, \dots, Y_p^{(K)} \right]^T.$$

Definition of graphical models on matrices like \mathbb{Y} is challenging. In standard settings, as we explained in the Chapter 1, each random variable corresponds to a node of the graph $G = (V, E)$ underlying the statistical model, such that the vertex set V contains the indices of the random variables. In our setting, however, whether we consider p vectors \mathbf{Y}_j , or K vectors $\mathbf{Y}^{(k)}$, nodes are multidimensional vectors, “hiding”, so to say, more random variables. That is why a more useful way to visualize this new setting is imagining \mathbb{Y} as a multilayer network with $d = 1$ aspects: both rows and columns of \mathbb{Y} can be interpreted as layers, meaning that the single random variables which are elements of \mathbb{Y} are now identified by node-layer couples and not as nodes any more. Following notation introduced in Chapter 1 and denoting the multilayer network on \mathbb{Y} as $M = (V_M, E_M, V, \mathbf{L})$, we can have, depending on whether we are reading \mathbb{Y} by row or by column, respectively,

- $V = \{1, \dots, K\}$ or $V = \{1, \dots, p\}$,
- $\mathbf{L} = \{1, \dots, p\}$ or $\mathbf{L} = \{1, \dots, K\}$,
- $V_M = \{(j, k) : j \in V \text{ and } k \in \mathbf{L}\}$ or $V_M = \{(k, j) : k \in V \text{ and } j \in \mathbf{L}\}$ and
- $E_M \subseteq V_M \times V_M$ such that $E_M = E_A \cup E_C$, $E_A \cap E_C = \emptyset$, with E_A the set of intra-layer edges and E_C the set of inter-layer edges.

In other words, the double perspective on matrix \mathbb{Y} translates into a double perspective on M . When we interpret the matrix as $\mathbb{Y} = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_j^T, \dots, \mathbf{Y}_p^T]^T$, we are facing p layers (p genes) containing the same set of K interconnected node-layer couples (K measurements), and we are supposing that inter-layer edges connect (possible) different node-layer couples, i.e., $E_C = \{((u, \alpha), (v, \beta)) : u, v \in V, \alpha \neq \beta, \alpha, \beta \in \mathbf{L}\}$, with $V = \{1, \dots, K\}$ and $\mathbf{L} = \{1, \dots, p\}$. On the contrary, when we read \mathbb{Y} as $[\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(k)}, \dots, \mathbf{Y}^{(K)}]^T$, we are using state-of-art models, where M is formed by K internally homogeneous layers, in the sense that the related random variables in the arbitrary $\mathbf{Y}^{(k)}$ have all the same domain and such that each of them is containing p node-layer couples, possible different intra-layer edges in each of them, and inter-layer edges of the form $E_C = \{((j, k), (j, h)) : j \in V, h, k \in \mathbf{L} \text{ and } k \neq h\}$, with $V = \{1, \dots, p\}$ and $\mathbf{L} = \{1, \dots, K\}$. Table 3.1 includes a graphical representation of what we described so far.

Random matrix	Multilayer network (by row)	Multilayer network (by column)
$\mathbb{Y} = \begin{bmatrix} Y_1^{(1)} & \dots & Y_1^{(k)} & \dots & Y_1^{(K)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_j^{(1)} & \dots & Y_j^{(k)} & \dots & Y_j^{(K)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_p^{(1)} & \dots & Y_p^{(k)} & \dots & Y_p^{(K)} \end{bmatrix}$		

TABLE 3.1: Illustration of equivalent mathematical structures discussed in the data format section: the random matrix \mathbb{Y} , its representation by multilayer network if it is read by rows (nodes in $\{1, \dots, K\}$ and layers in $\{1, \dots, p\}$) and its counterpart by columns (nodes in $\{1, \dots, p\}$ and layers in $\{1, \dots, K\}$).

3.2.2 The model

We put ourselves in the simplest possible situation: we will assume $K = 2$, i.e., each node is a two-dimensional random vector, and we will impose a complete intra-layer graph. Moreover, we will assume that each layer contains a continuous and a categorical random variable. In our setting, \mathbb{Y} becomes $\mathbb{Y} = [Y_1^{(1)}, \dots, Y_p^{(1)}, Y_1^{(2)}, \dots, Y_p^{(2)}]^T$, where the superscripts (1) and (2) denote the categorical and continuous variables, respectively. The vector takes values in the product space

$$\mathbf{y} = \mathbf{y}^{(1)} \times \mathbf{y}^{(2)}.$$

Since $\mathcal{Y}_i^{(1)}$ is equal for each $i \in \{1, \dots, p\}$, then $\mathbf{y}^{(1)} = (\mathcal{Y}^{(1)})^p$; similarly, $\mathbf{y}^{(2)} = \mathbb{R}^p$. Moreover, every categorical random variable $Y_i^{(1)}$, $i \in \{1, \dots, p\}$, takes at most r values, i.e., $\mathcal{Y}^{(1)} = \{1, \dots, r\}$, so that $\mathbf{y}^{(1)} = \{1, \dots, r\}^p$.

Because we are working with only two types of variables, categorical and continuous, we can drop the notation in $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, and resort to a different one, i.e., we can denote the two vectors as \mathbf{Z}_Δ and \mathbf{Y}_Γ , respectively, as in Chapter 2 for CG models. In this notation, $\Delta = \{1, \dots, q\}$ is the index set of the dummy variables coding the categorical variables, with q being a multiple of p (i.e., $q = r \times p$), and $\Gamma = \{1, \dots, p\}$ is the index set of the continuous variables, so that $V_M = \Delta \times \Gamma$ and $V = \Delta \cup \Gamma$. Here, $|V| = q + p$, where $q = |\Delta|$ and $p = |\Gamma|$.

The random vector $\mathbb{Y} = (\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})$ might therefore also be written as $\mathbb{Y} = (\mathbf{Z}_\Delta, \mathbf{Y}_\Gamma)$. Similarly, the two sample spaces corresponding to $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ will be denoted as \mathcal{Z}_Δ and \mathcal{Y}_Γ , respectively. Typical observed values of \mathbb{Y} will be denoted as $\mathbf{y} = (\mathbf{z}_\Delta, \mathbf{y}_\Gamma)$, where \mathbf{z}_Δ are the observed values of \mathbf{Z}_Δ , and \mathbf{y}_Γ are the observed values of \mathbf{Y}_Γ .

It is reasonable to assume for the vector \mathbb{Y} a joint CG distribution (Lauritzen and Wermuth, 1989), as described in Chapter 2, Section 2.2

$$f_{\mathbb{Y}}(\mathbf{y}; g, h, \Omega) = f_{\mathbf{Z}_\Delta, \mathbf{Y}_\Gamma}(\mathbf{z}_\Delta, \mathbf{y}_\Gamma; g, h, \Omega) = \exp \left\{ g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)^T \mathbf{y}_\Gamma - \frac{1}{2} \mathbf{y}_\Gamma^T \Omega(\mathbf{z}_\Delta) \mathbf{y}_\Gamma \right\},$$

or, equivalently

$$\log f_{\mathbb{Y}}(\mathbf{y}; g, h, \Omega) = \sum_{d \subseteq \Delta} \lambda_d(\mathbf{z}_d) + \sum_{d \subseteq \Delta} \boldsymbol{\eta}_d(\mathbf{z}_d)^T \mathbf{y}_\Gamma - \frac{1}{2} \sum_{d \subseteq \Delta} \mathbf{y}_\Gamma^T \boldsymbol{\Phi}_d(\mathbf{z}_d) \mathbf{y}_\Gamma.$$

As the number of interaction parameters scales exponentially with the number of categorical variables, thus with the number of p genes we are trying to model, the full mixed model is very complex and cannot be easily estimated from data without some additional assumptions. In what follows, therefore, we will limit ourselves to a simplified version of the CG distribution, in which the number of parameters is greatly reduced, and adopt the constraints also proposed in Cheng *et al.* (2017). Therefore, (i) we impose a maximum order of interaction effects (e.g., interactions like $Z_i Z_j Z_k$ are not considered), and (ii) we consider $\lambda_d(\mathbf{z}_\Delta)$, $\boldsymbol{\eta}_d(\mathbf{z}_\Delta)$ and $\boldsymbol{\Phi}_d(\mathbf{z}_\Delta)$ as linear functions of $\boldsymbol{\lambda}$, $\boldsymbol{\eta}$ and $\boldsymbol{\Phi}$, i.e.,

$$\begin{aligned} g(\mathbf{z}_\Delta) &= \sum_{\substack{d \subseteq \Delta \\ |d| \leq 2}} \lambda_d(\mathbf{z}_\Delta) = \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j; \\ h(\mathbf{z}_\Delta) &= \sum_{\substack{d \subseteq \Delta \\ |d| \leq 1}} \boldsymbol{\eta}_d(\mathbf{z}_\Delta) = \boldsymbol{\eta}_0 + \sum_{i \in \Delta} \boldsymbol{\eta}_i z_i = \begin{bmatrix} \eta_0^1 + \sum_{i \in \Delta} \eta_i^1 z_i \\ \vdots \\ \eta_0^p + \sum_{i \in \Delta} \eta_i^p z_i \end{bmatrix}; \\ \Omega(\mathbf{z}_\Delta) &= \sum_{\substack{d \subseteq \Delta \\ |d| \leq 1}} \boldsymbol{\Phi}_d(\mathbf{z}_\Delta) = \boldsymbol{\Phi}_0 + \sum_{i \in \Delta} \boldsymbol{\Phi}_i z_i = \begin{bmatrix} \phi_0^{11} + \sum_{i \in \Delta} \phi_i^{11} z_i & \cdots & \phi_0^{1p} + \sum_{i \in \Delta} \phi_i^{1p} z_i \\ \vdots & \ddots & \vdots \\ \phi_0^{p1} + \sum_{i \in \Delta} \phi_i^{p1} z_i & \cdots & \phi_0^{pp} + \sum_{i \in \Delta} \phi_i^{pp} z_i \end{bmatrix} \\ &= \begin{bmatrix} \phi_0^{11} & \phi_0^{12} + \sum_{i \in \Delta} \phi_i^{12} z_i & \cdots & \phi_0^{1p} + \sum_{i \in \Delta} \phi_i^{1p} z_i \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0^{p1} + \sum_{i \in \Delta} \phi_i^{p1} z_i & \cdots & \phi_0^{pp-1} + \sum_{i \in \Delta} \phi_i^{pp-1} z_i & \phi_0^{pp} \end{bmatrix}. \end{aligned}$$

In this specification, $\phi_i^{\gamma\gamma} = 0$ for all $i \in \Delta$ and $\gamma \in \Gamma$, and λ_0 is the normalizing constant (see also Cheng *et al.* (2017))

$$\lambda_0^{-1} = (2\pi)^{\frac{p}{2}} \sum_{\mathbf{z}_\Delta \in \{0,1\}^q} \frac{1}{\sqrt{\det(\Omega(\mathbf{z}_\Delta))}} \times \exp \left\{ \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i,j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \frac{1}{2} h(\mathbf{z}_\Delta)^T (\Omega(\mathbf{z}_\Delta))^{-1} h(\mathbf{z}_\Delta) \right\}.$$

Bearing in mind that $\Omega(\mathbf{z}_\Delta)$, being a concentration matrix, it is symmetric, it follows that

$$\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i = \phi_0^{\xi\gamma} + \sum_{i \in \Delta} \phi_i^{\xi\gamma} z_i \Leftrightarrow \left(\phi_0^{\gamma\xi} - \phi_0^{\xi\gamma} \right) + \sum_{i \in \Delta} \left(\phi_i^{\gamma\xi} - \phi_i^{\xi\gamma} \right) z_i = 0$$

for all $\gamma, \xi \in \Gamma$, $\gamma \neq \xi$. But this is equivalent to

$$\phi_0^{\gamma\xi} = \phi_0^{\xi\gamma} \wedge \phi_i^{\gamma\xi} = \phi_i^{\xi\gamma}$$

for all $i \in \Delta$, $\gamma, \xi \in \Gamma$, and $\gamma \neq \xi$.

Substituting in the expression of the CG density, we obtain

$$\begin{aligned} \log f_{\mathbb{Y}}(\mathbf{y}; g, h, \Omega) &= g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)^T \mathbf{y}_\Gamma - \frac{1}{2} \mathbf{y}_\Gamma^T \Omega(\mathbf{z}_\Delta) \mathbf{y}_\Gamma \\ &= \sum_{\substack{d \subset \Delta \\ |d| \leq 2}} \lambda_d(\mathbf{z}_d) + \sum_{\substack{d \subset \Delta \\ |d| \leq 1}} \eta_d(\mathbf{z}_d)^T \mathbf{y}_\Gamma - \frac{1}{2} \sum_{\substack{d \subset \Delta \\ |d| \leq 1}} \mathbf{y}_\Gamma^T \Phi_d(\mathbf{z}_d) \mathbf{y}_\Gamma \\ &= \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i,j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i \right) y_\gamma \\ &\quad - \frac{1}{2} \sum_{\gamma \in \Gamma} \phi_0^{\gamma\gamma} y_\gamma^2 - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\gamma y_\xi \end{aligned} \quad (3.2)$$

where $\{\lambda_i; i = 1, \dots, q\}$ are the main effects of the discrete variables, $\{\lambda_{ij}; i = 1, \dots, q, j = 1, \dots, p\}$ represent interaction parameters between pairs of discrete variables, $\{\eta_\gamma; \gamma = 1, \dots, p\}$ are the main effects of continuous variables, $\{\eta_i^\gamma; \gamma = 1, \dots, p, i = 1, \dots, q\}$ are the mixed linear interactions between one discrete and one continuous variables, $\{\phi^{\gamma\xi}; \gamma, \xi = 1, \dots, p\}$ are the pure quadratic interactions between pairs of continuous variables and $\{\phi_i^{\gamma\xi}; \gamma, \xi = 1, \dots, p, i = 1, \dots, q\}$ are the mixed quadratic interactions between pairs of continuous variables and a discrete variable. This model specification

allows for heteroschedasticity of the conditional Gaussian distributions, a feature that seems particularly suitable in the omics settings.

As stated in Cheng *et al.* (2017), this specification brings a consistent reduction in the number of parameters, that drop from an order of $O(2^{q+p}p^2)$ of the unconstrained specification to an order of $O(\max(q^2, qp^2))$; furthermore, forcing a maximum order of the interactions to two does not preclude any structure for the multilayer graph, provided that higher-order interactions, if any, imply the presence of corresponding pairwise interactions, a situation which is most common in applications.

By an application of Theorem 2.1 (see Chapter 2, Section 2.2), the following conditional independences hold (see also Cheng *et al.* (2017)):

- $Z_i \perp\!\!\!\perp Z_j \mid \mathbb{Y} \setminus \{Z_i, Z_j\}$ iff $\lambda_{ij} = 0$
- $Z_i \perp\!\!\!\perp Y_\gamma \mid \mathbb{Y} \setminus \{Z_i, Y_\gamma\}$ iff $\eta_i^\gamma = 0$ and $\phi_i^{\gamma\xi} = 0 \forall \xi \neq \gamma$
- $Y_\gamma \perp\!\!\!\perp Y_\xi \mid \mathbb{Y} \setminus \{Y_\gamma, Y_\xi\}$ iff $\phi^{\gamma\xi} = 0$ and $\phi_i^{\gamma\xi} = 0 \forall i \in \Delta$.

However, taking into account our multilayer set-up, and our assumptions about the nature and the structure of the layers, the conditional independences of core interest for us are those of type,

$$(Z_i, Y_\gamma) \perp\!\!\!\perp (Z_k, Y_\mu) \mid \mathbb{Y} \setminus \{Z_i, Z_k, Y_\gamma, Y_\mu\}. \quad (3.3)$$

Therefore, the crucial conditions in the development of structure learning algorithms are those involving the parameters that make such conditional independences to be true. Denote with

$$(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \left(\lambda_0, \{\lambda_i\}_{i \in \Delta}, \{\lambda_{ij}\}_{\substack{i, j \in \Delta \\ i > j}}, \{\eta_0^\gamma\}_{\gamma \in \Gamma}, \{\eta_i^\gamma\}_{\substack{i \in \Delta \\ \gamma \in \Gamma}}, \{\phi_0^{\gamma\xi}\}_{\gamma, \xi \in \Gamma}, \{\phi_i^{\gamma\xi}\}_{\substack{i \in \Delta \\ \gamma, \xi \in \Gamma}} \right).$$

To exploit such conditions, let $B = \{i, \gamma\}$ and $A = V \setminus \{i, \gamma\}$. From Chapter 2, Section 2.2, the conditional distribution

$$f_{\mathbb{Y}_B | \mathbb{Y}_A}(\mathbf{y}_B \mid \mathbf{y}_A; g_{B|A}, h_{B|A}, \Omega_{B|A}) = f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma \mid \mathbb{Y} \setminus \{z_i, y_\gamma\}; g_{i, \gamma | V \setminus \{i, \gamma\}}, h_{i, \gamma | V \setminus \{i, \gamma\}}, \Omega_{i, \gamma | V \setminus \{i, \gamma\}})$$

represents a CG regression, with canonical characteristics expressed in terms of $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$ as follows (see Appendix B for details):

$$\begin{aligned} \Omega_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) &= \phi_0^{\gamma\gamma}, \\ h_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) &= \eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu, \end{aligned}$$

$$g_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) = \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \lambda_0 - \tilde{\lambda}_0,$$

with $\tilde{\lambda}_0$ appropriately defined.

On writing

$$\begin{aligned} \ln f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | \mathbb{Y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &\propto \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu + \\ &- z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + y_\gamma \left(\eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2, \end{aligned}$$

then we can isolate the couple (z_k, y_μ) with respect to the other variables in the various summations, that is

$$\begin{aligned} \ln f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | \mathbb{Y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &\propto \lambda_i z_i + \lambda_{ik} z_i z_k + z_i \sum_{\substack{j \in \Delta \setminus \{i, k\} \\ i > j}} \lambda_{ij} z_j + \\ &+ \eta_i^\mu z_i y_\mu + z_i \sum_{\xi \in \Gamma \setminus \{\mu, \gamma\}} \eta_i^\xi y_\xi - z_i \sum_{\substack{\xi \in \Gamma \setminus \{\mu, \gamma\} \\ \mu > \xi}} \phi_i^{\mu\xi} y_\mu y_\xi - z_i \sum_{\substack{\zeta, \xi \in \Gamma \setminus \{\mu, \gamma\} \\ \zeta > \xi}} \phi_i^{\zeta\xi} y_\zeta y_\xi + \eta_0^\gamma y_\gamma + \eta_k^\gamma z_k y_\gamma + \\ &+ y_\gamma \sum_{j \in \Delta \setminus \{k\}} \eta_j^\gamma z_j - \phi_0^{\gamma\mu} y_\gamma y_\mu - y_\gamma \sum_{\substack{\xi \in \Gamma \setminus \{\mu, \gamma\} \\ \gamma > \xi}} \phi_0^{\gamma\xi} y_\xi - \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j y_\gamma y_\mu - y_\gamma \sum_{\substack{\xi \in \Gamma \setminus \{\mu, \gamma\} \\ \mu > \xi}} \phi_k^{\gamma\xi} z_k y_\xi + \\ &- y_\gamma \sum_{\substack{\xi \in \Gamma \setminus \{\mu, \gamma\} \\ \mu > \xi}} \sum_{j \in \Delta \setminus \{k\}} \phi_j^{\gamma\xi} z_j y_\xi - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2. \end{aligned}$$

It is then easy to state the conditions that guarantee the conditional independence of our interest to hold: for the discrete interactions, $\lambda_{ik} = 0$; for continuous interactions, $\phi_0^{\gamma\mu} = 0$ and $\phi_i^{\gamma\mu} = 0$ for all $i \in \Delta$; for “mixed” interactions, $\eta_i^\mu = 0$, $\eta_k^\gamma = 0$, $\phi_i^{\mu\xi} = 0$ for all $\xi \in \Gamma \setminus \{\gamma, \mu\}$ and $\phi_k^{\gamma\xi} = 0$ for all $\xi \in \Gamma \setminus \{\gamma, \mu\}$. Those are indeed similar to the conditions described in the paper of Cheng *et al.* (2017), meaning that the Markov properties stating the conditional independence conditions for bivariate conditional distributions can be expressed by means of Markov properties for conditional independence conditions for univariate conditional distributions, that must be simultaneously satisfied. Equivalently, to check if

$$(Z_i, Y_\gamma) \perp\!\!\!\perp (Z_k, Y_\mu) \mid \mathbb{Y} \setminus \{Z_i, Z_k, Y_\gamma, Y_\mu\},$$

the following requirements on the parameters of univariate conditional distributions must hold:

$$\begin{aligned}
Z_i \perp\!\!\!\perp Z_k \mid \mathbb{Y} \setminus \{Z_i, Z_k\} &\Leftrightarrow \lambda_{ik} = 0, \\
Z_i \perp\!\!\!\perp Y_\mu \mid \mathbb{Y} \setminus \{Z_i, Y_\mu\} &\Leftrightarrow \eta_i^\mu = 0 \text{ and } \phi_i^{\mu\xi} = 0 \forall \xi \neq \mu, \\
Z_k \perp\!\!\!\perp Y_\gamma \mid \mathbb{Y} \setminus \{Z_k, Y_\gamma\} &\Leftrightarrow \eta_k^\gamma = 0 \text{ and } \phi_k^{\gamma\xi} = 0 \forall \xi \neq \gamma, \\
Y_\gamma \perp\!\!\!\perp Y_\mu \mid \mathbb{Y} \setminus \{Y_\gamma, Y_\mu\} &\Leftrightarrow \phi_0^{\gamma\mu} = 0 \text{ and } \phi_i^{\gamma\mu} = 0 \forall i \in \Delta.
\end{aligned}$$

Those conclusions on the parameters of the bivariate conditional distributions will be essential in the next Chapter, because they represent the batteries of structure learning tests necessary to infer the multilayer graph underlying the CG model.

Chapter 4

Structure learning of undirected multilayer graphs

In this chapter we will talk about structural learning algorithms for undirected and simple multilayer networks. In particular, since a multilayer network can be seen as a re-labelled graph, as described in Chapter 1, we can adopt algorithms designed for graphs to our framework without loss of generality. Moreover, although structural learning is mostly associated with causal discovery, and we will use methods initially designed for that, in our algorithm we will not search or impose any type of directionality in the inferred graphs. Indeed, many structural learning algorithms deal with directed acyclic graphs, but their first step always requires to infer the undirected skeleton of the graphs: we will stop at this step to derive our structure.

In detail, the Chapter will be organised as follows: in Section 4.1 we will analyse the techniques for structural learning available in the literature for heterogeneous graphical models and their pros and cons, whilst in Section 4.2 we will present our approach. As a formal proof of consistency of the algorithm that we propose is outside the scope of this thesis, consistency of our proposal will be empirically explored in Section 4.3.

4.1 Literature review

There are essentially two classical families of methods for estimating (possibly sparse) graphical models: constraint-based (Spirtes *et al.*, 2000) and score-based (Chickering, 2002) methods. In a nutshell, constraint-based algorithms identify the (causal) structure by using conditional independence tests (under the assumption of faithfulness). Score-based approaches, instead, adopt a network score to measure the goodness of fit of the

network to the data – possibly applying a sparsity penalty to avoid over-fitting – and try to maximize such score.

The best established algorithms are developed for graphs defined of variables belonging to the same family of distributions. The challenges and difficulties rising from handling mixed types of data concern not only the definition of a joint distribution and related properties, as we saw in Chapter 2, but also the development of appropriate learning algorithms: constraint-based techniques must use independence tests suitable to ascertain the independence of a mixed set of variables, and score-based methods require an appropriate score function.

In the last eight years, few groups have developed algorithms to learn undirected graphs over heterogeneous variables (Tur and Castelo, 2011; Cheng *et al.*, 2017; Fellinghauer *et al.*, 2013; Lee and Hastie, 2015; Chen *et al.*, 2015; Yang *et al.*, 2014a). The approach of Tur and Castelo starts from a joint CG distribution and tries to estimate the precision matrix, since the latter is explicitly related to the adjacency matrix of the underlying graph; in particular, for every pair of available random variables, a linear measure of association over all marginal distributions of size lesser than the sample size is computed. But their algorithm, since it is calibrated for cases in which the number of variables is far greater than the sample size, has some limitations: discrete random variables are supposed marginally independent, thus their associations are not considered, and it does not take into account heteroscedasticity in the data.

Other proposals depend on node-wise regression models, i.e., local regressions built around the neighbourhood of a node and derived from a variety of distributions of continuous and discrete variables; in particular, of densities belonging to the exponential family of distributions, as in Cheng *et al.* (2017), Chen *et al.* (2015), (Yang *et al.*, 2014b) or Fellinghauer *et al.* (2013). Those models are called elementary block directed Markov random fields (EBDMRFs); the adjective “directed” is related to the orientation of edges between some blocks of nodes (the graph is not completely undirected, but only partially undirected). The parameters of these local regressions are directly related to the presence and absence of edges in a graph. In order to ensure sparsity in the inferred structure, the Authors use either a lasso penalty (Cheng *et al.*, 2017; Chen *et al.*, 2015; Yang *et al.*, 2014b) or local rankings for each regression, where “local” means that it is possible to rank the importance of predictors for every regression (Fellinghauer *et al.*, 2013).

One of the most popular methods for learning undirected graphical models is the one proposed by Lee and Hastie (Lee and Hastie, 2015): the Authors implements a special case of CG distributions, imposing a common covariance matrix (i.e., discrete random

variables do not have an impact on the covariance structure of the continuous random variables), an additive form for the mean of the distribution and a marginal distribution of the discrete random variables that factorizes as a pairwise discrete Markov random field. Its popularity stands in its computation and simplicity. Indeed, its computation is based on the pseudo-likelihood approach of Besag (1975), which is a computationally efficient and consistent estimator for a joint distribution, determined by products of all conditional distributions of the variables of interest; moreover, the conditional distributions of the model proposed by Lee and Hastie are two widely adopted and well understood models: a multiclass logistic regression and a Gaussian linear regression. Also, in this methodology, the conditional independence structure is captured by the conditional distributions via the regression coefficient of a variable on all others, as in the models of Cheng *et al.* (2017), Chen *et al.* (2015), and (Yang *et al.*, 2014b). The Authors maximize the pseudo-likelihood subject to a weighted penalization on the number of edges to obtain a sparse graphical model on mixed types of data, where the type of penalization is adapted to the type of parameters that are tested in the model (i.e., norm of degree 1 for scalar parameters, degree 2 for vectors of parameters and Frobenius norm for matrices of parameters).

All methods discussed so far are constraint-based. As stated in Sedgewick *et al.* (2017), adapting score-based methods to mixed data is an open problem, because there is the need to find a suitable score, factorizable and computationally efficient. This is why we decided to work within a constraint-based methodology, without searching to deepen and develop methods for score-based algorithms.

4.2 The PC-CGRM algorithm

We propose a new algorithm, called PC-CGRM, designed by exploiting three existing algorithms, i.e., the model parameterization of Cheng *et al.* (2017), the EBDMRF algorithm (Yang *et al.*, 2014b), and the PC algorithm (Spirtes *et al.*, 2000). Briefly, PC-CGRM combines the local approach of Cheng *et al.* (2017) and Yang *et al.* (2014b), i.e., it assumes, as EBDMRF does, that the distribution of each node variable, conditionally to neighbour node variables, belongs to an exponential family; in particular, each node conditional distribution can be expressed as a multinomial regression and a Gaussian linear regression model with the parameters specified as in Cheng *et al.* (2017). Then, the neighbourhood of each node is estimated in turn by a testing procedure on the parameters of the local regressions, following the lines of the PC algorithm. But let us be more specific.

As said at the beginning of the chapter, we will work with the graph-representation of an undirected (and simple) multilayer network. We start from the complete graph $G_M = (V_M, E_M)$, where $V_M \subseteq \Delta \times \Gamma$ by definition in Chapter 1, Section 1.2. We are considering the problem of learning an undirected (possibly sparse) graphical structure under model specification (3.2), which implies that each node-layer conditional distribution follows a CG distribution of the form,

$$f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | \mathbb{y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \propto \exp \left\{ \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu + \right. \\ \left. - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + y_\gamma \left(\eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 \right\}. \quad (4.1)$$

Structure learning can be performed by mean of a series of conditional independence tests. Starting from a complete graph, for each node-layer couple (i, γ) , for each node-layer couple $(k, \mu) \in V_M \setminus \{(i, \gamma)\}$ and for any set of indexes $\mathbf{S} \subseteq V_M \setminus \{(i, \gamma), (k, \mu)\}$, the following null hypotheses

$$H_0 : (Z_i, Y_\gamma) \perp\!\!\!\perp (Z_k, Y_\mu) \mid \mathbb{Y}_{\mathbf{S}},$$

i.e.,

$$H_0 : (Z_i, Y_\gamma) \perp\!\!\!\perp (Z_k, Y_\mu) \mid \mathbb{Y} \setminus \{Z_i, Z_k, Y_\gamma, Y_\mu\}.$$

can be tested to identify the missing edges, corresponding to non rejections of the null hypotheses. Log-likelihood ratio tests can be employed to this aim. Sample estimates $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Phi}})$ can be obtained within a maximum likelihood approach and test statistics can be built by exploiting the asymptotic behaviour of these estimators.

By what was seen at the end of Chapter 3, Section 3.2.2, these null hypotheses can be expressed by a series of simultaneous conditions on the parameters modelling the discrete interactions between (Z_i, Y_γ) and (Z_k, Y_μ) , the mixed interactions between the two couples of random variables and the purely continuous interactions between (Z_i, Y_γ)

and (Z_k, Y_μ) , i.e.,

$$H_0 : \begin{cases} \lambda_{ik} = 0 \\ \{\eta_i^\mu = 0, \phi_i^{\mu\xi} = 0 \forall \xi \neq \mu\} \\ \{\eta_k^\gamma = 0, \phi_k^{\gamma\xi} = 0 \forall \xi \neq \gamma\} \\ \{\phi_0^{\gamma\mu} = 0, \phi_i^{\gamma\mu} = 0 \forall i \in \Delta\} \end{cases}, \quad (4.2)$$

which can be tested by mean of an appropriate series of univariate conditional regressions.

Although we are dealing with models which are more tractable than the ones originally specified by Lauritzen (1996), the inference might be problematic when the number of random variables grows large. To face this problem, we borrow the solution offered by the PC algorithm, which relies on regulating the number of variables that can be part of the conditional sets. The cardinality of \mathbf{S} (i.e., the set of conditional variables) is regulated to remedy the problem of unstable estimations: either its dimension is between 0 and $(p+q) - 4$ (i.e., the total number of variable minus the four not included in \mathbf{S}), or it has an upper bound m , $m \ll (p+q) - 4$, representing the maximum number of neighbours that one node-layer tuple is allowed to have. Furthermore, by the rationale of our model (see Chapter 3, Section 3.1), if the node-layer couples considered belong to the same layer, i.e., we are testing the presence of an edge between (i, γ) and (j, γ) given \mathbf{S} ($(Z_i, Y_\gamma) \perp\!\!\!\perp (Z_j, Y_\gamma) \mid \mathbb{Y}_{\mathbf{S}}$, with $\mathbf{S} \subseteq V_M \setminus \{(i, \gamma), (j, \gamma)\}$), their connection is fixed (indeed, they represent different measurements on the same gene).

In detail, let $\mathbf{K} = V_M \setminus \{(i, \gamma)\}$, such that $\mathbb{Y}_{\mathbf{K}} = \mathbb{Y} \setminus \{Z_i, Y_\gamma\}$ and $\mathbf{y}_{\mathbf{K}} = \mathbf{y} \setminus \{z_i, y_\gamma\}$. Moreover, assume that

$$(Z_i, Y_\gamma) \mid \mathbb{Y}_{\mathbf{K}} = \mathbf{y}_{\mathbf{K}} \sim \text{CGR}(g_{i,\gamma|\mathbf{K}}, h_{i,\gamma|\mathbf{K}}, \Omega_{i,\gamma|\mathbf{K}}), \quad \forall i, \gamma \in V, \quad (4.3)$$

is a CG regression (CGR) distribution with $(g_{i,\gamma|\mathbf{K}}, h_{i,\gamma|\mathbf{K}}, \Omega_{i,\gamma|\mathbf{K}})$ characterized as in Chapter 3, Section 3.2.2:

$$\begin{aligned} \Omega_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{K}}}(z_i) &= \phi_0^{\gamma\gamma}, \\ h_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{K}}}(z_i) &= \eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu, \\ g_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{K}}}(z_i) &= \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \lambda_0 - \tilde{\lambda}_0, \end{aligned}$$

with $\tilde{\lambda}_0$ appropriately defined. Similarly, it is possible to define the generic node-layer conditional distribution under H_0 as

$$(Z_i, Y_\gamma) \mid \mathbb{Y}_{\mathbf{S}} = \mathbf{y}_{\mathbf{S}} \sim \text{CGR} (g_{i,\gamma|\mathbf{S}}, h_{i,\gamma|\mathbf{S}}, \Omega_{i,\gamma|\mathbf{S}}), \quad \forall i, \gamma \in V, \quad (4.4)$$

where $\mathbf{S} \subseteq V_M \setminus \{(i, \gamma), (k, \mu)\}$ and the canonical characteristics of (4.4) can again be expressed as

$$\begin{aligned} \Omega_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{S}}}(z_i) &= \phi_0^{\gamma\gamma}, \\ h_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{S}}}(z_i) &= \eta_0^\gamma + \sum_{\substack{j \in \Delta \setminus \{i, k\} \\ i > j}} \eta_j^\gamma z_j - \sum_{\substack{\xi \in \Gamma \setminus \{\gamma, \mu\} \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{\substack{j \in \Delta \setminus \{i, k\} \\ i > j}} \phi_j^{\gamma\xi} z_j \right) y_\xi, \\ g_{Z_i, Y_\gamma | \mathbb{Y}_{\mathbf{S}}}(z_i) &= \lambda_i z_i + z_i \sum_{\substack{j \in \Delta \setminus \{i, k\} \\ i > j}} \lambda_{ij} z_j + z_i \sum_{\xi \in \Gamma \setminus \{\mu, \gamma\}} \eta_i^\xi y_\xi - z_i \sum_{\substack{\zeta, \xi \in \Gamma \setminus \{\mu, \gamma\} \\ \zeta > \xi}} \phi_i^{\zeta\xi} y_\zeta y_\xi + \lambda_0 - \tilde{\lambda}_0, \end{aligned}$$

with $\tilde{\lambda}_0$ defined in an appropriate way.

Define as

$$\ell(0) = \ln f_{(Z_i, Y_\gamma) | \mathbb{Y}_{\mathbf{S}}} \left((z_i, y_\gamma) \mid \mathbf{y}_{\mathbf{S}}; \hat{g}_{i,\gamma|\mathbf{S}}, \hat{h}_{i,\gamma|\mathbf{S}}, \hat{\Omega}_{i,\gamma|\mathbf{S}} \right)$$

the maximised log-likelihood under H_0 of $(g_{i,\gamma|\mathbf{S}}, h_{i,\gamma|\mathbf{S}}, \Omega_{i,\gamma|\mathbf{S}})$ reparameterized in terms of interaction terms $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$. For convenience, denote as $\theta_{i,\gamma|\mathbf{K}} = (\boldsymbol{\lambda}_{i,\gamma|\mathbf{K}}, \boldsymbol{\eta}_{i,\gamma|\mathbf{K}}, \boldsymbol{\Phi}_{i,\gamma|\mathbf{K}})$, where $(\boldsymbol{\lambda}_{i,\gamma|\mathbf{K}}, \boldsymbol{\eta}_{i,\gamma|\mathbf{K}}, \boldsymbol{\Phi}_{i,\gamma|\mathbf{K}})$ is the parametrization in terms of interaction terms $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$ of the canonical characteristics $(g_{i,\gamma|\mathbf{K}}, h_{i,\gamma|\mathbf{K}}, \Omega_{i,\gamma|\mathbf{K}})$, and writes as $\ell(\theta_{i,\gamma|\mathbf{K}})$ the global maximised log-likelihood of $\theta_{i,\gamma|\mathbf{K}}$,

$$\ell(\theta_{i,\gamma|\mathbf{K}}) = \ln f_{(Z_i, Y_\gamma) | \mathbb{Y}_{\mathbf{K}}} \left((z_i, y_\gamma) \mid \mathbf{y}_{\mathbf{K}}; \hat{g}_{i,\gamma|\mathbf{K}}, \hat{h}_{i,\gamma|\mathbf{K}}, \hat{\Omega}_{i,\gamma|\mathbf{K}} \right).$$

A log-likelihood ratio test statistic for the hypothesis (4.2) can be obtained from standard asymptotic theory, giving

$$W(\theta_{i,\gamma|\mathbf{K}}) = -2 [\ell(\theta_{i,\gamma|\mathbf{K}}) - \ell(0)] \xrightarrow{d} \chi_\nu^2,$$

where $\nu = \dim(\Theta) - \dim(\Theta_0)$, Θ is the unrestricted parameter space of interest for the model under verification, and Θ_0 is the restricted parameter space under the null hypothesis.

By the conditional independence equivalences found in Chapter 3, Section 3.2.2, it is also possible to interpret the null hypothesis formulated in (4.2) as an intersection of

(null) sub-hypotheses, one for each univariate conditional regression that decomposes the bivariate conditional regression (4.1). Equivalently, since under the specification of the model in use,

$$(Z_i, Y_\gamma) \perp\!\!\!\perp (Z_k, Y_\mu) \mid \mathbb{Y} \setminus \{Z_i, Z_k, Y_\gamma, Y_\mu\} \Leftrightarrow \begin{cases} Z_i \perp\!\!\!\perp Z_k \mid \mathbb{Y} \setminus \{Z_i, Z_k\} \\ Z_i \perp\!\!\!\perp Y_\mu \mid \mathbb{Y} \setminus \{Z_i, Y_\mu\} \\ Z_k \perp\!\!\!\perp Y_\gamma \mid \mathbb{Y} \setminus \{Z_k, Y_\gamma\} \\ Y_\gamma \perp\!\!\!\perp Y_\mu \mid \mathbb{Y} \setminus \{Y_\gamma, Y_\mu\} \end{cases}, \quad (4.5)$$

then the global null hypothesis (4.2) can be re-postulated as an intersection of the following four null sub-hypothesis, one for each condition on the right side of the equivalence in (4.5):

$${}_1H_0 : \lambda_{ik} = 0, \quad (4.6)$$

$${}_2H_0 : \eta_i^\mu = 0 \text{ and } \{\phi_i^{\mu\xi} = 0 \forall \xi \neq \mu, \xi \in \Gamma\}, \quad (4.7)$$

$${}_3H_0 : \eta_k^\gamma = 0 \text{ and } \{\phi_k^{\gamma\xi} = 0 \forall \xi \neq \gamma, \xi \in \Gamma\}, \quad (4.8)$$

$${}_4H_0 : \phi_0^{\gamma\mu} = 0 \text{ and } \{\phi_i^{\gamma\mu} = 0 \forall i \in \Delta\}, \quad (4.9)$$

and

$$H_0 : {}_1H_0 \cap {}_2H_0 \cap {}_3H_0 \cap {}_4H_0.$$

For the tests (4.6) – (4.9), particular cases of CG regressions (see also Cheng *et al.* (2017) and Appendix B) are implied: when we tests for ${}_4H_0$, the univariate conditional distributions of $Y_\gamma \mid \mathbb{Y} \setminus \{Y_\gamma\} = \mathbf{y} \setminus \{y_\gamma\}$ and $Y_\mu \mid \mathbb{Y} \setminus \{Y_\mu\} = \mathbf{y} \setminus \{y_\mu\}$ are linear Gaussian regressions; when we test for ${}_1H_0$, the univariate conditional distributions of $Z_i \mid \mathbb{Y} \setminus \{Z_i\} = \mathbf{y} \setminus \{z_i\}$ and $Z_k \mid \mathbb{Y} \setminus \{Z_k\} = \mathbf{y} \setminus \{z_k\}$ are logistic regressions. When we are testing ${}_2H_0$ or ${}_3H_0$, depending on which variable we are interested, we have either a Gaussian linear regression or a logistic one.

In this latter characterization of the null hypothesis, it is explicit to see that we are in front of a multiple testing procedure. To avoid the problems deriving from not taking into account this structure of the testing procedure, we use a false discovery rate (FDR) approach to control the expected proportion of rejected null hypotheses that are false (Benjamini and Hochberg, 1995).

However, the conditional independence tests are prone to mistake. Incorrectly deleting or retaining an edge would result in different neighbourhoods of other node-layer couples, as the graph G_M is updated dynamically. Therefore, the resulting graph is dependent on the order in which the conditional independence tests are executed. To

bypass this issue, we employ the solution in Colombo and Maathuis (2014): the Authors modified the PC algorithm in such a way that it does not depend on the order of the variables, calling it PC-stable. To remove the order-dependence, the neighbours of all node-layer couples are searched for and kept unchanged at each particular cardinality c of the set \mathbf{K} . As a result, an edge deletion at one level does not impact on the conditioning set of the other node-layer couples, from which it follows that the output does not depend on the ordering of the random variables.

The pseudo-code for the algorithm can be written as in Algorithm 1, where

$$\text{adj}(G_M, (i, \gamma)) = \{(k, \mu) \in V_M : ((i, \gamma), (k, \mu)) \in E_M\}$$

denotes the set of all node-layer couples that are adjacent to (i, γ) on the graph G_M .

Algorithm 1 The PC-CGRM algorithm.

- 1: **Input:** n independent realizations of the random vector \mathbb{Y} ; an ordering $order(V_M)$ of the variables, with $V_M \subseteq \Delta \times \Gamma$ set of indexes of variables in \mathbb{Y} (and a stopping level m).
 - 2: **Output:** An estimated undirected (simple) graph \hat{G}_M .
 - 3: Form the complete undirected graph \tilde{G}_M on the vertex set V_M .
 - 4: Keep fixed edges on the same layer γ : $((i, \gamma), (k, \gamma)) \in E_M$.
 - 5: $c \leftarrow -1$; $\hat{G}_M \leftarrow \tilde{G}_M$.
 - 6: **repeat**
 - 7: $c \leftarrow c + 1$
 - 8: **for all** vertices $(i, \gamma) \in V_M$, **do**
 - 9: let $\mathbf{K}_{(i, \gamma)} = \text{adj}(\tilde{G}_M, (i, \gamma))$.
 - 10: **end for**
 - 11: **repeat**
 - 12: select a (new) ordered pair of nodes (i, γ) and (k, μ) that are adjacent in \hat{G}_M such that $|\mathbf{K}_{(i, \gamma)} \setminus \{(k, \mu)\}| \geq c$, using $order(V_M)$.
 - 13: **repeat**
 - 14: choose a (new) set $\mathbf{S} \subseteq \mathbf{K}_{(i, \gamma)} \setminus \{(k, \mu)\}$, with $|\mathbf{S}| = c$, using $order(V_M)$.
 - 15: **if** $H_0 : \theta_{i, \gamma | \mathbf{K}} = 0$ not rejected after FDR correction, **then**
 - 16: delete edge $((i, \gamma), (k, \mu))$ from \hat{G}_M .
 - 17: **end if**
 - 18: **until** edge $((i, \gamma), (k, \mu))$ is deleted or all $\mathbf{S} \subseteq \mathbf{K}_{(i, \gamma)} \setminus \{(k, \mu)\}$, with $|\mathbf{S}| = c$, have been considered.
 - 19: **until** all ordered pair of adjacent couples (i, γ) and (k, μ) such that $|\mathbf{K}_{(i, \gamma)} \setminus \{(k, \mu)\}| \geq c$ and $\mathbf{S} \subseteq \mathbf{K}_{(i, \gamma)} \setminus \{(k, \mu)\}$, with $|\mathbf{S}| = c$, have been tested for conditional independence.
 - 20: **until** $c \leftarrow m$ or, for each ordered pair of adjacent node-layer couples (i, γ) , (k, μ) : $|\text{adj}(\hat{G}_M, (i, \gamma)) \setminus \{(k, \mu)\}| < c$
-

We note that the pseudo-code is almost identical to Algorithm 4.1 in Colombo and Maathuis (2014). Indeed, the differences lie in line 4 of the above pseudo-code (it is not

present in the original PC-stable algorithm), where we specified the fixed edges related to the same layer, and in the statistical procedure used to test the hypothesis at line 15.

4.3 Simulation studies

In this section, we provide the results obtained in two different simulation studies, aimed at empirically evaluate consistency of PC-CGRM.

To generate synthetic data coming from a CG distribution, defined as in equation (3.2), we followed this procedure. Given the true structure G_M , we set all parameters corresponding to absent edges to 0. For the non-zero parameters, we set the parameters of both discrete and linear interaction effects corresponding to the intra-layer edges to 1 or -1 with equal probability, while the parameters of the quadratic interaction effects corresponding to intra-layer edges are set to 0. The non-zero off-diagonal elements of Φ_0 and Φ_i (for all dummy indices i) are defined in such a way that $\Omega(\mathbf{z}_\Delta) = \Phi_0 + \sum_{i \in \Delta} \Phi_i z_i$ is symmetric positive definite for all possible values of \mathbf{z}_Δ . We compute the marginal discrete probabilities for the possible values of \mathbf{z}_Δ and then, for each \mathbf{z}_Δ , we generate the (conditional) continuous component \mathbf{y}_Γ from a multivariate Gaussian distribution with mean vector $\mu(\mathbf{z}_\Delta)$ and variance-covariance matrix $\Sigma(\mathbf{z}_\Delta)$ as defined in points 1. and 2. in Section 2.2.

To measure the ability of PC-CGRM to reconstruct the true graph, the following table has been created at each replication of the simulations,

		G_M	
		<i>Edge present</i>	<i>Edge absent</i>
\hat{G}_M	<i>Edge present</i>	True positive (TP)	False positive (FP)
	<i>Edge absent</i>	False negative (FN)	True negative (TN)

where \hat{G}_M denotes the inferred multilayer network. From the table, three measures have been computed i.e., the true positives (TP), the positive predictive value (PPV) and the sensitivity, where

$$\text{PPV} = \frac{TP}{TP + FP}, \quad \text{Sensitivity} = \frac{TP}{TP + FN}.$$

The overall goodness of the structure learning has been evaluated on the basis of the number of true positive (TP), the positive predictive value (PPV) and sensitivity curves across different combinations of neighbourhood size (m) and nominal significance level for the conditional independence test (α), computed as an average over the values obtained in each of the B replications of the simulations.

Scenario 1. In this scenario, we considered a rather small multilayer network with $p = 3$ layers, pictured in Figure 4.1. In the plot, the node color code specified in Chapter 1, Section 1.1 was adopted. The inter-layer edges can be coloured in three possible colours: (i) black, if the two layers connected by it form a complete graph; (ii) blue, if an entire layer is connected with only the discrete variable of the other layer; (iii) dark orange, if an entire layer is connected with only the continuous variable of the other layer. In panel (B) of Figure 4.1, a more compacted design is proposed, where the nodes are now portraying the layers of the multilayer network in panel (A); this type of visualization will be useful for the real case studies which will be addressed in Chapter 5, and it is the one we use to display the simulation results.

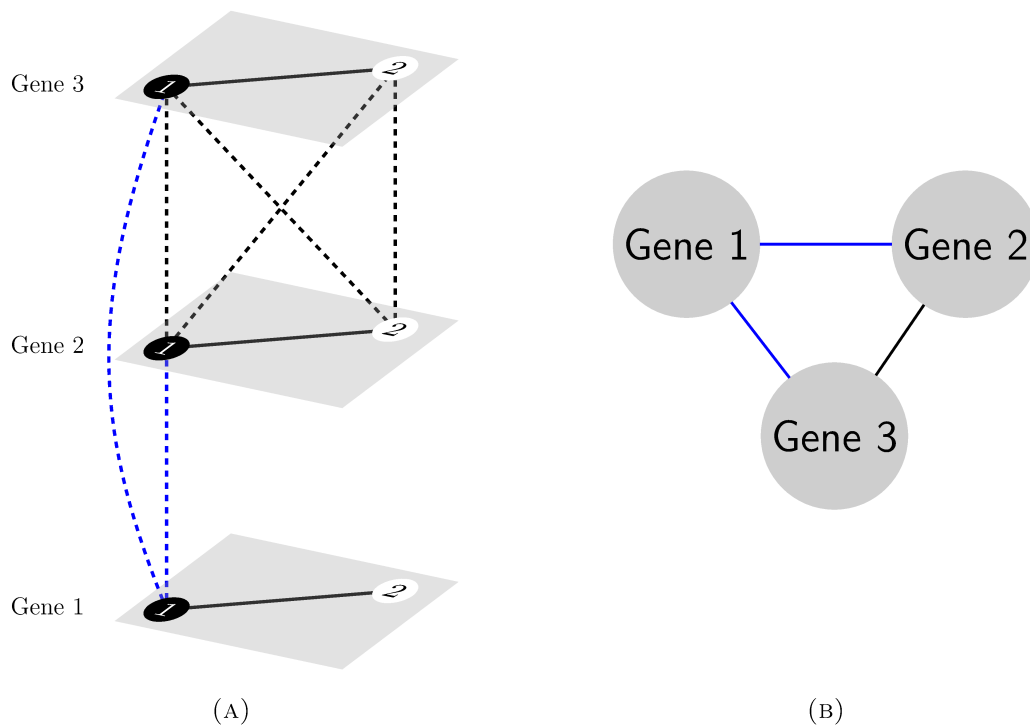


FIGURE 4.1: Multilayer network of scenario 1. Panel (B) shows the network in its entirety; black dots represent categorical variables and white dots represent continuous variables. Panel (B) contains the representation of the same network through its layers.

In this scenario, $B=100$ datasets were sampled for four sample sizes, $n = 100, 200, 500$ and 1000 . The results are summarized in Figure 4.3 for neighbourhood size $m = 1$, Figure 4.4 for neighbourhood size $m = 2$ and Figure 4.5 for $m = 5$; the different coloured curves in all three figures represent the three indices of goodness of fit for four nominal significance levels $\alpha = 0.001, 0.01, 0.05$ and 0.1 .

As the sample size grows, the algorithm converges to the true structure regardless of the neighbourhood size m . Given the sample size, decreasing the nominal levels

of significance causes the algorithm to find fewer edges, but those it finds are actual edges of the graph from which we are generating the data, as supported by the PPV panels of Figure 4.3, Figure 4.4 and Figure 4.5. This means that the only cases of misspecification of presence/absence of an edge are false negatives: indeed, from the sensitivity curves in the third panel of Figure 4.3, Figure 4.4 and Figure 4.5, the PC-CGRM algorithm tends to miss true edges at smaller values of sample size and smaller values of α , unconditionally from m .

Running times of the PC-CGRM algorithm for this scenario are reported in Table 4.1, Table 4.2, Table 4.3 and Table 4.4, one for each chosen sample size. All the times are in seconds. The column “User” displays the CPU time spent by the current process (i.e., PC-CGRM algorithm), the column “System” is the CPU time spent by the operating system (calculation server SGI¹) on behalf of the current process, and “Elapsed” is the wall clock time taken to execute the algorithm, plus some benchmarking code wrapping it (that’s why it is higher than the user time). At most, the algorithm spends almost 839 seconds (~ 14 minutes) to compute 100 inferred multilayer networks (see Table 4.3, case $m = 2$ and $\alpha = 0.001$), while the minimum runtime of the routine is 88 seconds (~ 1 minute and 46 seconds; see Table 4.1, case $m = 1$ and $\alpha = 0.01$).

Scenario 2. In this scenario, we considered a multilayer network with $p = 20$ layers, with a structure resembling that of some networks estimated during the real data analysis. The multilayer view for the second setting is not reported due to the difficulties given by the high number of layers; instead, we visualize it through its layer connections in Figure 4.2, similarly as the graph in panel (B) of Figure 4.1.

We generated $B=500$ datasets of sizes $n = 200, 500, 1000$ and 2000 . The results are summarized in Figure 4.6 for neighbourhood size $m = 1$, Figure 4.7 for neighbourhood size $m = 2$ and Figure 4.8 for neighbourhood size $m = 5$; as before, the different coloured curves represent the three indices of goodness of fit for the four nominal significance levels $\alpha = 0.001, 0.01, 0.05$ and 0.1 .

The situation is similar to what we saw for the first scenario. As the sample size grows, the algorithm converges to the true multilayer graph. This is confirmed by the increasing curves of TP and PPV values, the first and second panel of Figure 4.6, respectively. At the same level of sample size, smaller nominal levels of significance α cause the algorithm to find fewer edges, although these differences tends to decrease when the sample size is large. Higher values of α increase the number of false positives,

¹Calculation server SGI, model UV3000 (Department of Statistical Sciences, University of Padova), 8 CPU Intel®Xeon®E5-4650 v3 at 2.10 GHz, with 12 physical cores per CPU for a total of 192 logical CPU units; home space for a single user is 4 Gb at most.

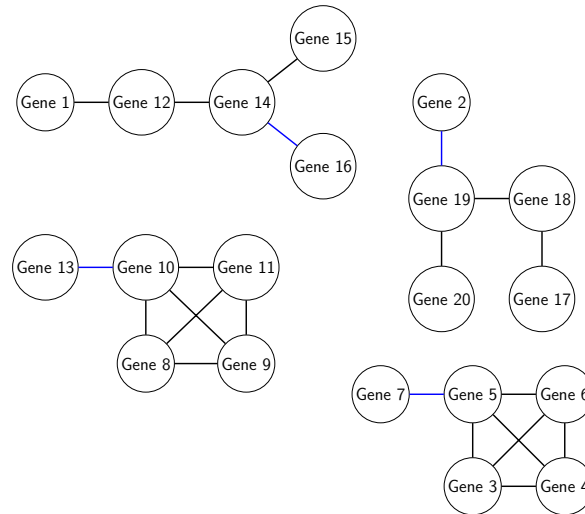


FIGURE 4.2: Compacted visualization of the multilayer network for $p = 20$ layers.

as it is shown in the PPV curves panel, although the number of false positives remains less than 20% of all edges inferred by the algorithm. The number of false negatives tends to zero the higher the sample size is, as the sensitivity panels show; however, increasing the neighbourhood size from $m = 1$ to $m = 2$ with a fixed small level of significance α and a fixed sample size n leads to an higher number of false positives: this is clearly visible from the sensitivity curves in Figure 4.6 and Figure 4.7 for $n = 2000$ and $\alpha = 0.001$ (grey curves). There are no evident differences, instead, when comparing Figure 4.7 and Figure 4.8.

In this scenario, due to an increased number of replications B and higher number of variables $p = 20$, the runtime of the PC-CGRM algorithm increases: the lowest running time is approximately 815 seconds (~ 14 minutes; see Table 4.6, case $m = 2$ and $\alpha = 0.001$ – notice that, in the previous scenario, 14 minutes were the longest runtime), while the highest is 6225 seconds (~ 2 hours and 13 minutes; see Table 4.7, case $m = 5$ and $\alpha = 0.1$).

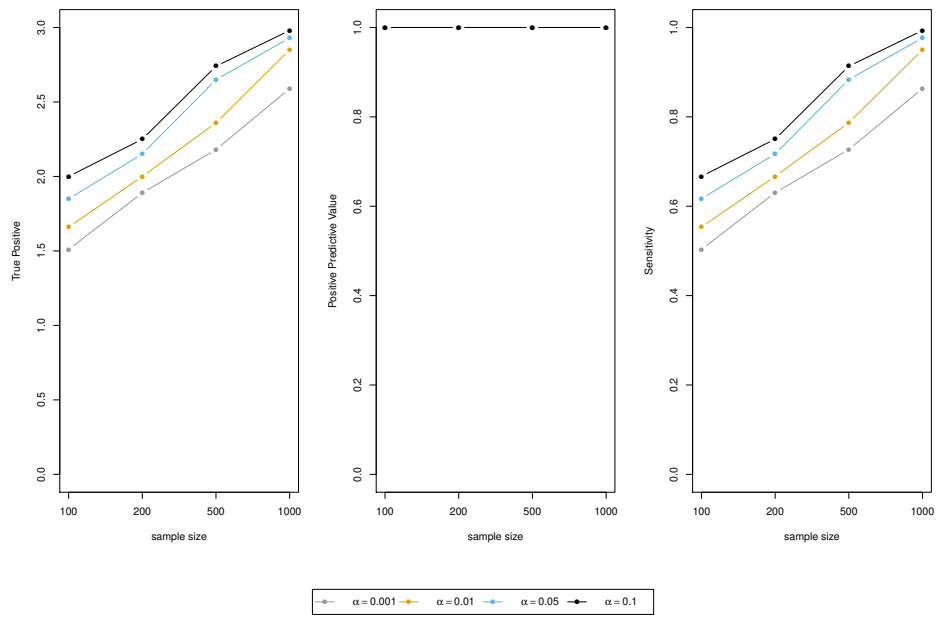


FIGURE 4.3: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 1$.

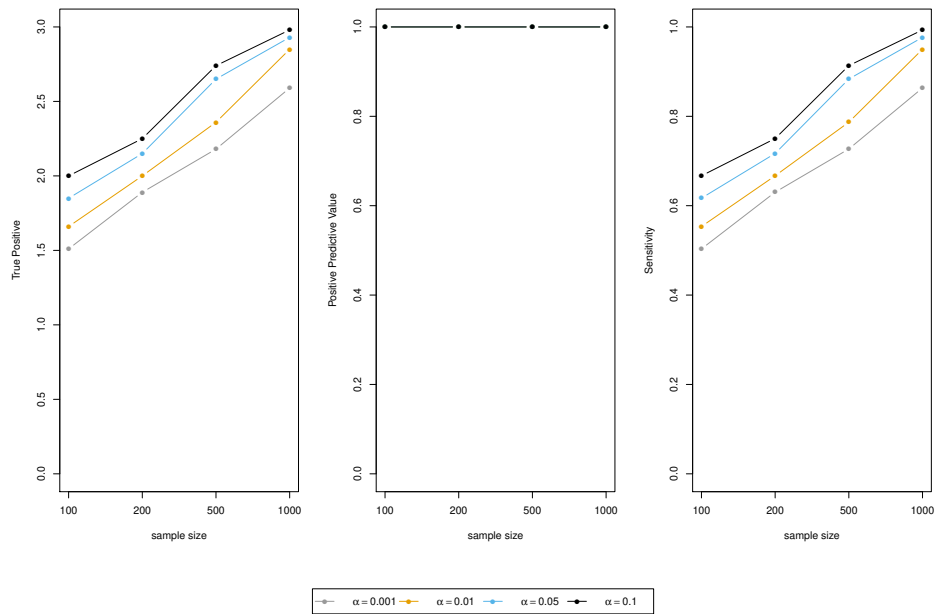


FIGURE 4.4: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 2$.

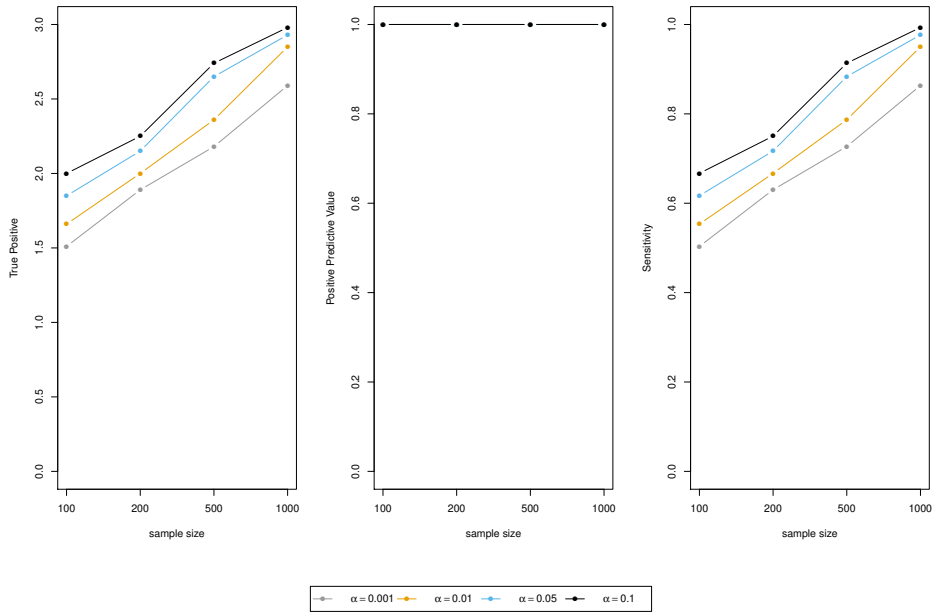


FIGURE 4.5: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the first scenario in case $m = 5$.

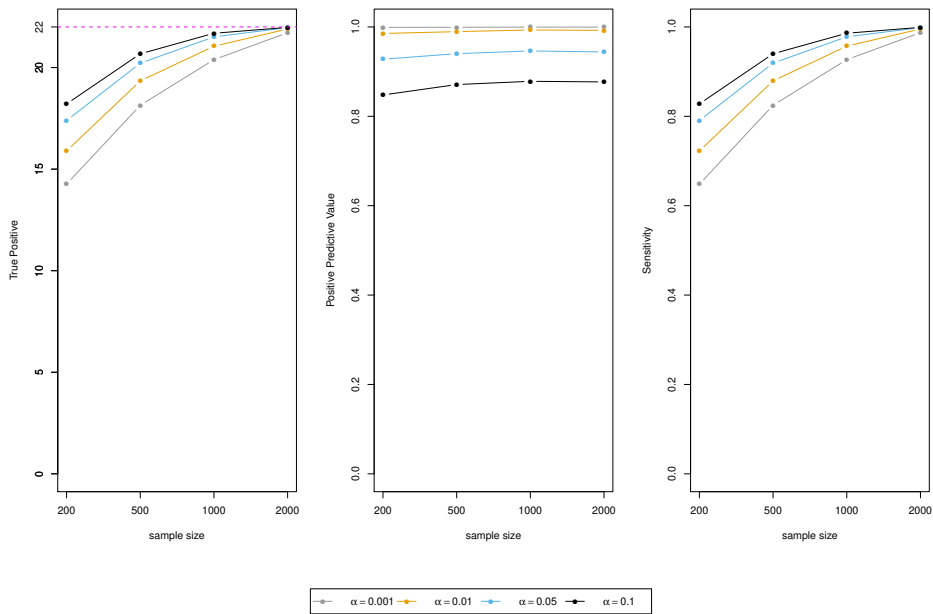


FIGURE 4.6: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 1$.

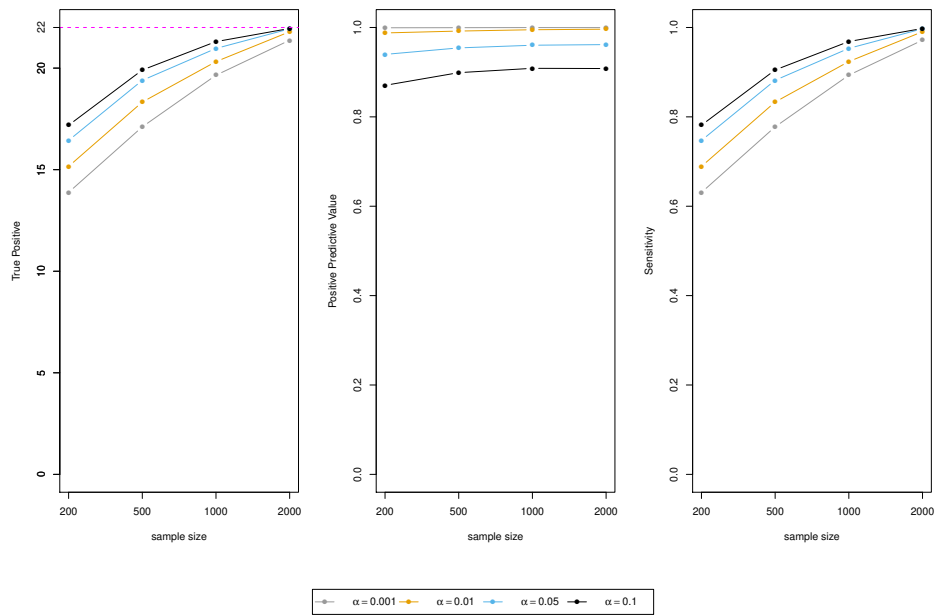


FIGURE 4.7: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 2$.

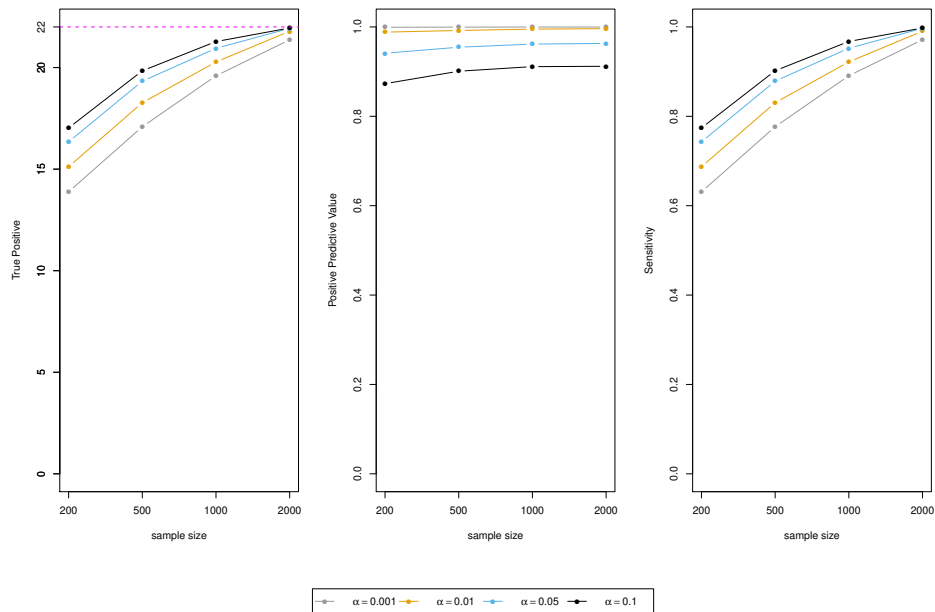


FIGURE 4.8: TP (first panel), PPV (second panel) and sensitivity (third panel) curves for the second scenario in case $m = 5$.

Sample size n	m level	α level	User	System	Elapsed
$n = 100$	$m = 1$	$\alpha = 0.001$	99.206	0.671	154.769
		$\alpha = 0.01$	87.691	0.539	141.43
		$\alpha = 0.05$	89.322	0.59	140.063
		$\alpha = 0.1$	89.15	0.567	148.985
	$m = 2$	$\alpha = 0.001$	124.624	0.829	205.865
		$\alpha = 0.01$	127.641	0.758	216.626
		$\alpha = 0.05$	130.259	0.745	220.967
		$\alpha = 0.1$	439.459	1.706	661.343
	$m = 5$	$\alpha = 0.001$	252.214	1.317	407.466
		$\alpha = 0.01$	258.955	1.225	412.475
		$\alpha = 0.05$	258.604	1.278	388.874
		$\alpha = 0.1$	260.298	1.293	395.615

TABLE 4.1: Runtime of PC-CGRM for the first scenario with sample size $n = 100$.

Sample size n	m level	α level	User	System	Elapsed
$n = 200$	$m = 1$	$\alpha = 0.001$	674.177	1.660	1262.110
		$\alpha = 0.01$	692.942	1.620	1308.248
		$\alpha = 0.05$	717.681	1.542	1497.368
		$\alpha = 0.1$	634.865	1.580	1234.836
	$m = 2$	$\alpha = 0.001$	709.454	2.029	1297.795
		$\alpha = 0.01$	603.329	1.807	983.119
		$\alpha = 0.05$	606.608	2.24	986.811
		$\alpha = 0.1$	578.744	2.009	907.084
	$m = 5$	$\alpha = 0.001$	249.117	1.38	409.091
		$\alpha = 0.01$	259.259	1.251	420.166
		$\alpha = 0.05$	695.887	2.701	990.223
		$\alpha = 0.1$	725.258	2.583	1030.912

TABLE 4.2: Runtime of PC-CGRM for the first scenario with sample size $n = 200$.

Sample size n	m level	α level	User	System	Elapsed
$n = 500$	$m = 1$	$\alpha = 0.001$	530.648	1.689	1043.932
		$\alpha = 0.01$	348.340	1.556	654.363
		$\alpha = 0.05$	335.631	1.364	622.568
		$\alpha = 0.1$	380.450	1.864	758.051
	$m = 2$	$\alpha = 0.001$	838.519	2.474	1604.554
		$\alpha = 0.01$	352.747	1.283	620.384
		$\alpha = 0.05$	283.402	1.45	495.616
		$\alpha = 0.1$	293.467	1.587	496.425
	$m = 5$	$\alpha = 0.001$	257.487	1.418	447.896
		$\alpha = 0.01$	265.877	1.301	452.003
		$\alpha = 0.05$	265.576	1.412	416.932
		$\alpha = 0.1$	267.649	1.388	435.486

TABLE 4.3: Runtime of PC-CGRM for the first scenario with sample size $n = 500$.

Sample size n	m level	α level	User	System	Elapsed
$n = 1000$	$m = 1$	$\alpha = 0.001$	691.994	1.899	1374.859
		$\alpha = 0.01$	720.625	1.796	1449.787
		$\alpha = 0.05$	467.813	1.497	975.132
		$\alpha = 0.1$	410.406	1.421	828.006
	$m = 2$	$\alpha = 0.001$	857.192	2.258	1620.118
		$\alpha = 0.01$	793.357	1.95	1477.179
		$\alpha = 0.05$	711.085	2.221	1287.723
		$\alpha = 0.1$	619.946	2.415	1044.749
	$m = 5$	$\alpha = 0.001$	360.687	1.521	649.935
		$\alpha = 0.01$	378.477	1.484	636.127
		$\alpha = 0.05$	376.537	1.468	626.616
		$\alpha = 0.1$	370.461	1.406	593.43

TABLE 4.4: Runtime of PC-CGRM for the first scenario with sample size $n = 1000$.

Sample size n	m level	α level	User	System	Elapsed
$n = 200$	$m = 1$	$\alpha = 0.001$	1210.121	8.613	6139.524
		$\alpha = 0.01$	1584.453	8.908	8923.999
		$\alpha = 0.05$	1461.194	9.203	10172.78
		$\alpha = 0.1$	1354.268	8.193	11582.809
	$m = 2$	$\alpha = 0.001$	1930.088	10.156	8236.267
		$\alpha = 0.01$	2157.803	10.785	10208.798
		$\alpha = 0.05$	2193.696	13.18	12167.233
		$\alpha = 0.1$	2467.654	15.045	16632.3
	$m = 5$	$\alpha = 0.001$	3738.357	17.791	10548.293
		$\alpha = 0.01$	3800.856	19.514	11642.281
		$\alpha = 0.05$	3612.652	18.33	12960.912
		$\alpha = 0.1$	3589.931	18.808	15342.041

TABLE 4.5: Runtime of PC-CGRM for the second scenario with sample size $n = 200$.

Sample size n	m level	α level	User	System	Elapsed
$n = 500$	$m = 1$	$\alpha = 0.001$	814.879	8.558	6209.553
		$\alpha = 0.01$	966.509	8.159	7118.472
		$\alpha = 0.05$	1103.067	8.258	9580.179
		$\alpha = 0.1$	587.374	7.17	7767.304
	$m = 2$	$\alpha = 0.001$	2275.297	11.772	12569.532
		$\alpha = 0.01$	2241.991	13.09	12732.5
		$\alpha = 0.05$	2180.011	12.908	14671.809
		$\alpha = 0.1$	1967.251	15.712	16460.93
	$m = 5$	$\alpha = 0.001$	3802.763	25.622	13942.087
		$\alpha = 0.01$	4470.317	29.161	16204.63
		$\alpha = 0.05$	4534.307	29.615	19274.855
		$\alpha = 0.1$	5506.27	35.397	26239.631

TABLE 4.6: Runtime of PC-CGRM for the second scenario with sample size $n = 500$.

Sample size n	m level	α level	User	System	Elapsed
$n = 1000$	$m = 1$	$\alpha = 0.001$	956.961	9.976	9489.524
		$\alpha = 0.01$	885.907	8.329	9407.603
		$\alpha = 0.05$	1056.963	10.422	13048.994
		$\alpha = 0.1$	1285.6	11.623	17899.866
	$m = 2$	$\alpha = 0.001$	2019.06	15.375	15512.391
		$\alpha = 0.01$	2745.872	18.136	19604.643
		$\alpha = 0.05$	2612.64	18.265	22269.231
		$\alpha = 0.1$	2188.466	15.401	25081.264
	$m = 5$	$\alpha = 0.001$	6107.822	36.237	24596.756
		$\alpha = 0.01$	4423.099	32.382	19876.933
		$\alpha = 0.05$	5841.274	37.712	28037.871
		$\alpha = 0.1$	6224.754	36.665	36167.65

TABLE 4.7: Runtime of PC-CGRM for the second scenario with sample size $n = 1000$.

Sample size n	m level	α level	User	System	Elapsed
$n = 2000$	$m = 1$	$\alpha = 0.001$	1607.649	12.082	19226.876
		$\alpha = 0.01$	1896.3	14.692	21833.858
		$\alpha = 0.05$	1708.984	15.031	25368.721
		$\alpha = 0.1$	2476.122	16.199	38923.027
	$m = 2$	$\alpha = 0.001$	2563.887	19.889	27034.149
		$\alpha = 0.01$	3117.381	23.32	30791.561
		$\alpha = 0.05$	3164.782	24.56	36937.39
		$\alpha = 0.1$	2758.219	23.165	42565.881
	$m = 5$	$\alpha = 0.001$	5374.867	43.549	32497.898
		$\alpha = 0.01$	4587.224	36.183	30255.56
		$\alpha = 0.05$	5882.468	39.684	40860.518
		$\alpha = 0.1$	6030.108	49.928	51738.957

TABLE 4.8: Runtime of PC-CGRM for the second scenario with sample size $n = 2000$.

Chapter 5

Application to Omics data

In this Chapter, we will apply our approach to a real case study in which different genomic activities are recorded in ovarian cancer cells. The Chapter is organised in a first section, Section 5.1, that illustrates the data at hand, followed by Section 5.2, where we will present our results, with emphasis on some pathways which result influential in ovarian cancer.

5.1 Ovarian cancer data

5.1.1 Why ovarian cancer

The ovaries are two small intra-abdominal organs that, during a woman's reproductive age, cyclically produce the egg cell or ova. Over the years, the ovary gradually decreases its activity until it becomes dormant, determining the advent of the menopause. Ovarian cancer originates usually around this time from the surface of this organ, and can be either benign or malignant. As stated in an informative website page of the European Institute of Oncology (IEO) (Istituto Europeo di Oncologia, 2019), the latter type of cancer can metastasise (i.e., cancer cells move from the area where they formed) to other parts of the female body. Between all malignant ovarian tumours, epithelial ovarian carcinoma accounts for 90% of them: this is due to its intrinsic complexity, since it is not a single disease, but combines different diseases with different biological behaviour. From the IEO's website again, we report

Ovarian cancer is the leading cause of death from gynaecological cancer and it is fifth most common cancer in the female population in developed countries.

Each year, it is estimated that 65,000 cases are diagnosed in Europe, including almost 5,000 in Italy. Despite the relatively low incidence of ovarian cancer, it is burdened by high mortality.

About 5%-10% of women diagnosed with ovarian cancer present a genetic mutation (BRCA1, BRCA2, Lynch syndrome) of hereditary/family nature, which increases the risk of this and other types of cancer. Furthermore, several specific genes involved in ovarian carcinogenesis have been identified in the literature, including the p53 tumor suppressor gene and ERBB2 and PIK3CA oncogenes (Wenham *et al.*, 2002; Katabuchi and Okamura, 2003; Shih and Kurman, 2004; Desai *et al.*, 2014).

The danger of such form of carcinoma is the reason we decided to apply our methodology on data related to it.

5.1.2 Preprocessing

Our study uses the data from the ovarian (OV) project available at the Cancer Genome Atlas (TCGA) website. TCGA is the main public database for cancer genomics since 2006, when the National Cancer Institute and the National Human Genome Research Institute (both part of the National Institutes of Health, U.S. Department of Health and Human Services) combined their efforts to persecute a better understanding of the molecular basis of cancer through the application of genome analysis technologies. Nowadays, it has more than 20000 primary cancer cases (i.e., first-time cases of that disease) and 33 macro-types of tumours, for a total of over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data.

The original OV dataset contains different measurements of genetic activity inside the cells, as expression, methylation status, somatic mutations and copy number variation profiles, for a total of 20801 genes recorded. By expression, it is intended the levels at which a particular set of genes is expressed within a cell, tissue or organism. DNA methylation is a process by which methyl groups (i.e., a chemical compound composed by one carbon atom bonded to three hydrogen atoms) are added to the DNA molecule, changing the latter activity without altering its genomic sequence; in the case the methylation occurs in a gene promoter, it usually represses gene transcription. Somatic mutations are alterations in DNA that occurs after conception in any of the cells of the body except the germ cells (sperm and egg). Copy number variations correspond to a structural process where parts of the genome are duplicated or deleted, which differs between individuals in the human population.

In this thesis, we focused on gene expression levels (GELs) and copy number variations (CNVs). The latter classifies the severity of deletions or additions of DNA material

as “-2”, “-1”, “0”, “+1”, “+2”. As data on GELs and CNVs were not simultaneously available for all 608 patients and for all 20K genes, only genes and patients were retained for which both measurements were available along with the related SYMBOL identifier, resulting in a total of 217 patients and 15587 genes.

A portion of the (huge) table showing the frequency distribution of CNV per gene can be seen in Table 5.1. For each gene, the table shows how patients are distributed over the five classes representing severity of deletions or additions.

<i>Gene ID</i>	“-2”	“-1”	“0”	“+1”	“+2”	Patients
<i>ACAP3</i>	1	83	89	84	10	267
<i>ACTRT2</i>	1	83	89	84	10	267
<i>PRDM16</i>	2	82	89	84	10	267
<i>ARHGEF16</i>	0	84	89	83	11	267
<i>TP53BP1</i>	7	161	85	13	1	267
<i>BRCA2</i>	4	153	73	31	6	267
<i>BRCA1</i>	7	202	40	17	1	267
<i>CSMD1</i>	30	147	58	29	3	267
<i>MYC</i>	0	10	41	94	122	267

TABLE 5.1: Frequency distribution of CNVs in the 267 available patients for 9 out of 15587 genes

The distribution of the frequencies of the five CNV classes over all genes can be explored in Figure 5.1. It is evident that extreme categories “-2” and “+2” are much rarer than the others. The same information can be deduced by Table 5.2.

<i>Class</i>	Min.	1 st Qu.	Median	Mean	3 rd Qu.	Max.	Std.	Genes
“-2”	0	0	1	6.6	2	30	3.24	15587
“-1”	1	44	93	100.6	133	232	57.68	15587
“0”	13	74	95	90.4	113	157	28.24	15587
“+1”	4	40	61	73.0	97	163	36.94	15587
“+2”	0	3	7	29.4	15	122	14.42	15587

TABLE 5.2: Summary indices for each category of CNV values.

On only 5958 genes (38.22% of all genes) the distribution of CNV satisfied the condition of having a frequency count larger than 4 in each class (minimum sample size in each class to obtain stable estimates). This aspect is particularly relevant for following analyses, because estimates are sensitive to sample size. To avoid limiting the analysis to consideration of only 5958 genes, we considered the solution to collapse one or both the extreme classes “-2” and “+2” into the adjacent category. As a first step, we collapsed the classes “-2” and “-1”, being “-2” the least represented category. As a result, 9612

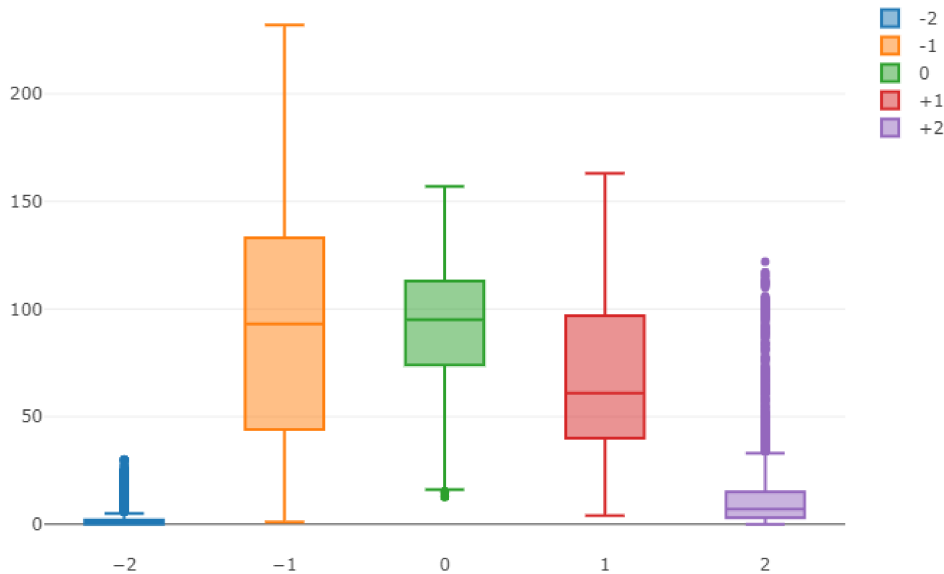


FIGURE 5.1: Boxplots of the absolute frequencies of each CNVs category computed on all 15587 genes.

genes (61.67%), showed at least 4 occurrences on all four classes, a figure that seemed reasonable for the following analyses.

Although the hypothesis of completeness of intra-layer graphs is plausible in our setting (Chapter 3 and Chapter 4), we decided to empirically validate this intuition on this reduced dataset. In other words, we checked if dependency between CNVs and GELs within the same gene (i.e., layer) was reasonable by mean of linear regressions. Overall, at the nominal level 0.05, in about 80% of the layers a significant association was found between these different measurements, providing a convincing evidence in favour of completeness of intra-layer graphs.

Subsequently, we carried out a second step of preprocessing, and selected genes annotated on biological pathways which are deemed important in OV tumour survival. The list of pathways was extracted by a MOSClip analysis (Martini *et al.*, 2019), which is a methodology able to combine survival analysis and graphical model theory to test the survival association of pathways, or of their connected components, in a multi-omics framework. The physical structure of the pathways can be downloaded either from the REACTOME website (Fabregat *et al.*, 2018) or using package `graphite` (Sales *et al.*, 2012, 2018) of Bioconductor repository (Huber *et al.*, 2015).

As not all genes belonging to pathways are usually measured (see Table 5.3), we selected for the following analyses the two biological pathways with the highest number of genes with available measurements, i.e., *Ca2+ pathway* and *Binding and Uptake of Ligands by Scavenger Receptors*.

REACTOME pathway	Annotated genes	Available genes
Ca2+ pathway	55	29
Binding and Uptake of Ligands by Scavenger Receptors	40	22
Constitutive Signaling by AKT1 E17K in Cancer	23	14
Cobalamin (Cbl, vitamin B12) transport and metabolism	20	13
Nucleobase catabolism	28	13
AKT phosphorylates targets in the cytosol	14	11
Interleukin-35 Signalling	12	9
Purine catabolism	16	7
DSCAM interactions	10	5
Erythrocytes take up carbon dioxide and release oxygen	12	4

TABLE 5.3: Number of annotated and available genes for each REACTOME pathway resulting from MOSClip.

These two steps allowed us to set up the two case studies on 267 patients that will be considered in the following Section, one with 29 genes (layers), called *Ca2+*, and one with 22 genes (layers), called *Binding and Uptake of Ligands by Scavenger Receptors*. For these two case studies, we will now present the main results obtained running PC-CGRM for performing structure learning.

5.2 Results

For each of the two cases of interest, we estimated twenty possible graphical models, corresponding to different choices for parameters m and α , where m determines the dimension of the neighbourhood of a node-layer couple, and α is the nominal level of the tests made by PC-CGRM (not to be confused with the symbol denoting a layer, i.e., α , - see Chapter 4). We set $m = 1, 2, 3, 4, 5$ (i.e., a node-layer couple (i, γ) cannot have more than 5 neighbours, where by neighbours we intend node-layer couples adjacent to (i, γ) belonging to layers different from γ), which seemed sensible in a biological pathway of only 29 or 22 layers, while for α we set 0.001, 0.01, 0.05, 0.1. In the following, we will report some of resulting estimated models; the complete display can be found in Appendix C.

The edges are coloured in three possible colours, symbolising the nature of the estimated inter-layer edges, as we already discussed in Chapter 4, Section 4.3. In particular, we remember that two layers are connected by: (i) a black edge, when a complete graph is estimated over the nodes in the layers; (ii) a blue edge, when both an entire layer is connected with only the discrete variable of the other layer, CNV in our case; (iii) a

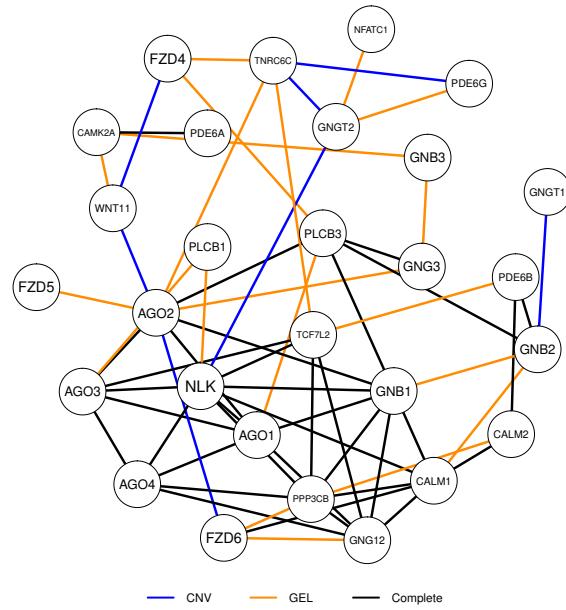
dark orange edge when an entire layer is connected with only the continuous variable of the other layer, GEL in our case. We colour only these cases because we are interested in what type of relation entire layers and single node-layer couples share.

Figure 5.2 shows the estimated multilayer network for the first case study of interest, Ca^{2+} , when $m = 1$ (panel (A)) and $m = 5$ (panel (B)), with a nominal level of 0.001. It is evident that a greater value of the cardinality implies a larger sparsity in the graph, as it was expected from the model, once the nominal level is fixed. Obviously, incrementing α will reduce the sparsity (Appendix C for details).

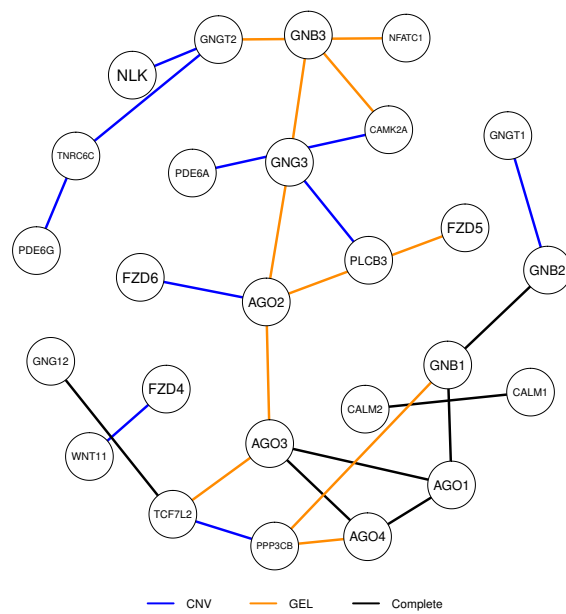
From a biological point of view, interpretation of these results deserves care. Indeed, we are considering only 29 of the original 55 genes belonging to the pathway. Hence, one should not expect to find what it is represented in the original pathway. Nevertheless, it is worth highlighting that some of the connections that are known to be influential in ovarian cancer, such as the WNT / FZD-4 (Sugimura and Li, 2010; Ricken *et al.*, 2002), appear in the inferred multilayer networks. Furthermore, when amplifications / deletions (or other aberrant events) happen between WNT and FZD-4, an increase of intracellular abundance of Ca^{2+} follows, which in turn affects a wide range of cellular processes, including gene expression (Huang *et al.*, 2016). This type of process between WNT and FZD-4 is well captured by the color of the edge connecting the two layers in Figure 5.2: indeed, its blue colour shows a connection between abnormalities in a gene and other types of events in the other gene. Similar conclusions can be drawn for other families of genes, as AGO and NLK ones (Di Leva and Croce, 2013; Zhang *et al.*, 2011).

Similar considerations can be done for the second case study, *Binding and Uptake of Ligands by Scavenger Receptors*. Figure 5.3 shows the inferred multilayer network in the previously considered setting.

At a biological level, most of the genes displayed in both panels of Figure 5.3 play a role in scavenging aberrant molecules which participate in pathological process, the ovarian cancer being one (PrabhuDas *et al.*, 2017; Li *et al.*, 2019; Hyter *et al.*, 2018). Unfortunately, eighteen genes in the original pathway are missing, hence spurious connections could happen, like the one between APOB and COLEC11 (not present in the original pathway). On second thought, other connections are preserved, like COL4A1 with COL4A2 (their position in the genome is close) or MARCO with CD163. The latter couple is related to inflammatory responses when pathogens threaten normal cell activities (Novakowski *et al.*, 2016): although MARCO does not directly cause an inflammatory response, it does help other receptors in initiating it, resulting in up-regulated values for CD163 gene; this process is synthesized in both multilayer graphs in Figure 5.3 by the dark orange color of the edge connecting these two genes.

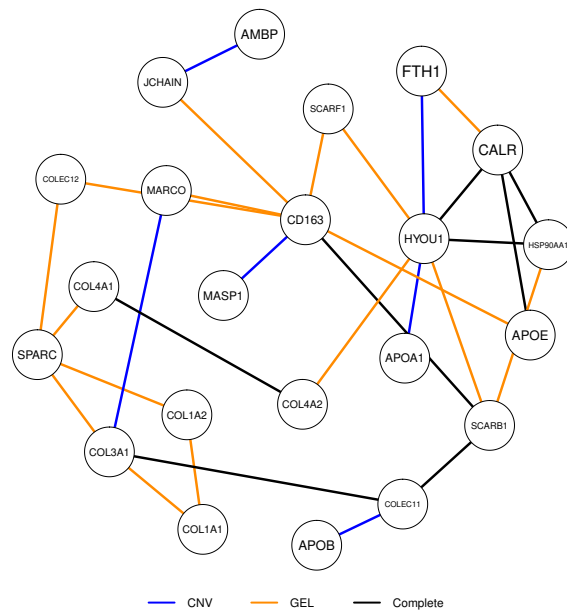


(A)

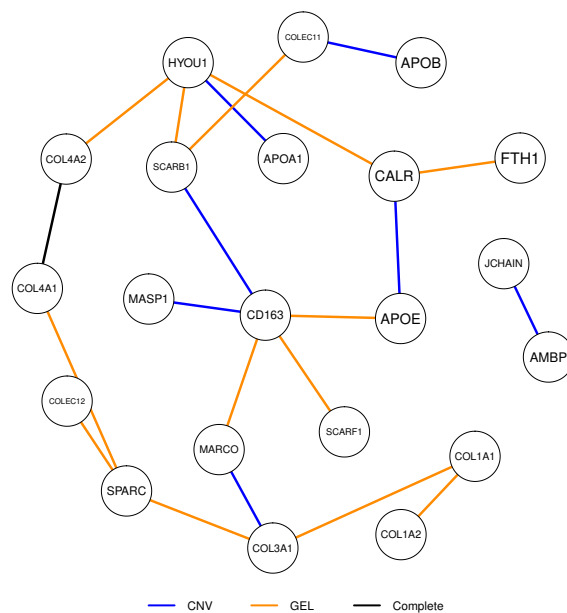


(B)

FIGURE 5.2: Inferred undirected multilayer graph for *Ca2+* pathway with cardinality $m = 1$ (panel (A)) in comparison with cardinality $m = 5$ (panel (B)); the nominal level in both cases is $\alpha = 0.001$.



(A)



(B)

FIGURE 5.3: Inferred undirected multilayer graph for *Binding and Uptake of Ligands by Scavenger Receptors* pathway, with cardinality $m = 1$ (panel (A)) in comparison with cardinality $m = 5$ (panel (B)); the nominal level in both cases is $\alpha = 0.001$.

Chapter 6

Conclusions

The problem of learning the structure of a multilayer network from a set of given heterogeneous data is a recent research area in graphical modelling. While most of the effort is devoted in models able to respect Markov properties between edges (i.e., random graph modelling theory), very few works attempted to develop models able to include Markov properties between node-layer tuples, due mainly to the difficulties in implementing a joint distributions of variables with different domains. Moreover, in case of multidimensional measurements on the same entity, as often happens with multi-omics data, the state-of-the-art modelling does not take into account the implicit dependencies between these measurements. Here, we tackled such problem when continuous and categorical data are available. Our first results, presented in this thesis, show promise. Many open problems await future research, and we mention some of them here.

Firstly, the proposed algorithm could be extended to other type of distributions, as, for example, all the members of the exponential family (e.g., binomial, Poisson or exponential distribution). The same algorithm could be indeed customized making use of the exponential family assumption in the conditional distributions of a node-layer tuple given the remaining variables. This type of approach would follow the footsteps of the EBDMRFs of Yang *et al.* (2014b), and would be interesting to obtain statistical results to guarantee consistency of the estimators in this wider setting. At a practical level, developing such a method would allow to explore in a more complete manner the relations between expressions, somatic mutations, methylation status, RNA-seq arrays and other types of biological data, in order to be able to (hopefully) advance in cancer research.

Secondly, when we treat categorical data, we suppose implicitly to have enough sample size in each class to obtain consistent maximum likelihood estimators for the interaction parameters of the model proposed. From a biological point of view, that

could be a potential limitation, especially in cancer studies, since some forms of cancers are due to rare somatic mutations in the genome (Marx, 2016; Gkolfinopoulos and Mountzios, 2018; Brown *et al.*, 2019; Scholl and Fröhling, 2019).

Thirdly, in this thesis, we considered structure learning only for undirected multilayer networks. Since there could be a causality dependence between the different measurements (e.g., copy number variation profiles can alter the expression of a gene, but the vice versa is not biologically feasible), it would be of interest to take into account the topological ordering of the variables, expanding to directed multilayer networks or partially directed multilayer networks.

Another possible development is to take into account missing data, in particular missing information on some measurements on genes. In this thesis project, we assumed to possess for each gene all possible measurements of interest, but this assumption is unlikely to be true. Structure learning in presence of missing data is an unexplored topic with an enormous potential in the omics setting we are considering.

It would be of interest also to configure other simulation experiments apart from the ones presented in Section 4.3 to check the model's ability to infer the desired multilayer structure, possibly combined with a formal analysis on the consistency of the algorithm. Moreover, we did not discuss in a deep manner run-time complexity of our algorithm. Clearly, further work should be developed to explore these issues.

On a technical note, all computing routines for the PC-CGRM were implemented in the R statistical software (R Core Team, 2018); in particular, of importance was the use of the routines “multinom” (in the “nnet” library) and “glm” for computing the node-layer conditional distributions, while the function “qgraph” of the homonymous library was used to display the network structures (Venables and Ripley, 2002; Epskamp *et al.*, 2012). At a software level, it could be of interest to develop an R library targeted to practitioners and researchers of this field.

Appendix A

Properties of conditional independence

In what follows, we prove implications (P1), (P2), (P3) and (P4).

Proof. We will prove them point by point.

(P1) *Statement:* if $X \perp\!\!\!\perp Y \mid Z$, also $Y \perp\!\!\!\perp X \mid Z$ (symmetry).

By hypothesis, $X \perp\!\!\!\perp Y \mid Z$. Then, since the product between two functions is a symmetric operator,

$$\begin{aligned} f_{XY|Z}(x, y \mid z) &= f_{X|Z}(x \mid z)f_{Y|Z}(y \mid z) \\ &= f_{Y|Z}(y \mid z)f_{X|Z}(x \mid z) \\ &= f_{YX|Z}(y, x \mid z), \end{aligned}$$

which is equivalent to saying $Y \perp\!\!\!\perp X \mid Z$.

(P2) *Statement:* if $X \perp\!\!\!\perp Y \mid Z$ and $U = h(X)$, also $U \perp\!\!\!\perp Y \mid Z$.

Starting from the conditional distribution of U, Y given Z , we have

$$f_{UY|Z}(u, y \mid z) = f_{h(X)Y|Z}(h(x), y \mid z);$$

by the property of inversion for transformations of random variables,

$$f_{h(X)Y|Z}(h(x), y \mid z) = f_{XY|Z}(h^{-1}(u), y \mid z) \left| \frac{d}{du} h^{-1}(u) \right|.$$

By hypothesis, $X \perp\!\!\!\perp Y \mid Z$, so

$$\begin{aligned} f_{XY|Z}(h^{-1}(u), y \mid z) \left| \frac{d}{du} h^{-1}(u) \right| &= f_{X|Z}(h^{-1}(u) \mid z) f_{Y|Z}(y \mid z) \left| \frac{d}{du} h^{-1}(u) \right| \\ &= f_{X|Z}(h^{-1}(u) \mid z) \left| \frac{d}{du} h^{-1}(u) \right| f_{Y|Z}(y \mid z) \end{aligned}$$

$$= f_{U|Z}(u | z)f_{Y|Z}(y | z).$$

Putting all together, we have that

$$f_{UY|Z}(u, y | z) = f_{U|Z}(u | z)f_{Y|Z}(y | z)$$

or, equivalently, $U \perp\!\!\!\perp Y | Z$.

(P3) *Statement:* if $X \perp\!\!\!\perp Y | Z$ and $U = h(X)$, also $X \perp\!\!\!\perp Y | (Z, U)$.

By (P2), since $X \perp\!\!\!\perp Y | Z$ and $U = h(X)$, we know that $U \perp\!\!\!\perp Y | Z$, meaning that $X, U \perp\!\!\!\perp Y | Z$. By (P1), it is possible to write $Y \perp\!\!\!\perp X, U | Z$. Then

$$f_{XYUZ}(x, y, u, z) = f_{XYU|Z}(x, y, u | z)f_Z(z) = f_{XU|Z}(x, u | z)f_{Y|Z}(y | z)f_Z(z);$$

But

$$f_{XU|Z}(x, u | z) = \frac{f_{XUZ}(x, u, z)}{f_Z(z)} = \frac{f_{X|UZ}(x | u, z)f_{UZ}(u, z)}{f_Z(z)},$$

so

$$\begin{aligned} f_{XYUZ}(x, y, u, z) &= \frac{f_{X|UZ}(x | u, z)f_{UZ}(u, z)}{f_Z(z)}f_{Y|Z}(y | z)f_Z(z) \\ &= f_{X|UZ}(x | u, z)f_{UZ}(u, z)f_{Y|Z}(y | z), \end{aligned}$$

which is equivalent to

$$f_{XY|UZ}(x, y | u, z) = f_{X|UZ}(x | u, z)f_{Y|Z}(y | z).$$

We only need to prove that $f_{Y|Z}(y | z) = f_{Y|UZ}(y | u, z)$. But

$$\begin{aligned} f_{Y|Z}(y | z) &= \frac{f_{YZ}(y, z)}{f_Z(z)} \frac{f_{UZ}(u, z)}{f_Z(z)} \frac{f_Z(z)}{f_{UZ}(u, z)} \\ &= f_{Y|Z}(y | z)f_{U|Z}(u | z) \frac{f_Z(z)}{f_{UZ}(u, z)} \\ &= f_{UY|Z}(u, y | z) \frac{f_Z(z)}{f_{UZ}(u, z)} \\ &= \frac{f_{UYZ}(u, y, z)}{f_Z(z)} \frac{f_Z(z)}{f_{UZ}(u, z)} \\ &= \frac{f_{UYZ}(u, y, z)}{f_{UZ}(u, z)} = f_{Y|UZ}(y | u, z). \end{aligned}$$

(P4) *Statement:* if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, also $X \perp\!\!\!\perp (W, Y) \mid Z$.

By hypotheses,

$$\begin{aligned}
f_{XYZW}(x, y, z, w) &= f_{XW|YZ}(x, w \mid y, z) f_{YZ}(y, z) \\
&= f_{X|YZ}(x \mid y, z) f_{W|YZ}(w \mid y, z) f_{YZ}(y, z) \\
&= \frac{f_{XY|Z}(x, y \mid z) f_Z(z)}{f_{YZ}(y, z)} f_{W|YZ}(w \mid y, z) f_{YZ}(y, z) \\
&= f_{XY|Z}(x, y \mid z) f_{W|YZ}(w \mid y, z) f_Z(z) \\
&= f_{X|Z}(x \mid z) f_{Y|Z}(y \mid z) f_{W|YZ}(w \mid y, z) f_Z(z) \\
&= f_{X|Z}(x \mid z) f_{Y|Z}(y \mid z) \frac{f_{WY|Z}(w, y \mid z) f_Z(z)}{f_{YZ}(y, z)} f_Z(z) \\
&= f_{X|Z}(x \mid z) f_{Y|Z}(y \mid z) \frac{f_{WY|Z}(w, y \mid z)}{f_{Y|Z}(y \mid z)} f_Z(z) \\
&= f_{X|Z}(x \mid z) f_{WY|Z}(w, y \mid z) f_Z(z),
\end{aligned}$$

which is equivalent to

$$f_{XYW|Z}(x, y, w \mid z) = f_{X|Z}(x \mid z) f_{WY|Z}(w, y \mid z),$$

hence $X \perp\!\!\!\perp (W, Y) \mid Z$.

□

Appendix B

Conditioning and marginalizing on CG distributions

We start from \mathbb{Y} random vector, following a CG-distribution reparameterized as in Chapter 3:

$$\begin{aligned} \ln f_{\mathbb{Y}}(y; g, h, \Omega) = & \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i \right) y_\gamma + \\ & - \frac{1}{2} \sum_{\gamma \in \Gamma} \phi_0^{\gamma\gamma} y_\gamma^2 - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\gamma y_\xi, \end{aligned}$$

where $\Delta = \{1, \dots, q\}$, $\Gamma = \{1, \dots, p\}$, $V = \Delta \cup \Gamma$, $V_M = \Delta \times \Gamma$, and the parameters have the same meaning as in Section 3.2.2 in Chapter 3. Denote with

$$(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \left(\lambda_0, \{\lambda_i\}_{i \in \Delta}, \{\lambda_{ij}\}_{\substack{i, j \in \Delta \\ i > j}}, \{\eta_0^\gamma\}_{\gamma \in \Gamma}, \{\eta_i^\gamma\}_{\substack{i \in \Delta \\ \gamma \in \Gamma}}, \{\phi_0^{\gamma\xi}\}_{\gamma, \xi \in \Gamma}, \{\phi_i^{\gamma\xi}\}_{\substack{i \in \Delta \\ \gamma, \xi \in \Gamma}} \right).$$

We compute now the following log-conditional distributions:

- (a) **Univariate conditional logistic regression:** $\ln f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i | y \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$;
- (b) **Univariate conditional linear regression:** $\ln f_{Y_\gamma | \mathbb{Y} \setminus Y_\gamma}(y_\gamma | y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$;
- (c) **Bivariate CG-regression:** $\ln f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | y \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$.

Case (a): Univariate conditional logistic regression

By definition of conditional probability, we have

$$\ln f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i | y \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \underbrace{\ln f_{\mathbb{Y}}(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (1)}} - \underbrace{\ln f_{\mathbb{Y} \setminus Z_i}(y \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (2)}}. \quad (\text{B.1})$$

We can rewrite Part (1) of equation (B.1) isolating all the terms related to Z_i :

$$\begin{aligned} \ln f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \lambda_0 + \lambda_i z_i + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \\ &+ \sum_{\gamma \in \Gamma} \eta_0^\gamma y_\gamma + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma + \sum_{\gamma \in \Gamma} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \\ &- \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi. \end{aligned}$$

Instead, Part (2) can be computed using Proposition 6.2 of Lauritzen (1996), reported in Chapter 2, Section 2.2, since the hypothesis of the proposition are satisfied. Indeed, the joint density of \mathbb{Y} is a CG distribution by construction; moreover, if we call $B = \{i\}$ (where $i \in \Delta$ corresponds to the index of the variable Z_i), and A the set of indices of all the other variables in \mathbb{Y} which are not Z_i , then the condition $B \perp\!\!\!\perp \Gamma \mid \Delta \setminus B$ is satisfied, meaning that $f_{\mathbb{Y} \setminus Z_i}(\mathbf{y} \setminus z_i; \mathbf{g}_{\mathbb{Y} \setminus Z_i}, h_{\mathbb{Y} \setminus Z_i}, \Omega_{\mathbb{Y} \setminus Z_i})$ is itself a CG distribution, where its canonical characteristics depend on \mathbf{z}_Δ only through $\mathbf{z}_{\Delta \setminus \{i\}}$ (see the proof of Proposition 6.2 in Lauritzen (1996) for further details). The canonical characteristics can be rewritten in the parametrization $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$ as defined in Chapter 3, Section 3.2, where now the different summations depend only on the elements in $\Delta \setminus \{i\}$, obtaining:

$$\begin{aligned} \ln f_{\mathbb{Y} \setminus Z_i}(\mathbf{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma + \\ &- \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi, \end{aligned}$$

where $\tilde{\lambda}_0$ is the normalizing constant when Z_i is not considered inside the set of variables. Notice that this can be proved also directly summarizing out Z_i from the joint distribution of \mathbb{Y} :

$$\begin{aligned} f_{\mathbb{Y} \setminus Z_i}(\mathbf{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \sum_{z_i \in \{0,1\}} f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \\ &= \sum_{z_i \in \{0,1\}} \exp \left\{ \lambda_0 + \lambda_i z_i + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \eta_0^\gamma y_\gamma + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma + \right. \end{aligned}$$

$$\begin{aligned}
& \left. + \sum_{\gamma \in \Gamma} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi \right\} \\
& = \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi \right\} \times \\
& \times \sum_{z_i=0,1} \exp \left\{ \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi \right\} \\
& = \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi \right\} \times \\
& \times \left\{ 1 + \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma y_\gamma - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \phi_i^{\gamma\xi} y_\gamma y_\xi \right\} \right\}.
\end{aligned}$$

If we define as

$$\tilde{\lambda}_0 = \lambda_0 + \ln \left(1 + \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma y_\gamma - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \phi_i^{\gamma\xi} y_\gamma y_\xi \right\} \right),$$

then

$$\begin{aligned}
& f_{\mathcal{Y} \setminus Z_i}(\mathbf{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \\
& = \exp \left\{ \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi \right\},
\end{aligned}$$

so that indeed we obtain the same quantity as before when we apply the logarithm:

$$\begin{aligned}
& \ln f_{\mathcal{Y} \setminus Z_i}(\mathbf{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \\
& = \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma > \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi
\end{aligned}$$

$$= \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\gamma \in \Gamma} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j \right) y_\gamma - \frac{1}{2} \sum_{\gamma, \xi \in \Gamma} \phi_0^{\gamma\xi} y_\gamma y_\xi - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\xi} z_j y_\gamma y_\xi.$$

Then (B.1) can be computed as

$$\begin{aligned} \ln f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \ln f_{\mathbb{Y}}(\mathbb{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) - \ln f_{\mathbb{Y} \setminus Z_i}(\mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \\ &= \lambda_0 - \tilde{\lambda}_0 + \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi \\ &\propto \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi, \end{aligned}$$

so that

$$f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \propto \exp \left\{ \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma z_i y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi \right\}.$$

In particular, since $z_i \in \{0, 1\}$,

$$\begin{aligned} f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 0 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &\propto \exp \left\{ \lambda_i \cdot 0 + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} \cdot 0 \cdot z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma \cdot 0 \cdot y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} \cdot 0 \cdot y_\gamma y_\xi \right\} \\ &= 1, \end{aligned}$$

and

$$f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 1 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \propto \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} y_\gamma y_\xi \right\}.$$

Computing the conditional log-odds, we find that they are linear in parameters:

$$\begin{aligned} \ln \frac{f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 1 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}{f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 0 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})} &= \ln f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 1 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) - \ln f_{Z_i | \mathbb{Y} \setminus Z_i}(z_i = 0 | \mathbb{y} \setminus z_i; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \\ &\propto \ln \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} y_\gamma y_\xi \right\} - \ln 1 \end{aligned}$$

$$= \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\gamma \in \Gamma} \eta_i^\gamma y_\gamma - \frac{1}{2} \sum_{\substack{\gamma, \xi \in \Gamma \\ \gamma \neq \xi}} \phi_i^{\gamma\xi} y_\gamma y_\xi.$$

Hence, maximizing this conditional log-likelihood function can be done via fitting a logistic regression having Z_i as response, and $\mathbf{Z}_{\Delta \setminus \{i\}}$, \mathbf{Y}_Γ and $\mathbf{Y}_\Gamma^T \mathbf{Y}_\Gamma$ as predictors.

Case (b): Univariate conditional linear regression

By definition of conditional probability, we have

$$\ln f_{Y_\gamma | \mathbb{Y} \setminus Y_\gamma}(y_\gamma | \mathbf{y} \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \underbrace{\ln f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (1)}} - \underbrace{\ln f_{\mathbb{Y} \setminus Y_\gamma}(\mathbf{y} \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (2)}}. \quad (\text{B.2})$$

We can rewrite Part (1) of the above equation isolating all the terms related to Y_γ :

$$\begin{aligned} \ln f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \eta_0^\gamma y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{i \in \Delta} \eta_i^\gamma z_i y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{i \in \Delta} \eta_i^\mu z_i y_\mu + \\ &- \frac{1}{2} \left[\phi_0^{\gamma\gamma} y_\gamma^2 + \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \phi_0^{\gamma\xi} y_\gamma y_\xi + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \phi_0^{\mu\xi} y_\mu y_\xi + \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \sum_{i \in \Delta} \phi_i^{\mu\xi} z_i y_\mu y_\xi \right]. \end{aligned}$$

Instead, Part (2) can be computed using Proposition 6.1 of Lauritzen (1996), reported in Chapter 2, Section 2.2. Indeed, the joint density of \mathbb{Y} is a CG distribution by construction; let us call $B = \{\gamma\}$ (where $\gamma \in \Gamma$ corresponds to the index of the variable Y_γ), and A the set of indices of all the other variables in \mathbb{Y} which are not Y_γ . Since $B \subseteq \Gamma$, all the hypotheses of the Proposition 6.1 are satisfied, i.e., $f_{\mathbb{Y} \setminus Y_\gamma}(\mathbf{y} \setminus y_\gamma; g_{\mathbb{Y} \setminus Y_\gamma}, h_{\mathbb{Y} \setminus Y_\gamma}, \Omega_{\mathbb{Y} \setminus Y_\gamma})$ is itself a CG distribution.

We do not replace the quantities defining (g_A, h_A, Ω_A) with the corresponding values of $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$ as we did in the discrete case, because it is easier to compute the marginal of $\mathbb{Y} \setminus Y_\gamma$ with directly integrating out Y_γ from the joint distribution of \mathbb{Y} :

$$\begin{aligned} f_{\mathbb{Y} \setminus Y_\gamma}(\mathbf{y} \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \int_{-\infty}^{+\infty} f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \partial y_\gamma \\ &= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j \right\} \times \\ &\times \int_{-\infty}^{+\infty} \exp \left\{ \sum_{\mu \in \Gamma} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\xi, \mu \in \Gamma} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \partial y_\gamma \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \times \\
&\times \int_{-\infty}^{+\infty} \exp \left\{ \left(\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i \right) y_\gamma - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 - \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi y_\gamma \right\} \partial y_\gamma \\
&= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \times \\
&\times \int_{-\infty}^{+\infty} \exp \left\{ \left(\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right) y_\gamma - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 \right\} \partial y_\gamma \\
&= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \times \\
&\times \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} \phi_0^{\gamma\gamma} \left[y_\gamma^2 - \frac{2}{\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right) y_\gamma \right] \right\} \partial y_\gamma.
\end{aligned}$$

Now denote with

$$\kappa = \left[\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right] / \phi_0^{\gamma\gamma};$$

then

$$f_{Y \setminus Y_\gamma}(y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) =$$

$$\begin{aligned}
&= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \times \\
&\times \exp \left\{ \frac{1}{2} \phi_0^{\gamma\gamma} \kappa^2 \right\} \times \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} \phi_0^{\gamma\gamma} [y_\gamma^2 - 2\kappa y_\gamma + \kappa^2] \right\} \partial y_\gamma \\
&= \exp \left\{ \lambda_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\} \times
\end{aligned}$$

$$\times \exp \left\{ \frac{1}{2} \phi_0^{\gamma\gamma} \kappa^2 \right\} \times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}}.$$

Let

$$\tilde{\lambda}_0 = \lambda_0 + \frac{1}{2} \phi_0^{\gamma\gamma} \kappa^2 + \frac{1}{2} \ln(2\pi) - \ln \phi_0^{\gamma\gamma},$$

so that

$$f_{Y \setminus Y_\gamma}(y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \exp \left\{ \tilde{\lambda}_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu + \right. \\ \left. - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu \right\}$$

and

$$\ln f_{Y \setminus Y_\gamma}(y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \tilde{\lambda}_0 + \sum_{i \in \Delta} \lambda_i z_i + \sum_{\substack{i, j \in \Delta \\ i > j}} \lambda_{ij} z_i z_j + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\eta_0^\mu + \sum_{i \in \Delta} \eta_i^\mu z_i \right) y_\mu + \\ - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu, \xi \neq \gamma}} \left(\phi_0^{\xi\mu} + \sum_{i \in \Delta} \phi_i^{\xi\mu} z_i \right) y_\xi y_\mu.$$

Then (B.2) can be computed as

$$\begin{aligned} \ln f_{Y_\gamma | Y \setminus Y_\gamma}(y_\gamma | y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \ln f_Y(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) - \ln f_{Y \setminus Y_\gamma}(y \setminus y_\gamma; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \\ &= \lambda_0 - \tilde{\lambda}_0 + \eta_0^\gamma y_\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i y_\gamma - \frac{1}{2} \left[\phi_0^{\gamma\gamma} y_\gamma^2 + \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \phi_0^{\gamma\xi} y_\gamma y_\xi + \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i y_\gamma y_\xi \right] \\ &= \lambda_0 - \tilde{\lambda}_0 + \left[\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \frac{1}{2} \sum_{\substack{\xi \in \Gamma \\ \xi \neq \gamma}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right] y_\gamma - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 \\ &= \lambda_0 - \tilde{\lambda}_0 + \left[\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right] y_\gamma - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 \\ &\propto \left[\eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi \right] y_\gamma - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2, \end{aligned}$$

which is the kernel of a Gaussian distribution with conditional mean

$$E(Y_\gamma | \mathbb{Y} \setminus Y_\gamma) = \eta_0^\gamma + \sum_{i \in \Delta} \eta_i^\gamma z_i - \sum_{\substack{\xi \in \Gamma \\ \gamma > \xi}} \left(\phi_0^{\gamma\xi} + \sum_{i \in \Delta} \phi_i^{\gamma\xi} z_i \right) y_\xi$$

and conditional variance (not depending from any categorical variable)

$$\text{Var}(Y_\gamma | \mathbb{Y} \setminus Y_\gamma) = 1/\phi_0^{\gamma\gamma}.$$

Hence, maximizing this conditional log-likelihood function can be done via fitting a linear regression having Y_γ as response, and \mathbf{Z}_Δ , $\mathbf{Y}_{\Gamma \setminus \{\gamma\}}$ and their interactions as predictors.

Case (c): Bivariate CG-regression

By theory (Chapter 2, Section 3.2), we suppose that $f_{\mathbb{Y}}(\mathbf{y}; g, h, \Omega)$ is a CG distribution. If we let $B = \{i, \gamma\}$ and $A = V \setminus \{i, \gamma\}$, then, by Proposition 6.6, we can deduce that

$$f_{\mathbb{Y}_B | \mathbb{Y}_A}(\mathbf{y}_B | \mathbf{y}_A; g_{B|A}, h_{B|A}, \Omega_{B|A}) = f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | \mathbb{Y} \setminus \{z_i, y_\gamma\}; g_{i, \gamma | V \setminus \{i, \gamma\}}, h_{i, \gamma | V \setminus \{i, \gamma\}}, \Omega_{i, \gamma | V \setminus \{i, \gamma\}})$$

is a CG regression. The next step is to compute its reparameterization in $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})$. By definition of conditional probability, we have

$$\ln f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | \mathbb{Y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \underbrace{\ln f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (1)}} - \underbrace{\ln f_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(\mathbb{Y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi})}_{\text{Part (2)}}. \quad (\text{B.3})$$

We can rewrite Part (1) of the above equation isolating all the terms related to the couple (Z_i, Y_γ) :

$$\begin{aligned} \ln f_{\mathbb{Y}}(\mathbf{y}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \lambda_0 + \lambda_i z_i + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \sum_{\substack{j, k \in \Delta \\ j > k, \\ j, k \neq i}} \lambda_{jk} z_j z_k + \eta_0^\gamma y_\gamma + \eta_i^\gamma z_i y_\gamma + \\ &+ \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu z_i y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \left[\phi_0^{\gamma\gamma} y_\gamma^2 + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 + y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\gamma\mu} y_\mu + \right. \\ &+ y_\gamma z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu + y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_0^{\mu\xi} y_\mu y_\xi + z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j y_\mu y_\xi \left. \right]. \end{aligned}$$

Now, to compute Part (2) of (B.3), we have to compute the following:

$$f_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(y \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \int_{-\infty}^{+\infty} \left(\sum_{z_i \in \{0,1\}} f_{\mathbb{Y}}(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \right) \partial y_\gamma.$$

In particular,

$$\begin{aligned} & \sum_{z_i \in \{0,1\}} f_{\mathbb{Y}}(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) = \\ & = \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j,k \in \Delta \\ j > k \\ j,k \neq i}} \lambda_{jk} z_j z_k + \eta_0^\gamma y_\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu + \right. \\ & \left. - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \left[y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right] \right\} \times \\ & \times \sum_{z_i=0,1} \exp \left\{ \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + z_i \left(\eta_i^\gamma y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu \right) - \frac{1}{2} z_i \left(y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi \right) \right\} \\ & = \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j,k \in \Delta \\ j > k \\ j,k \neq i}} \lambda_{jk} z_j z_k + \eta_0^\gamma y_\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu + \right. \\ & \left. - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \left[y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right] \right\} \times \\ & \times \left\{ 1 + \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \eta_i^\gamma y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - \frac{1}{2} \left(y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi \right) \right\} \right\}. \end{aligned}$$

This result is of the form

$$\exp\{A(y_\gamma)\} (1 + \exp\{B(y_\gamma)\}) = \exp\{A(y_\gamma)\} + \exp\{(A+B)(y_\gamma)\},$$

where A and B are additive functions, so that by the linearity properties of the integrals we obtain

$$\int_{\mathbb{R}} \exp\{A(y_\gamma)\} (1 + \exp\{B(y_\gamma)\}) \partial y_\gamma = \int_{\mathbb{R}} \exp\{A(y_\gamma)\} \partial y_\gamma + \int_{\mathbb{R}} \exp\{(A+B)(y_\gamma)\} \partial y_\gamma.$$

The first integral on the right side of the last equation corresponds to

$$\begin{aligned} \int_{\mathbb{R}} \exp\{A(y_\gamma)\} \partial y_\gamma &= \int_{\mathbb{R}} \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \eta_0^\gamma y_\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu + \right. \\ &\quad \left. - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \left[y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu + \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right] \right\} \partial y_\gamma \\ &= \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right\} \times \\ &\quad \times \int_{\mathbb{R}} \exp \left\{ \eta_0^\gamma y_\gamma + y_\gamma \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 - \frac{1}{2} y_\gamma \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right\} \partial y_\gamma \\ &= \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right\} \times \\ &\quad \times \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \phi_0^{\gamma\gamma} \left[y_\gamma^2 - \frac{2}{\phi_0^{\gamma\gamma}} y_\gamma \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) \right] \right\} \partial y_\gamma \\ &= \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right\} \times \\ &\quad \times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}} \exp \left\{ \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right)^2 \right\} \end{aligned}$$

because of similar counts to the ones we saw in the continuous case. With similar computations we obtain also the second integral on the right side of the previously seen equation:

$$\int_{\mathbb{R}} \exp\{(A+B)(y_\gamma)\} \partial y_\gamma = \int_{\mathbb{R}} \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \eta_0^\gamma y_\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu \right\} \partial y_\gamma$$

$$\begin{aligned}
& -\frac{1}{2}\phi_0^{\gamma\gamma}y_\gamma^2 - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_0^{\mu\mu}y_\mu^2 - \frac{1}{2}\left[y_\gamma\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\left(\phi_0^{\gamma\mu} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\gamma\mu}z_j\right)y_\mu + \sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\left(\phi_0^{\mu\xi} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\mu\xi}z_j\right)y_\mu y_\xi\right] + \\
& + \lambda_i + \sum_{\substack{k\in\Delta \\ i>k}}\lambda_{ik}z_k + \eta_i^\gamma y_\gamma + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_i^\mu y_\mu - \frac{1}{2}\left(y_\gamma\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_i^{\gamma\mu}y_\mu + \sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\phi_i^{\mu\xi}y_\mu y_\xi\right)\Big\}\partial y_\gamma \\
& = \exp\left\{\lambda_0 + \sum_{\substack{j\in\Delta \\ j\neq i}}\lambda_j z_j + \sum_{\substack{j,k\in\Delta \\ j>k \\ j,k\neq i}}\lambda_{jk}z_j z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_0^\mu y_\mu + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\sum_{j\in\Delta} \eta_j^\mu z_j y_\mu - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_0^{\mu\mu}y_\mu^2 + \right. \\
& \left. - \frac{1}{2}\sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\left(\phi_0^{\mu\xi} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\mu\xi}z_j\right)y_\mu y_\xi + \lambda_i + \sum_{\substack{k\in\Delta \\ i>k}}\lambda_{ik}z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_i^\mu y_\mu - \frac{1}{2}\sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\phi_i^{\mu\xi}y_\mu y_\xi\right\}\times \\
& \times \int_{\mathbb{R}} \exp\left\{\eta_0^\gamma y_\gamma + y_\gamma\sum_{\substack{j\in\Delta \\ j\neq i}}\eta_j^\gamma z_j - \frac{1}{2}\phi_0^{\gamma\gamma}y_\gamma^2 - \frac{1}{2}y_\gamma\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\left(\phi_0^{\gamma\mu} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\gamma\mu}z_j\right)y_\mu + \eta_i^\gamma y_\gamma - \frac{1}{2}y_\gamma\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_i^{\gamma\mu}y_\mu\right\}\partial y_\gamma \\
& = \exp\left\{\lambda_0 + \sum_{\substack{j\in\Delta \\ j\neq i}}\lambda_j z_j + \sum_{\substack{j,k\in\Delta \\ j>k \\ j,k\neq i}}\lambda_{jk}z_j z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_0^\mu y_\mu + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\sum_{j\in\Delta} \eta_j^\mu z_j y_\mu - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_0^{\mu\mu}y_\mu^2 + \right. \\
& \left. - \frac{1}{2}\sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\left(\phi_0^{\mu\xi} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\mu\xi}z_j\right)y_\mu y_\xi + \lambda_i + \sum_{\substack{k\in\Delta \\ i>k}}\lambda_{ik}z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_i^\mu y_\mu - \frac{1}{2}\sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\phi_i^{\mu\xi}y_\mu y_\xi\right\}\times \\
& \times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}} \exp\left\{\frac{1}{2\phi_0^{\gamma\gamma}}\left(\eta_0^\gamma + \sum_{\substack{j\in\Delta \\ j\neq i}}\eta_j^\gamma z_j - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\left(\phi_0^{\gamma\mu} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\gamma\mu}z_j\right)y_\mu + \eta_i^\gamma - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_i^{\gamma\mu}y_\mu\right)^2\right\}.
\end{aligned}$$

Summing up these two results we obtain

$$\begin{aligned}
& f_{\mathbb{V}\setminus\{z_i, Y_\gamma\}}(\mathfrak{y}\setminus\{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \Phi) = \int_{\mathbb{R}} \exp\{A(y_\gamma)\}\partial y_\gamma + \int_{\mathbb{R}} \exp\{(A+B)(y_\gamma)\}\partial y_\gamma \\
& = \exp\left\{\lambda_0 + \sum_{\substack{j\in\Delta \\ j\neq i}}\lambda_j z_j + \sum_{\substack{j,k\in\Delta \\ j>k \\ j,k\neq i}}\lambda_{jk}z_j z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_0^\mu y_\mu + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\sum_{j\in\Delta} \eta_j^\mu z_j y_\mu - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_0^{\mu\mu}y_\mu^2 + \right. \\
& \left. - \frac{1}{2}\sum_{\substack{\mu,\xi\in\Gamma \\ \mu\neq\xi\neq\gamma}}\left(\phi_0^{\mu\xi} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\mu\xi}z_j\right)y_\mu y_\xi\right\}\times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}} \exp\left\{\frac{1}{2\phi_0^{\gamma\gamma}}\left(\eta_0^\gamma + \sum_{\substack{j\in\Delta \\ j\neq i}}\eta_j^\gamma z_j - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\left(\phi_0^{\gamma\mu} + \sum_{\substack{j\in\Delta \\ j\neq i}}\phi_j^{\gamma\mu}z_j\right)y_\mu\right)^2\right\} + \\
& + \exp\left\{\lambda_0 + \sum_{\substack{j\in\Delta \\ j\neq i}}\lambda_j z_j + \sum_{\substack{j,k\in\Delta \\ j>k \\ j,k\neq i}}\lambda_{jk}z_j z_k + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\eta_0^\mu y_\mu + \sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\sum_{j\in\Delta} \eta_j^\mu z_j y_\mu - \frac{1}{2}\sum_{\substack{\mu\in\Gamma \\ \mu\neq\gamma}}\phi_0^{\mu\mu}y_\mu^2 + \right.
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi + \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi \Big\} \times \\
& \times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}} \exp \left\{ \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu + \eta_i^\gamma - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu \right)^2 \right\} \\
& = \exp \left\{ \lambda_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right\} \times \\
& \times \frac{\sqrt{2\pi}}{\phi_0^{\gamma\gamma}} \exp \left\{ \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right)^2 \right\} \times \\
& \times \left[1 + \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_i^\gamma - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu \right)^2 \right. \right. \\
& \left. \left. + \frac{1}{\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) \left(\eta_i^\gamma - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu \right) \right\} \right].
\end{aligned}$$

Calling

$$\begin{aligned}
\tilde{\lambda}_0 & = \lambda_0 + \frac{1}{2} \ln(2\pi) - \ln \phi_0^{\gamma\gamma} + \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right)^2 + \\
& + \ln \left\{ 1 + \exp \left\{ \lambda_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \frac{1}{2\phi_0^{\gamma\gamma}} \left(\eta_i^\gamma - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu \right)^2 \right. \right. \\
& \left. \left. + \frac{1}{\phi_0^{\gamma\gamma}} \left(\eta_0^\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \left(\phi_0^{\gamma\mu} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) \left(\eta_i^\gamma - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu \right) \right\} \right\},
\end{aligned}$$

we can simplify the exponentials above, leading to

$$\begin{aligned}
f_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(\mathbf{y} \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) & = \exp \left\{ \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \right. \\
& \left. + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi \right\}
\end{aligned}$$

and

$$\begin{aligned} \ln f_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(y \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \tilde{\lambda}_0 + \sum_{\substack{j \in \Delta \\ j \neq i}} \lambda_j z_j + \sum_{\substack{j, k \in \Delta \\ j > k \\ j, k \neq i}} \lambda_{jk} z_j z_k + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_0^\mu y_\mu + \\ &+ \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\mu z_j y_\mu - \frac{1}{2} \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\mu\mu} y_\mu^2 - \frac{1}{2} \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \left(\phi_0^{\mu\xi} + \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\mu\xi} z_j \right) y_\mu y_\xi. \end{aligned}$$

Now it's possible to compute the formula of the conditional distribution (B.3):

$$\begin{aligned} \ln f_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i, y_\gamma | y \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) &= \ln f_{\mathbb{Y}}(y; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) - \ln f_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(y \setminus \{z_i, y_\gamma\}; \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\Phi}) \\ &= \lambda_0 - \tilde{\lambda}_0 + \lambda_i z_i + \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_i z_k + \eta_0^\gamma y_\gamma + \eta_i^\gamma z_i y_\gamma + \sum_{\substack{j \in \Delta \\ j \neq i}} \eta_j^\gamma z_j y_\gamma + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu z_i y_\mu - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 + \\ &- \frac{1}{2} y_\gamma \left[\sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_0^{\gamma\mu} y_\mu + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \phi_i^{\gamma\mu} y_\mu + \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \sum_{\substack{j \in \Delta \\ j \neq i}} \phi_j^{\gamma\mu} z_j y_\mu \right] - \frac{1}{2} z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu \neq \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi \\ &= \lambda_0 - \tilde{\lambda}_0 + \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \\ &+ y_\gamma \left(\eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) - \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2 \\ &\propto \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + y_\gamma \left(\eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu \right) + \\ &- \frac{1}{2} \phi_0^{\gamma\gamma} y_\gamma^2, \end{aligned}$$

which is indeed the reparameterization required:

$$\begin{aligned} \Omega_{B|A}(\mathbf{z}_{\Delta_B}) &= \Omega_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) = \phi_0^{\gamma\gamma}, \\ h_{B|A}(\mathbf{z}_{\Delta_B}) &= h_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) = h(\mathbf{z}_\Delta)_{Z_i, Y_\gamma} - \Omega(\mathbf{z}_\Delta)_{[Z_i, Y_\gamma] | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}} \mathbf{y}_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}} \\ &= \eta_0^\gamma + \sum_{j \in \Delta} \eta_j^\gamma z_j - \sum_{\substack{\mu \in \Gamma \\ \gamma > \mu}} \left(\phi_0^{\gamma\mu} + \sum_{j \in \Delta} \phi_j^{\gamma\mu} z_j \right) y_\mu, \\ g_{B|A}(\mathbf{z}_{\Delta_B}) &= g_{Z_i, Y_\gamma | \mathbb{Y} \setminus \{Z_i, Y_\gamma\}}(z_i) \\ &= g(\mathbf{z}_\Delta) + h(\mathbf{z}_\Delta)_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}^T \mathbf{y}_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}} - \frac{1}{2} \mathbf{y}_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}}^T \Omega(\mathbf{z}_\Delta)_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}} \mathbf{y}_{\mathbb{Y} \setminus \{Z_i, Y_\gamma\}} + \\ &- \ln \kappa(y \setminus \{z_i, y_\gamma\}) \end{aligned}$$

$$= \lambda_i z_i + z_i \sum_{\substack{k \in \Delta \\ i > k}} \lambda_{ik} z_k + z_i \sum_{\substack{\mu \in \Gamma \\ \mu \neq \gamma}} \eta_i^\mu y_\mu - z_i \sum_{\substack{\mu, \xi \in \Gamma \\ \mu > \xi \\ \mu, \xi \neq \gamma}} \phi_i^{\mu\xi} y_\mu y_\xi + \lambda_0 - \tilde{\lambda}_0.$$

Appendix C

Output of PC-CGRM on real data

We display all twenty graphical configurations we computed, according by maximum cardinality of the neighbourhood of a node-layer couple and the nominal level required by the PC-CGRM algorithm. Values of these tuning parameters are in Chapter 5, Section 5.2.

Ca²⁺ pathway

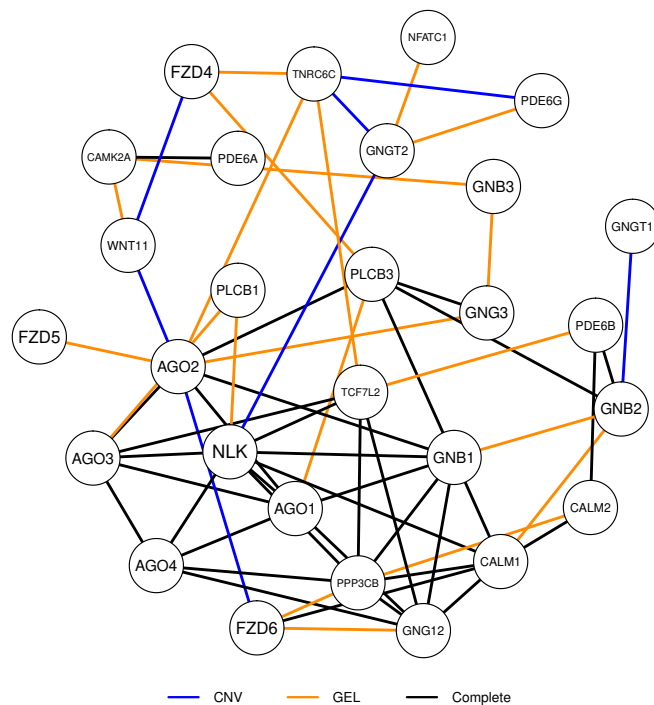


FIGURE C.1: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.001$.

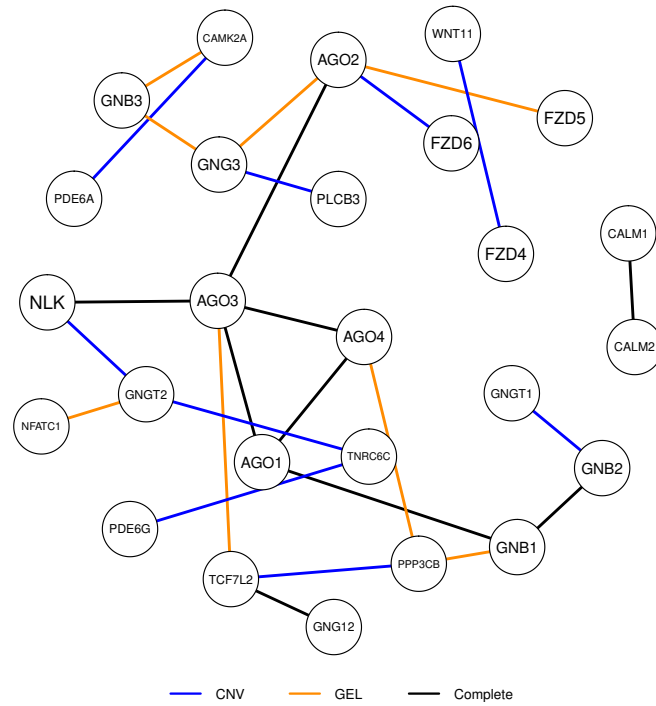


FIGURE C.2: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.001$.

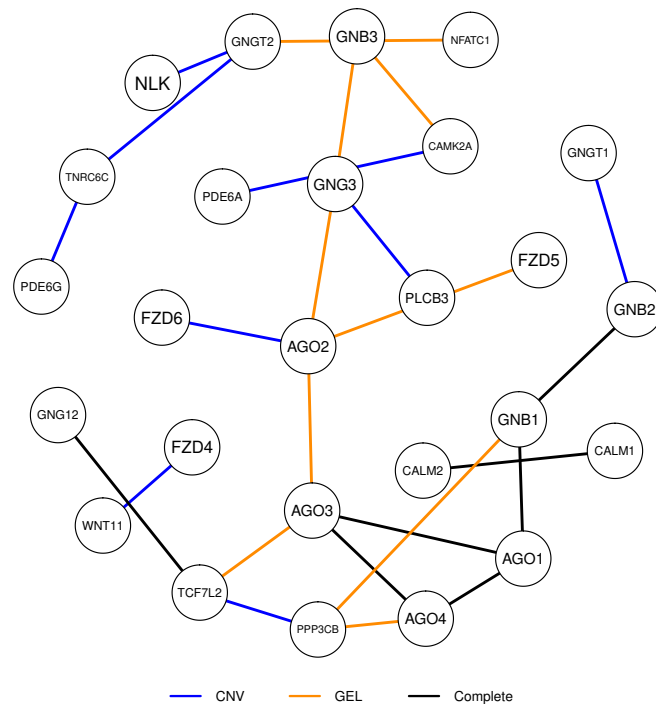


FIGURE C.3: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.001$.

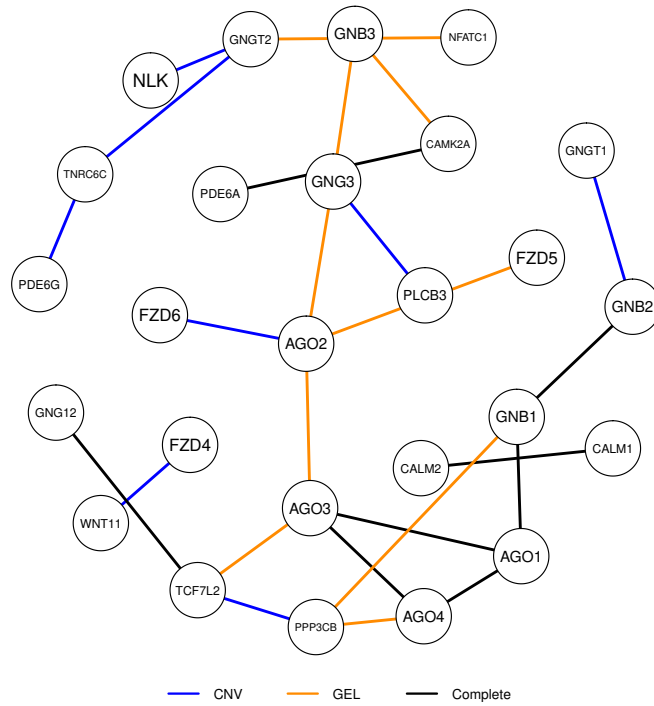


FIGURE C.4: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.001$.

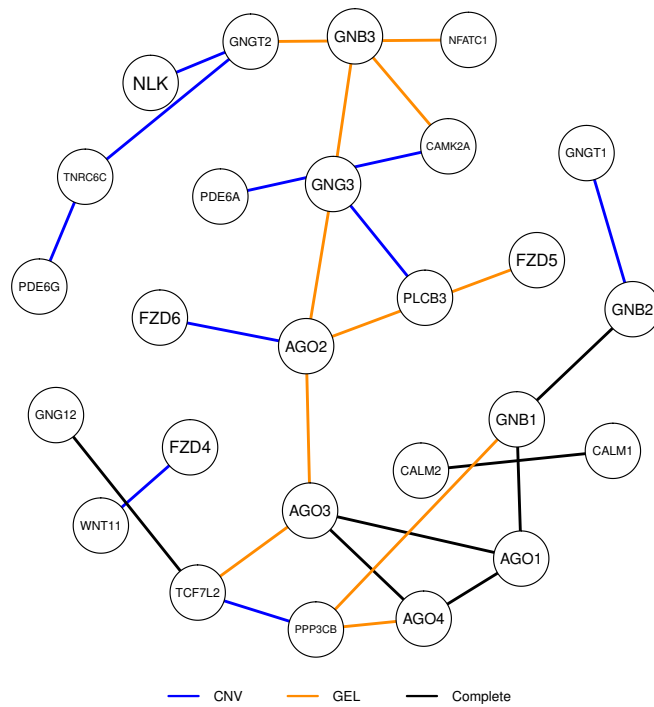


FIGURE C.5: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.001$.

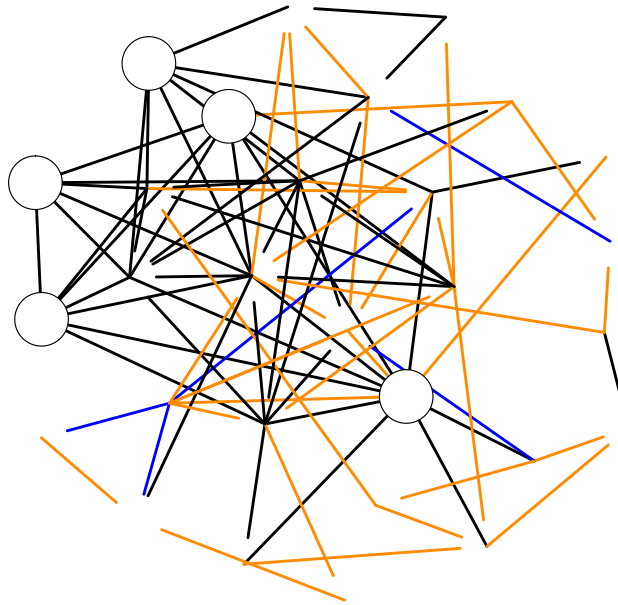


FIGURE C.6: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.01$.

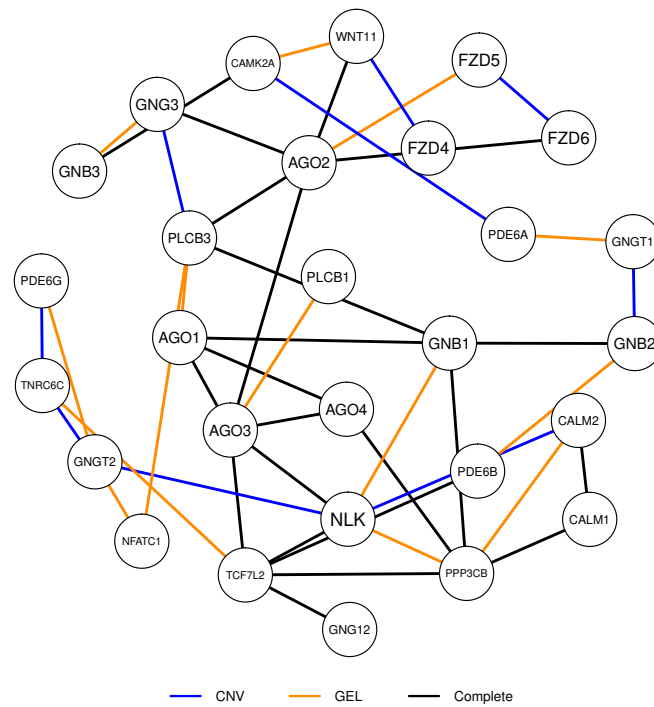


FIGURE C.7: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.01$.

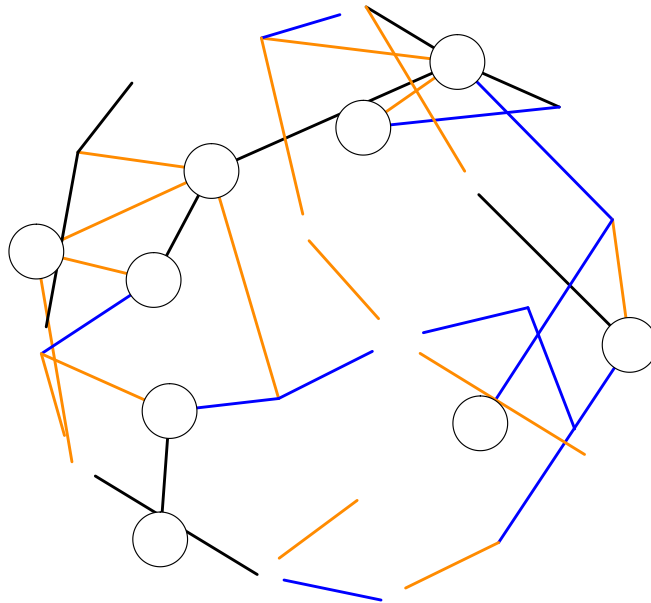


FIGURE C.8: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.01$.

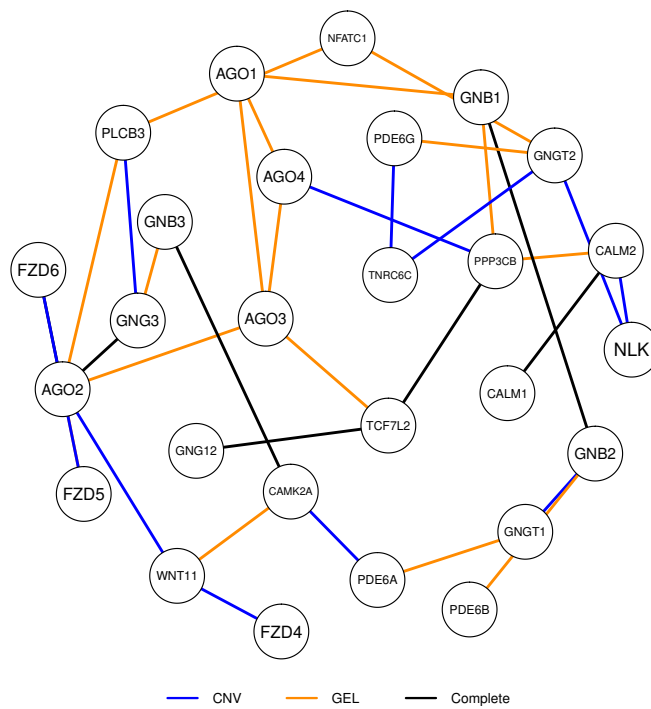


FIGURE C.9: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.01$.

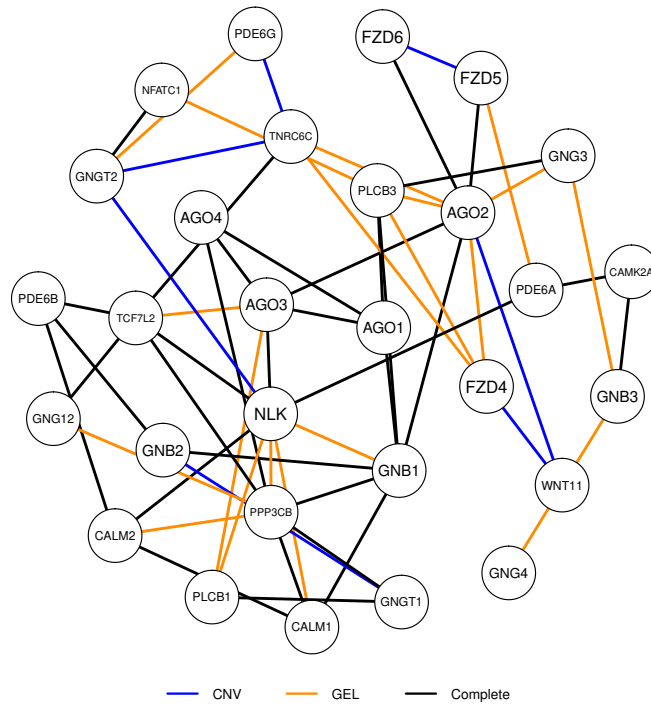


FIGURE C.12: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.05$.

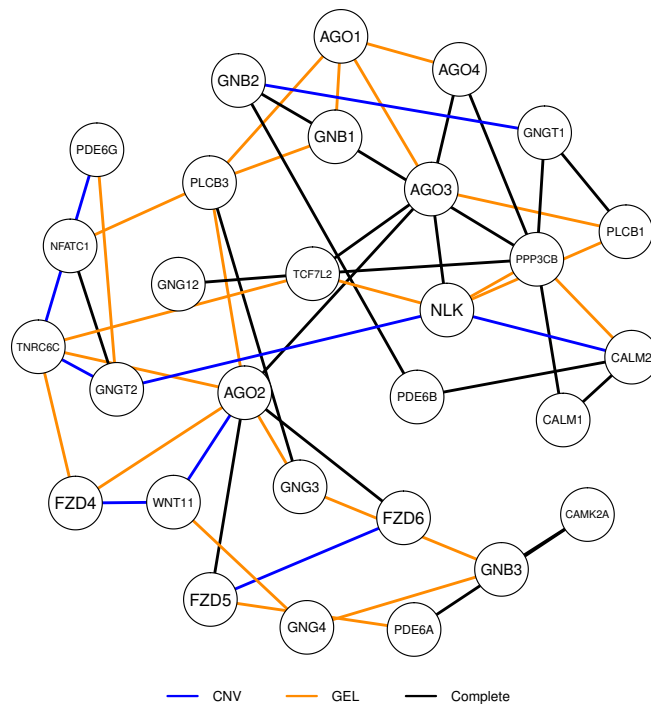


FIGURE C.13: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.05$.

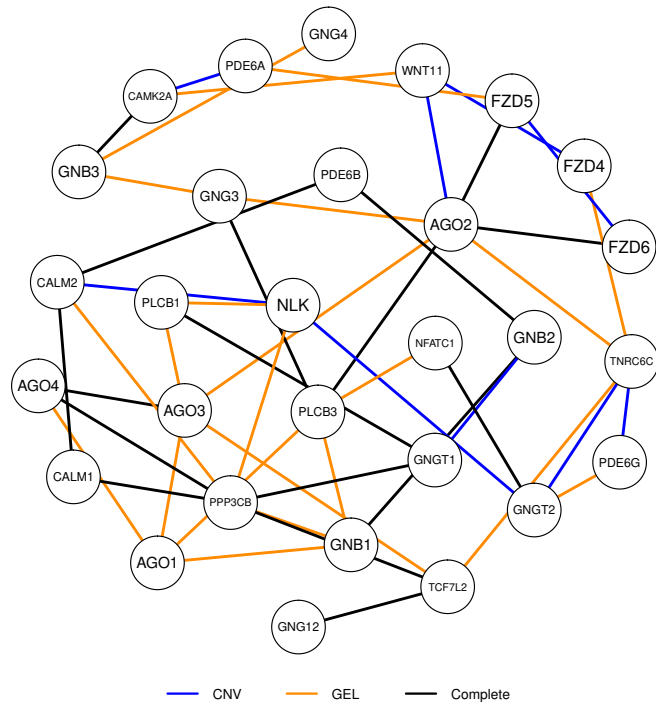


FIGURE C.14: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.05$.

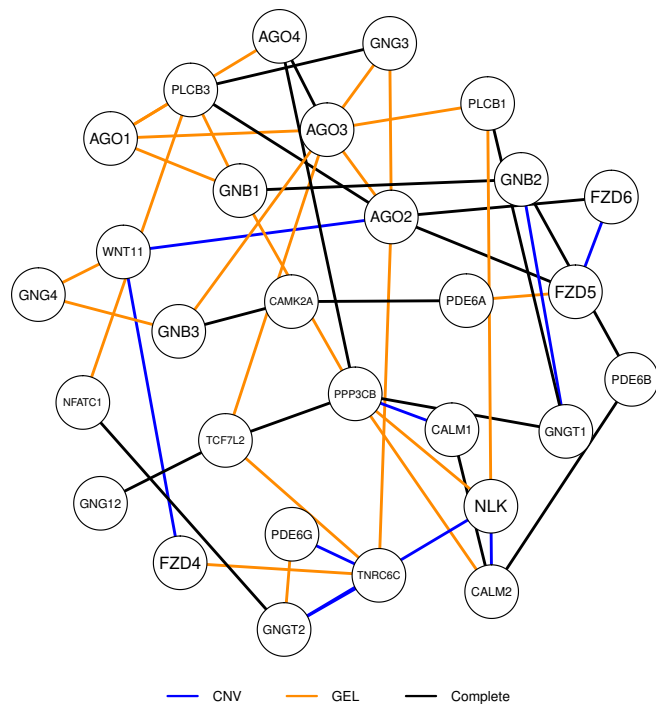


FIGURE C.15: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.05$.

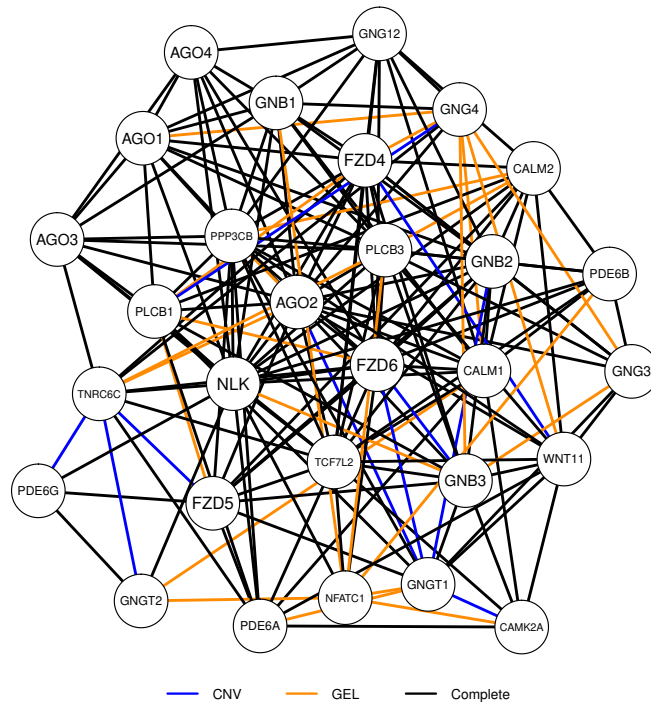


FIGURE C.16: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.1$.

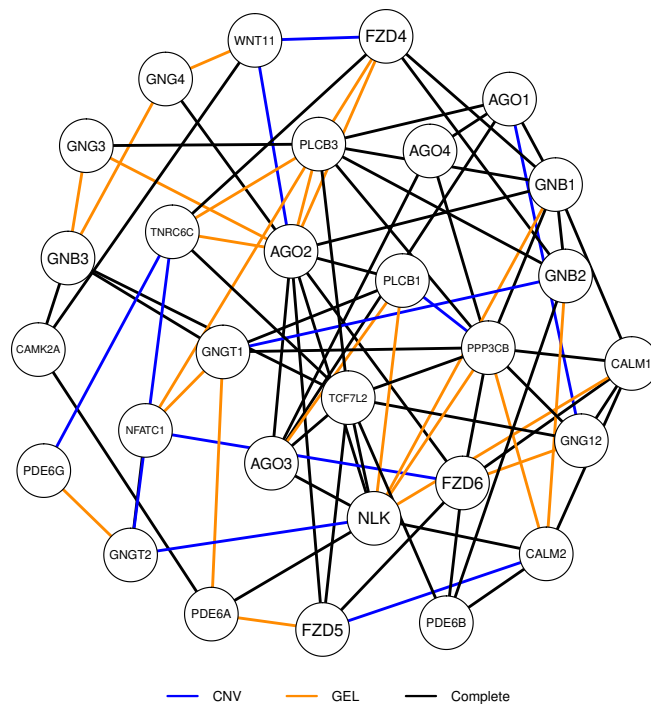


FIGURE C.17: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.1$.

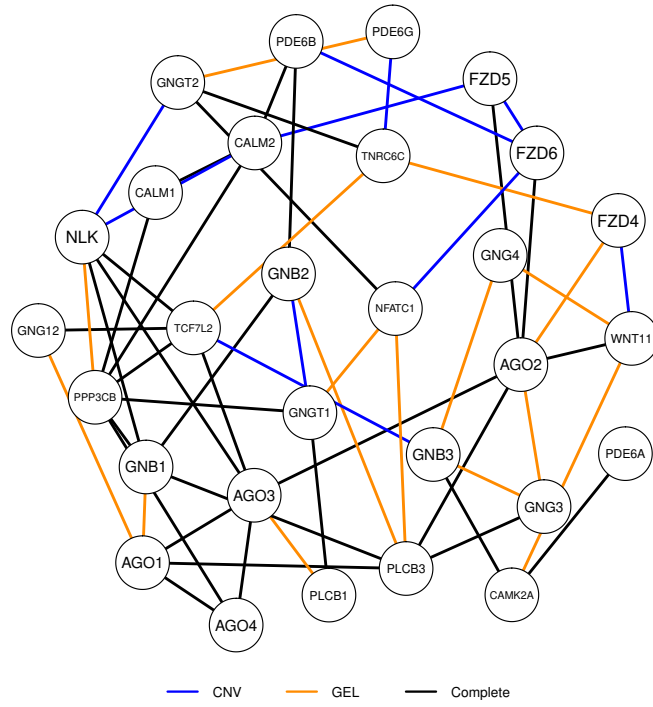


FIGURE C.18: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.1$.

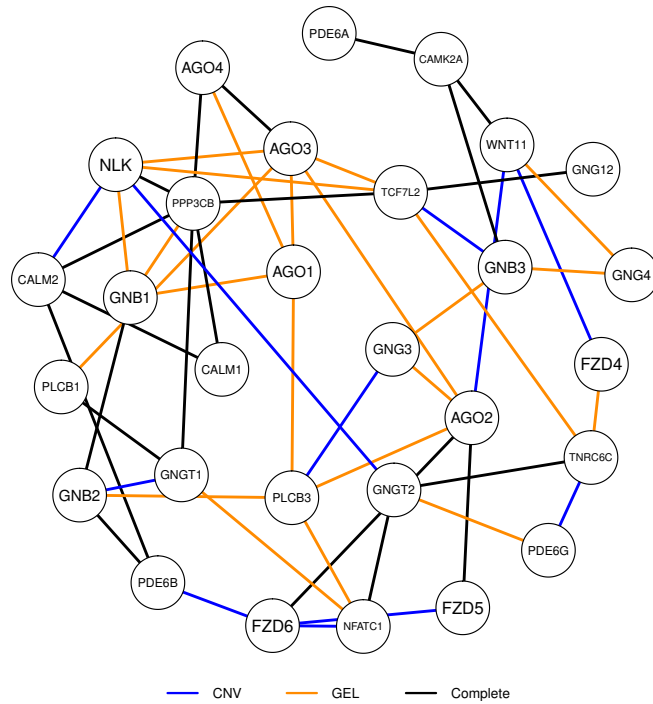


FIGURE C.19: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.1$.

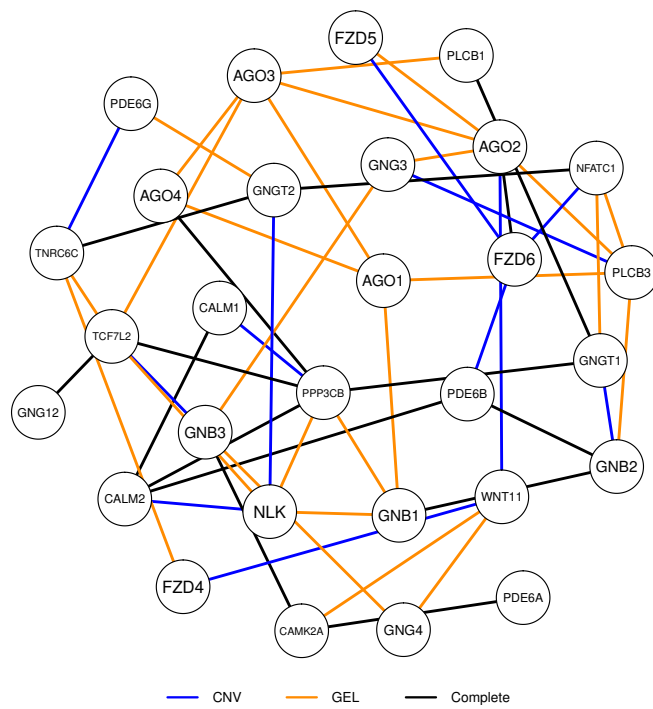


FIGURE C.20: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.1$.

Binding and Update of Ligands by Scavenger Receptors

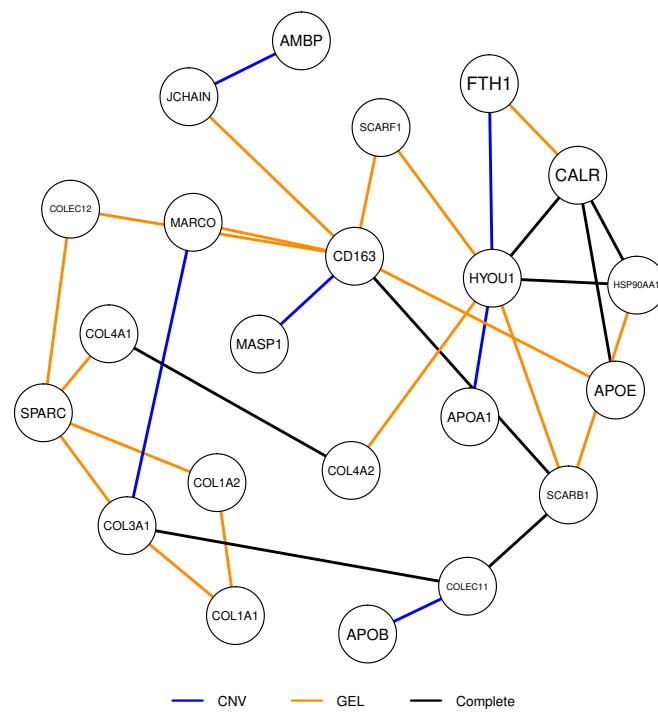


FIGURE C.21: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.001$.

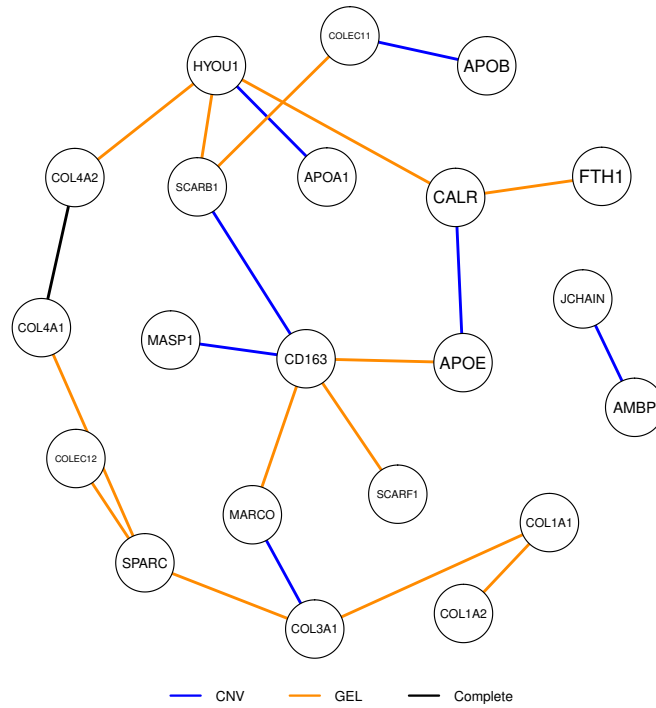


FIGURE C.24: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.001$.

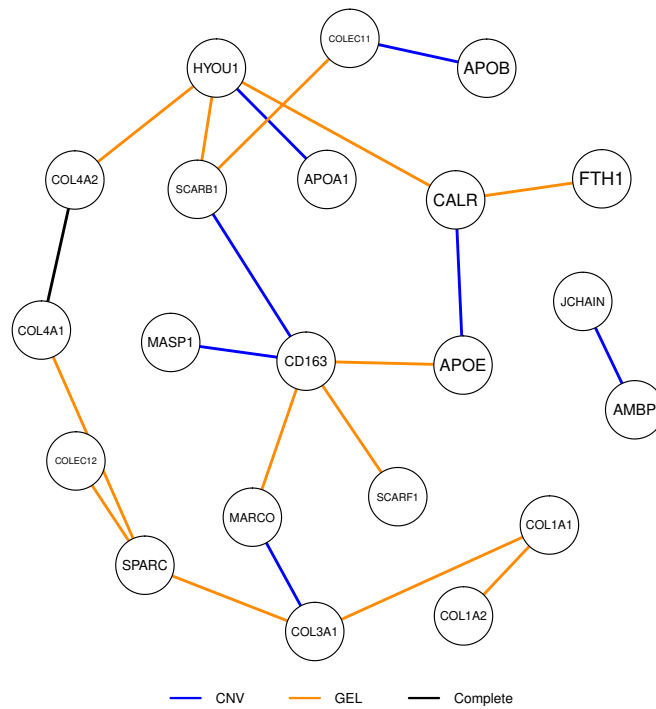


FIGURE C.25: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.001$.

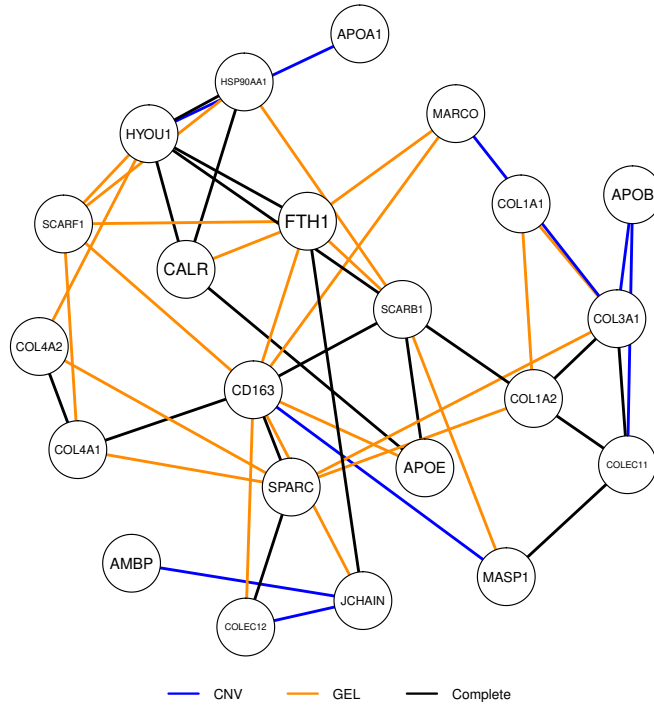


FIGURE C.26: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.01$.

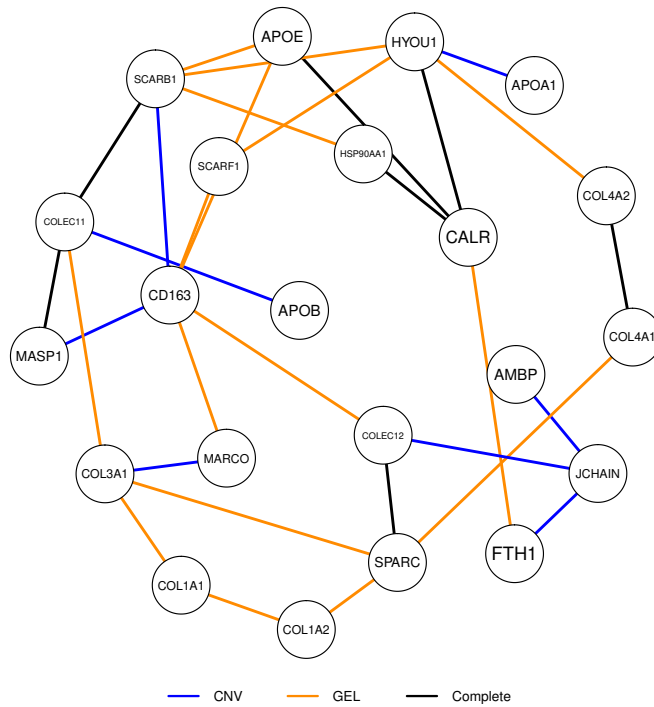
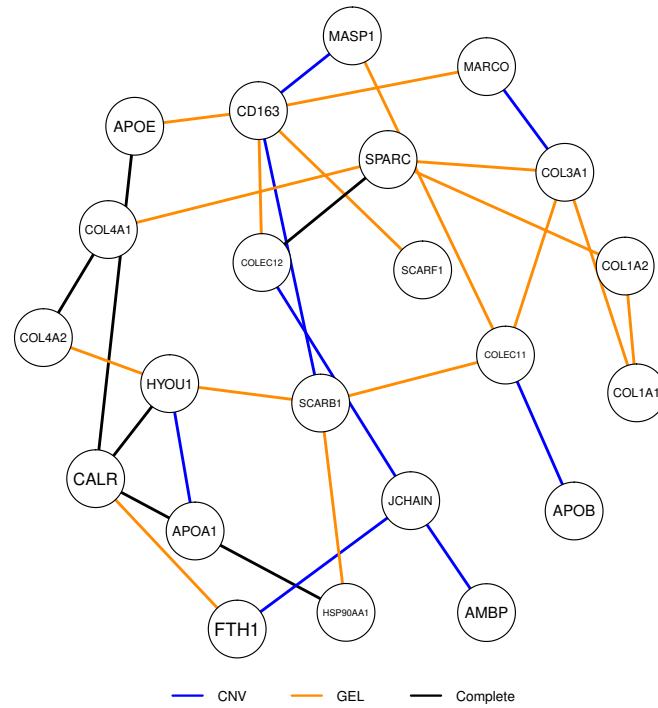
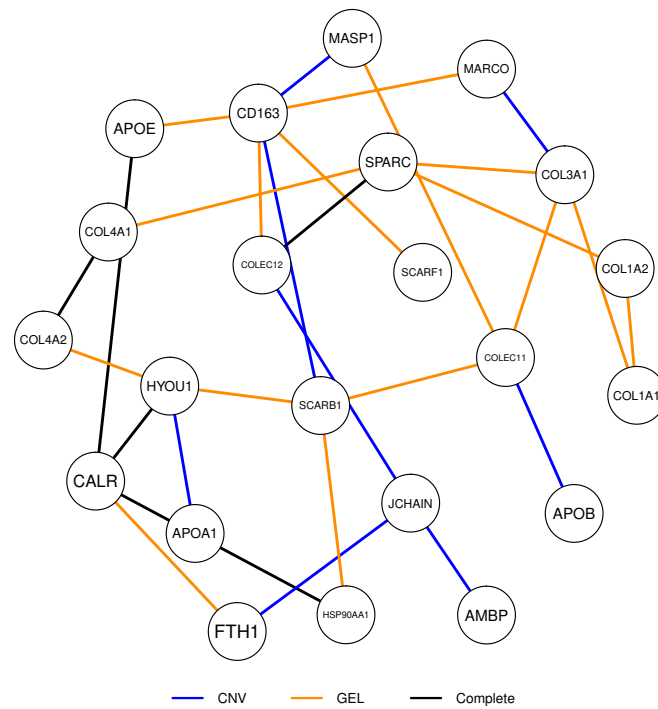


FIGURE C.27: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.01$.

FIGURE C.28: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.01$.FIGURE C.29: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.01$.

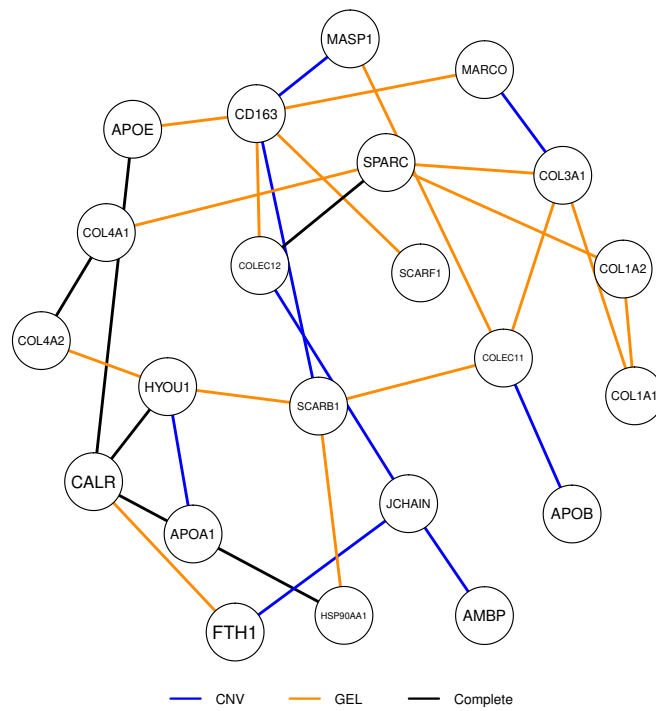


FIGURE C.30: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.01$.

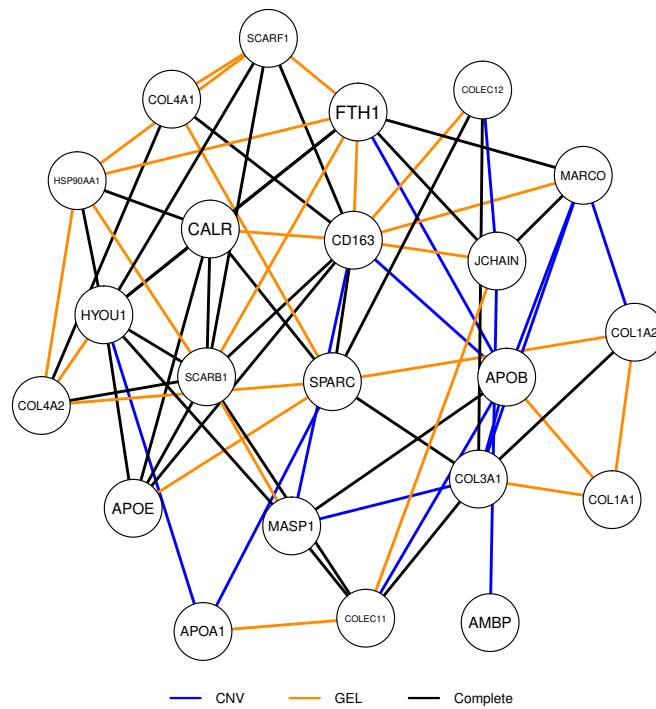


FIGURE C.31: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.05$.

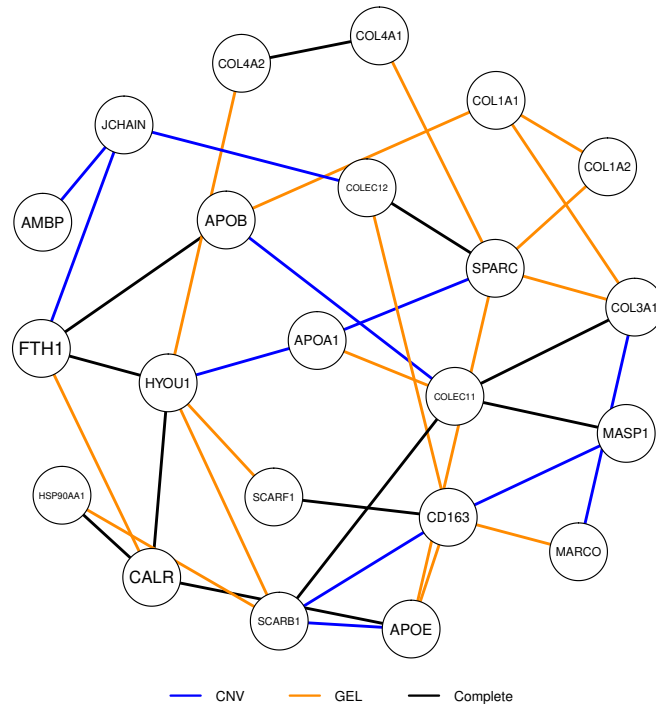


FIGURE C.32: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.05$.

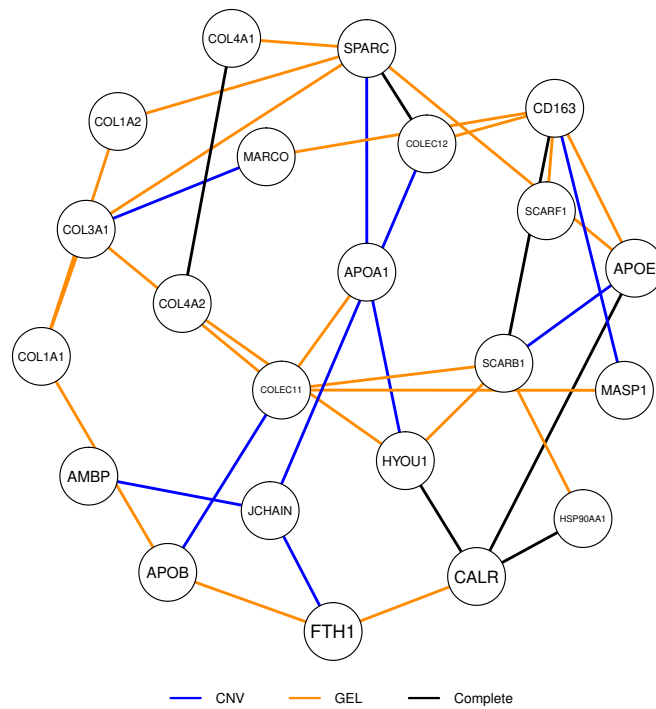


FIGURE C.33: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.05$.

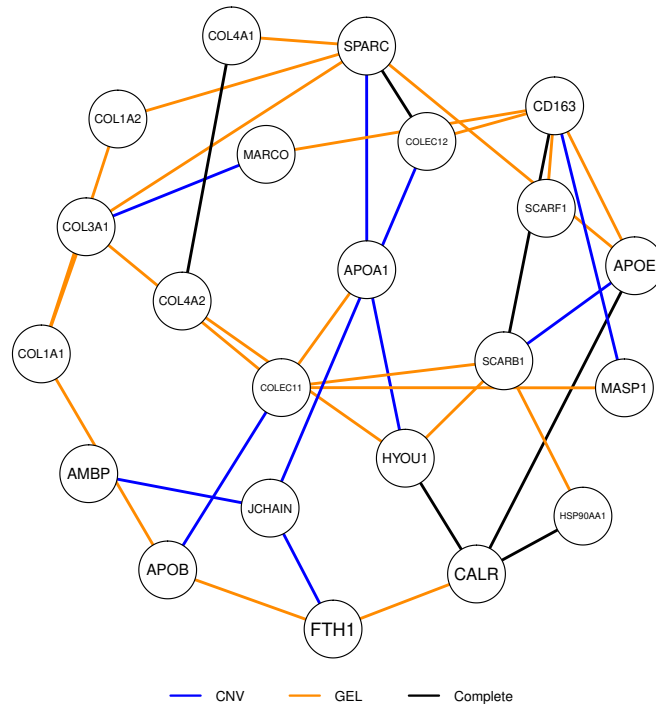


FIGURE C.34: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.05$.

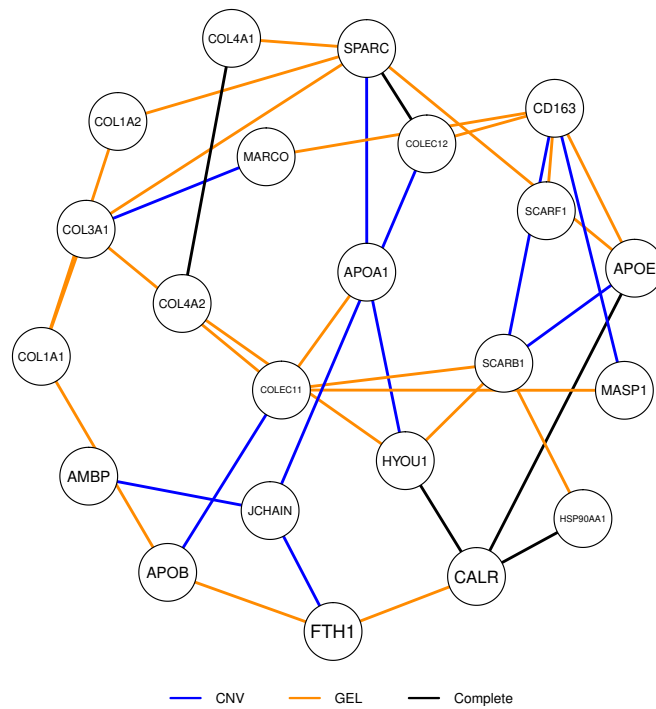
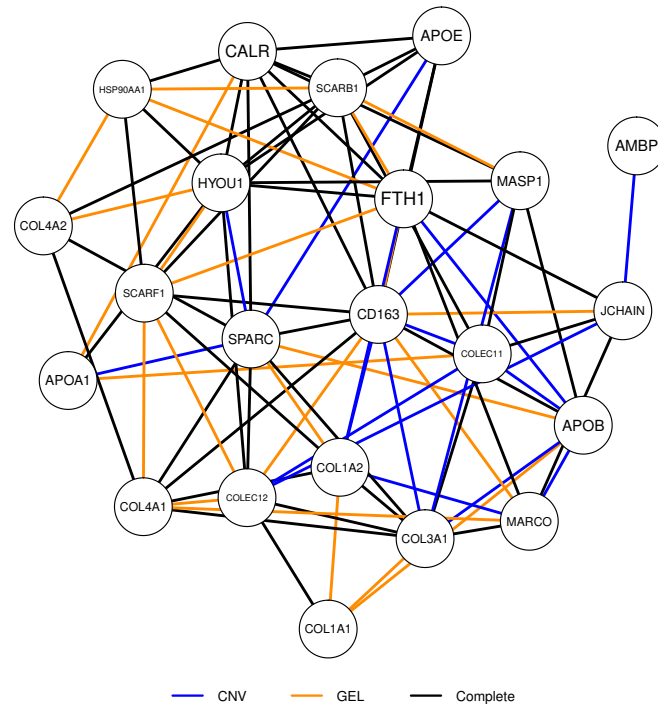
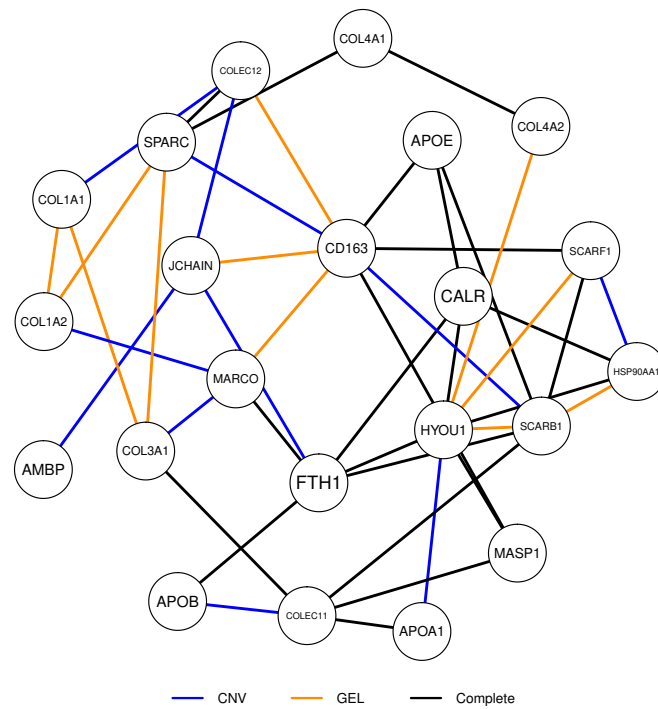


FIGURE C.35: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.05$.

FIGURE C.36: Estimated undirected multilayer network with $m = 1$ and $\alpha = 0.1$.FIGURE C.37: Estimated undirected multilayer network with $m = 2$ and $\alpha = 0.1$.

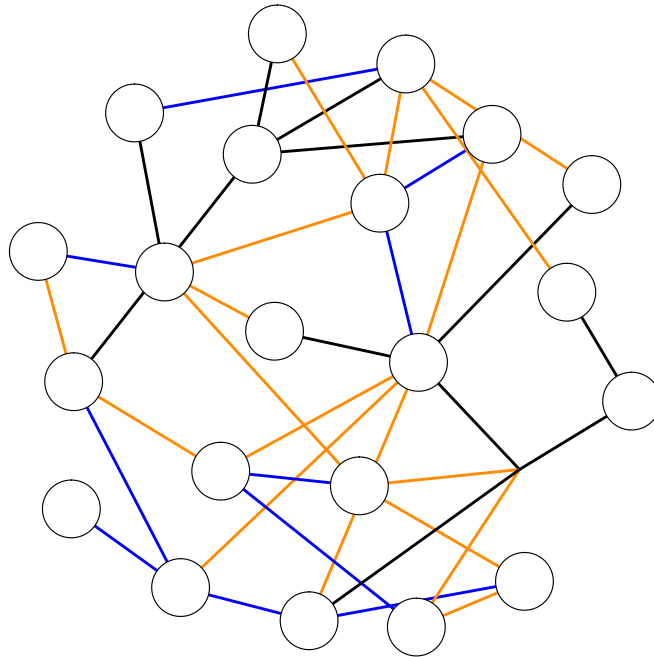


FIGURE C.38: Estimated undirected multilayer network with $m = 3$ and $\alpha = 0.1$.

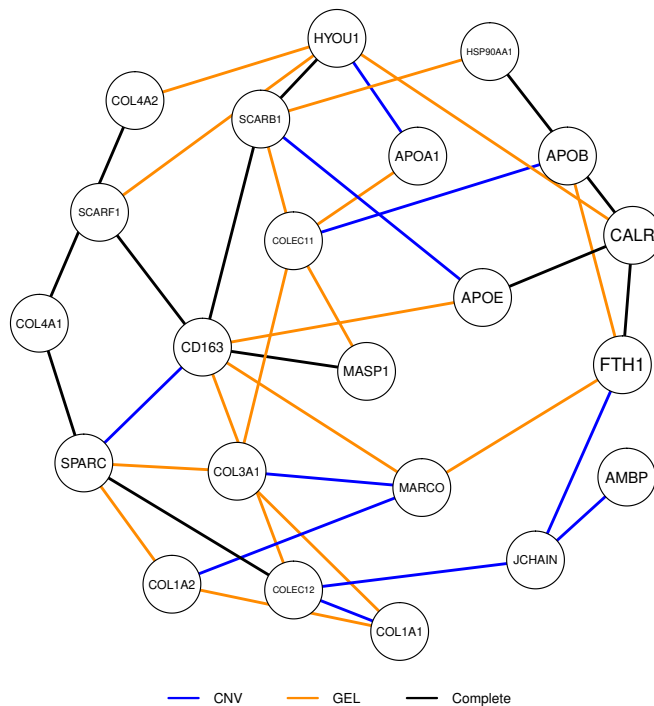
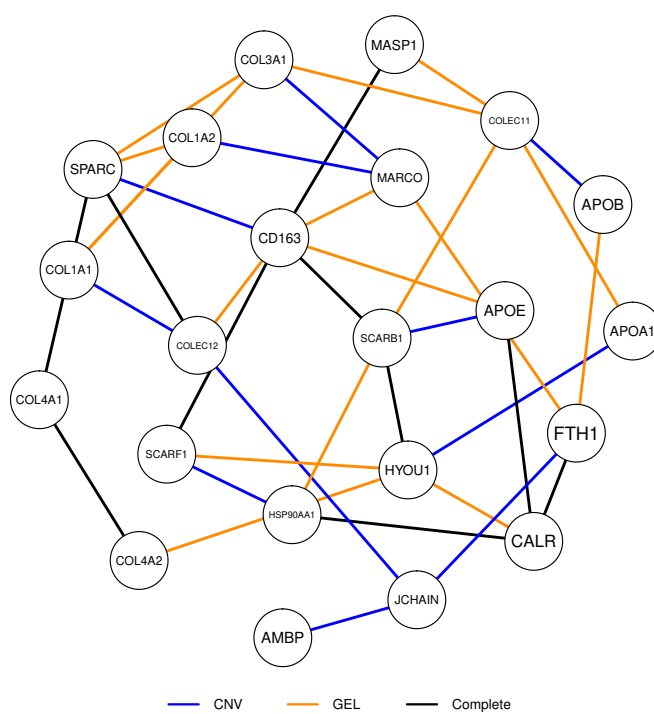


FIGURE C.39: Estimated undirected multilayer network with $m = 4$ and $\alpha = 0.1$.

FIGURE C.40: Estimated undirected multilayer network with $m = 5$ and $\alpha = 0.1$.

Bibliography

- Barabási, A. L. (2016) *Network Science*. Cambridge, UK: Cambridge University Press.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful approach to Multiple Testing. *Journal of the Royal Statistical Society B* **57**, 289–300.
- Besag, J. (1975) Statistical Analysis of Non-Lattice Data. *Journal of the Royal Statistical Society D* **24**, 179–195.
- Bhadra, A., Rao, A. and Baladandayuthapani, V. (2018) Inferring network structure in non-normal and mixed discrete-continuous genomic data. *Biometrics* **74**, 185–195.
- Bondy, J. A. and Murty, U. S. R. (2008) *Graph Theory*. London, UK: Springer.
- Brown, A. L., Li, M., Goncarenco, A. and Panchenko, A. R. (2019) Finding driver mutations in cancer: Elucidating the role of background mutational processes. *PLOS Computational Biology* **15**, 1–25.
- Chen, S., Witten, D. M. and Shojaie, A. (2015) Selection and estimation for mixed graphical models. *Biometrika* **102**, 47–64.
- Cheng, J., Li, T., Levina, E. and Zhu, J. (2017) High-Dimensional Mixed Graphical Models. *Journal of Computational and Graphical Statistics* **26**, 367–378.
- Chickering, D. M. (2002) Optimal Structure Identification With Greedy Search. *Journal of Machine Learning Research* **3**, 507–554.
- Colombo, D. and Maathuis, M. H. (2014) Order-Independent Constraint-Based Causal Structure Learning. *Journal of Machine Learning Research* **15**, 3741–3782.
- Desai, A., Xu, J., Aysola, K., Qin, Y., Okoli, C., Hariprasad, R., Chinemerem, U., Gates, C., Reddy, A., Danner, O., Franklin, G., Ngozi, A., Cantuaria, G., Singh, K., Grizzle, W., Landen, C., Partridge, E. E., Rice, V. M., Reddy, E. S. P. and Rao, V. N. (2014) Epithelial ovarian cancer: An overview. *World Journal of Translational Medicine* **3**, 1–8.

- Di Leva, G. and Croce, C. M. (2013) The Role of microRNAs in the Tumorigenesis of Ovarian Cancer. *Frontiers in Oncology* **3**, 1–10.
- Dimitrova, T. and Kocarev, L. (2018) Graphical Models Over Heterogeneous Domains and for Multilevel Networks. *IEEE Access* **6**, 69682–69701.
- Dobra, A. and Lenkoski, A. (2011) Copula Gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics* **5**, 969–993.
- Edwards, D. (1990) Hierarchical interaction models. *Journal of the Royal Statistical Society B* **52**, 3–20.
- Edwards, D., de Abreu, G. C. and Labouriau, R. (2010) Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests. *BMC Bioinformatics* **11**, 18.
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. and Borsboom, D. (2012) qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software* **48**, 1–18.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H. and D’Eustachio, P. (2018) The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **46**, D649–D655.
- Fan, J., Liu, H., Ning, Y. and Zou, H. (2016) High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society B* **79**, 405–421.
- Fellinghauer, B., Bühlmann, P., Ryffel, M., von Rhein, M. and Reinhardt, J. D. (2013) Stable graphical model estimation with Random Forests for discrete continuous and mixed variables. *Computational Statistics & Data Analysis* **64**, 132–152.
- Gkolfinopoulos, S. and Mountzios, G. (2018) Beyond EGFR and ALK: targeting rare mutations in advanced non-small cell lung cancer. *Annals of Translational Medicine* **6**, 1–12.
- Griffin, P. J., Zhang, Y., E.Johnson, W. and Kolaczyk, E. D. (2018) Detection of Multiple Perturbations in Multi-omics Biological Networks. *Biometrics* **74**, 1351–1361.

- Haslbeck, J. M. B. and Waldorp, L. J. (2015) Structure estimation for mixed graphical models in high-dimensional data. Available at: <https://arxiv.org/abs/1510.05677>.
- Huang, L., Jin, Y., Feng, S., Zou, Y., Xu, S., Qiu, S., Li, L. and Zheng, J. (2016) Role of Wnt/beta-catenin, Wnt/c-Jun N-terminal kinase and Wnt/Ca²⁺ pathways in cisplatin-induced chemoresistance in ovarian cancer. *Experimental and Therapeutic Medicine* **12**, 3851–3858.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Ole's, A. K., Pag'es, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L. and Morgan, M. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12**, 115–121.
- Hyter, S., Hirst, J., Pathak, H., Pessetto, Z. Y., Koestler, D. C., Raghavan, R., Pei, D. and Godwin, A. K. (2018) Developing a genetic signature to predict drug response in ovarian cancer. *Oncotarget* **9**, 14828–14848.
- Istituto Europeo di Oncologia (2019) *Ovarian Tumors*. <https://www.ieo.it/en/Patient-Care/Medical-Care/Ovarian-Tumors/> (last accessed: 25.09.2019).
- Kanehisa, M. (2019) Toward understanding the origin and evolution of cellular organisms. *Protein Science* **29**, 1–5. In Press.
- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019) New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**, D590–D595.
- Katabuchi, H. and Okamura, H. (2003) Cell biology of human ovarian surface epithelial cells and ovarian carcinogenesis. *Medical Electron Microscopy* **36**, 74–86.
- KEGG (2019) *DISEASE: Ovarian cancer*. https://www.genome.jp/dbget-bin/www_bget?ds:H00027 (last accessed: 25.09.2019).
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. and Porter, M. A. (2014) Multilayer Networks. *Journal of Complex Networks* **2**, 203–271.

- Lauritzen, S. L. (1996) *Graphical Models*. Oxford, UK: Clarendon Press.
- Lauritzen, S. L. and Wermuth, N. (1989) Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative. *Annal of Statistics* **17**, 31–57.
- Lederer, J. (2016) Graphical Models for Discrete and Continuous Data. Available at: <https://arxiv.org/pdf/1609.05551.pdf>.
- Lee, J. D. and Hastie, T. J. (2015) Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics* **24**, 230–253.
- Li, X., Zhang, N. X., Ye, H. Y., Song, P. P., Chang, W., Chen, L., Wang, Z., Zhang, L. and Wang, N. N. (2019) HYOU1 promotes cell growth and metastasis via activating PI3K/AKT signaling in epithelial ovarian cancer and predicts poor prognosis. *European Review for Medical and Pharmacological Sciences* **23**, 4126–4135.
- Li, Z. R., McCormick, T. H. and Clark, S. J. (2017) Using Bayesian latent Gaussian graphical models to infer symptom associations in verbal autopsies. Available at: <https://arxiv.org/abs/1711.00877>.
- Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012) HIGH-dimensional semiparametric Gaussian copula graphical models. *Annals of Statistics* **40**, 2293–2326.
- Liu, H., Lafferty, J. and Wasserman, L. (2009) The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research* **10**, 2295–2328.
- Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., Kovatich, A. J., Benz, C. C., Levine, D. A., Lee, A. V., Omberg, L., Wolf, D. M., Shriver, C. D. and Thorsson, V. (2018) An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400–416.
- Martini, P., Chiogna, M., Calura, E. and Romualdi, C. (2019) MOSClip: multi-omic and survival pathway analysis for the identification of survival associated gene and modules. *Nucleic Acids Research* **47**, e80.1–e80.13.
- Marx, V. (2016) Cancer: hunting rare somatic mutations. *Nature Methods* **13**, 295–299.
- National Cancer Institute and the National Human Genome Research Institute (2019) *Cancer Genome Atlas (TCGA)*. <https://www.cancer.gov/tcga> (last accessed: 25.09.2019).

- NIH - GDC Data Portal (2019) *TCGA-OV*. <https://portal.gdc.cancer.gov/projects/TCGA-OV> (last accessed: 25.09.2019).
- Novakowski, K. E., Huynh, A., Han, S., Dorrington, M. G., Yin, C., Tu, Z., Pelka, P., Whyte, P., Guarné, A., Sakamoto, K. and Bowdish, D. M. (2016) A naturally occurring transcript variant of MARCO reveals the SRCR domain is critical for function. *Immunology and Cell Biology* **94**, 646–655.
- Pearl, J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Francisco, CA, USA: Morgan Kaufmann.
- Pearl, J. and Paz, A. (1987) GRAPHOIDS: A graph-based logic for reasoning about relevance relations. In B. Du Boulay, D. Hogg, and L. Steels (Eds.). *Advances in Artificial Intelligence* **2**, 357–363. North-Holland, Amsterdam.
- PrabhuDas, M. R., Baldwin, C. L., Bollyky, P. L., Bowdish, D. M. E., Drickamer, K., Febbraio, M., Herz, J., Kobzik, L., Krieger, M., Loike, J., McVicker, B., Means, T. K., Moestrup, S. K., Post, S. R., Sawamura, T., Silverstein, S., Speth, R. C., Telfer, J. C., Thiele, G. M., Wang, X. Y., Wright, S. D. and Khoury, J. E. (2017) A Consensus Definitive Classification of Scavenger Receptors and Their Roles in Health and Disease. *Journal of Immunology* **198**, 3775–3789.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raghu, V. K., Poon, A. and Benos, P. V. (2018) Evaluation of Causal Structure Learning Methods on Mixed Data Types. *Proceedings of machine learning research* **92**, 48–65.
- Ricken, A., Lochhead, P., Kontogiannia, M. and Farookhi, R. (2002) Wnt signaling in the ovary: identification and compartmentalized expression of wnt-2, wnt-2b, and frizzled-4 mRNAs. *Endocrinology* **143**, 2741–2749.
- Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B* **71**, 319–392.
- Sales, G., Calura, E., Cavalieri, D. and Romualdi, C. (2012) graphite: a Bioconductor package to convert pathway topology to gene network. *BMC Bioinformatics* **13**, 1–12.
- Sales, G., Calura, E. and Romualdi, C. (2018) metaGraphite: a new layer of pathway annotation to get metabolite networks. *Bioinformatics* **35**, 1258–1260.

- Scholl, C. and Fröhsling, S. (2019) Exploiting rare driver mutations for precision cancer medicine. *Current Opinion in Genetics and Development* **54**, 1–6.
- Sedgewick, A. J., Ramsey, J. D., Spirtes, P., Glymour, C. and Benos, P. V. (2017) Mixed Graphical Models for Causal Analysis of Multi-modal Variables. Available at: <https://arxiv.org/abs/1704.02621>.
- Sedgewick, A. J., Shi, I., Donovan, R. M. and Benos, P. (2016) Learning mixed graphical models with separate sparsity parameters and stability-based model selection. *BMC Bioinformatics* **17**, Suppl 5, 175.
- Shih, I. and Kurman, R. J. (2004) Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *American Journal of Pathology* **164**, 1511–1518.
- Spirtes, P., Glymour, C. N. and Scheines, R. (2000) *Causation, Prediction, and Search*. Cambridge, MA, US: MIT Press.
- Sugimura, R. and Li, L. (2010) Noncanonical Wnt signaling in vertebrate development, stem cells, and diseases. *Birth Defects Research* **90**, 243–256.
- Tansey, W., Padilla, O. H. M., Suggala, A. S. and Ravikumar, P. (2015) Vector-Space Markov Random Fields via Exponential Families. In *ICML*. Available at: <http://proceedings.mlr.press/v37/tansey15.pdf>.
- Tsagris, M., Borboudakis, G., Lagani, V. and Tsamardinos, I. (2018) Constraint-based causal discovery with mixed data. *International Journal of Data Science and Analytics* **6**, 19–30.
- Tur, I. and Castelo, R. (2011) Learning mixed graphical models from data with p larger than n . *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence* pp. 389–697.
- Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. New York, US: Springer.
- Wainwright, M. J. and Jordan, M. I. (2008) Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning* **1**, 1–305.
- Wenham, R. M., Lancaster, J. M. and Berchuck, A. (2002) Molecular aspects of ovarian cancer. *Best Practice and Research. Clinical Obstetrics and Gynaecology* **16**, 483–497.

- Xue, L. and Zou, H. (2012) Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Annals of Statistics* **40**, 2541–2571.
- Yang, E., Baker, Y., Ravikumar, P., Allen, G. I. and Liu, Z. (2014a) Mixed Graphical Models via Exponential Families. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*. Available at: <http://proceedings.mlr.press/v33/yang14a.html>.
- Yang, E., Ravikumar, P., Allen, G. I., Baker, Y., Wan, Y. W. and Liu, Z. (2014b) A General Framework for Mixed Graphical Models. Available at: <https://arxiv.org/abs/1411.0288>.
- Yang, E., Ravikumar, P., Allen, G. I. and Liu, Z. (2012) Graphical Models via Generalized Linear Models. In *Advances in Neural Information Processing Systems 25*. Available at: <http://papers.nips.cc/paper/4617-graphical-models-via-generalized-linear-models.pdf>.
- Yang, E., Ravikumar, P., Allen, G. I. and Liu, Z. (2013) Conditional Random Fields via Univariate Exponential Families. In *Advances in Neural Information Processing Systems 26*. Available at: <http://papers.nips.cc/paper/5154-conditional-random-fields-via-univariate-exponential-families.pdf>.
- Yang, E., Ravikumar, P., Allen, G. I. and Liu, Z. (2015) Graphical models via univariate exponential family distributions. *Journal of Machine Learning Research* **16**, 3813–3847.
- Yang, Z., Ning, Y. and Liu, H. (2014c) On semiparametric exponential family graphical models. Available at: <https://arxiv.org/abs/1412.8697>.
- Zhang, Y., and G. Wu, C. P., Wang, Y., Liu, R., Yang, S., He, S., He, F., Yuan, Q., Huang, Y., Shen, A. and Cheng, C. (2011) Expression of NLK and its potential effect in ovarian cancer chemotherapy. *International Journal of Gynecological Cancer* **21**, 1380–1387.

Ilaria Bussoli

CURRICULUM VITAE

Contact Information

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

Tel. +39 049 827 4111
e-mail: bussoli@stat.unipd.it

Current Position

Since October 2016; (expected completion: October 2019 / March 2020)

PhD Student in Statistical Sciences, University of Padova.

Thesis title: "Heterogeneous Graphical Models with Applications to Omics Data"

Supervisor: Prof. Monica Chiogna

Research interests

- Biostatistics
- Graphical models
- Multi-omics integration

Education

September 2013 – September 2016

Master (laurea specialistica/magistrale) degree in Statistical Sciences.

University of Padova, School of Science

Title of dissertation: "Eventi di legame rari in dinamica molecolare: il caso del dominio SH2 e del peptide pYEEI"

Supervisor: Prof. Alessandra R. Brazzale

September 2010 – September 2013

Bachelor degree (laurea triennale) in Statistica e trattamento Informatico dei Dati (SMID).

University of Genova, School of Science

Title of dissertation: "Statistica e Sistemi Informativi Territoriali nella pianificazione e gestione di interventi della Protezione Civile sul territorio comunale"

Supervisor: Prof. Eva Riccomagno

Visiting periods

April 2018 – September 2018

Big Data Institute,

Oxford, England.

Supervisor: Prof. Sofia Massa

Awards and Scholarship

2016–2019

Three-year Ph.D. scholarship, University of Padova

Computer skills

- R, Python

Language skills

Italian: native; English: fluent.

Conference presentations

Bussoli, I., Massa, M.S., Chiogna, M. (2018). Analysing the structure of cancer data using graphical models. (poster) *44th Young Statistician Meeting*, Oxford, England, 29 - 31 July.

Bussoli, I., Chiogna, M. (2018). Multi-omics integrated analysis by means of graphical models. *Computational Methods in Statistics (ERCIM)*, Pisa, Italy, 16 - 16 December.

Teaching experience

October 2014 – June 2015

Calculus, Mathematical Statistics

Degree in Statistical Science

Teaching task (exercises/tutoring), 150 hours

Departments of Statistical Sciences, University of Padova

Instructor: Prof. Alessandra R. Brazzale

November 2018 year – January 2019

Statistical learning 1

Degree in Data Science

Teaching task (exercises/lab), 40 hours

Departments of Mathematics, University of Padova

Instructor: Prof. Alberto Roverato

References

Prof. Monica Chiogna

Department of Statistical Sciences

University of Bologna

Via Belle Arti, 41

40126 Bologna, Italy

e-mail: monica.chiogna2@unibo.it

Prof. Alessandra R. Brazzale

Department of Statistical Sciences

University of Padova

Via Cesare Battisti, 241

35121 Padova, Italy

e-mail: brazzale@stat.unipd.it

