Sapienza, University of Rome
Department of Statistical Sciences

Ph.D. in Methodological Statistics

Latent space models for multidimensional network data

Candidate:
Silvia D'Angelo

Thesis advisor:
Prof. Marco Alfò

Thesis submitted in 2018

Alla pizza napoletana, per aver contribuito a rendermi ciò che sono oggi.

Pizza brought me so far, but does not get all the credits.
During my master's degree, Alessandra, Valentina, Giulia and Alessia made my stay at the fourth floor so nice that me and Tullia decided not to leave for another three years. In all these years of friendship, they still have not understood why we did it. These last years would not have been the same without Tullia, Carlo and Marco. We have shared a lot, especially Stanza 41. This room taught us much about coexistence and helped us becoming friends. With them, Lampros, Riccardo, Francesca, Julianne, Brendan and all the others, from Rome to Dublin, contributed making PhD life a great adventure. My experience could have not been the same without Marco, who let me work freely, and never neglected me. There is much in life outside Statistics and I am happy to have found a supervisor who stands by that.
Laura, Elena, Giorgia and Caterina, friends of a lifetime, accompanied me through this PhD, as they did with all other episodes in my life before this one and will do with next to come.
This thesis is dedicated to my parents, Antonietta and Paolo, for giving me an inspiring life and for all the journeys we made together. You always wanted what was best for me, and met me halfway all the times our definitions of best did not match. Alessandro, my brother, this thesis is also for you, for turning the volume so loud while playing videogames that even people in Ariccia could hear you and for caring so much about me to come and visit me in Dublin, even though is Dublin.
Last, this is for you Michael, and for us. We know why.

Contents

List of Figures

List of Tables

Abstract

Network data are any relational data recorded among a group of individuals, the nodes. When multiple relations are recorded among the same set of nodes, a more complex object arises, which we refer to as "multidimensional network", or "multiplex", where different relations corresponding to different networks.
In the past, statistical analysis of networks has mainly focused on single-relation network data, referring to a single relation of interest. Only in recent years statistical models specifically tailored for multiplex data begun to be developed. In this context, only a few works have been introduced in the literature with the aim at extending the latent space modeling framework to multiplex data. Such framework postulates that nodes may be characterized by latent positions in a p-dimensional Euclidean space and that the presence/absence of an edge between any two nodes depends on such positions. When considering multidimensional network data, latent space models can help capture the associations between the nodes and summarize the observed structure in the different networks composing a multiplex.
This dissertation discusses some latent space models for multidimensional network data, to account for different features that observed multiplex data may present. A first proposal allows to jointly represent the different networks into a single latent space, so that average similarities between the nodes may be captured as proximities in such space. A second work introduces a class of latent space models with node-specific effects, in order to deal with different degrees of heterogeneity within and between networks in multiplex data, corresponding to different types of node-specific behaviours. A third work addresses the issue of clustering of the nodes in the latent space, a frequently observed feature in many real world network and multidimensional network data. Here, clusters of nodes in the latent space correspond to communities of nodes in the multiplex.
The proposed models are illustrated both via simulation studies and real world applications, to study their perfomances and abilities.

Acknowledgments

A special thank goes to the reviewers, Veronica Vinciotti and Giuliano Galimberti, for their precious comments and feedback.

Chapter 1

Introduction

In the last decades, statisticians showed an increasing interest raised in the analysis of network data, and a large variety of models has been developed for this task. In the beginning, network analysis was deeply linked with social network analysis, as first network data were collected and analysed by quantitative social scientists (Jacob Moreno and Helen Jennings, 1930). Nowadays, network data come from many other different fields: biology, telecommunications, economic and behavioural sciences, and so on. While first models focused mainly on studying global network properties and network summary statistics behaviour, more recent ones have also migrated to the analysis of single relationships. Indeed, many modelling techniques have been proposed: from visualization to link prediction, a lot of different approaches have been developed. In particular, recent developments in network analysis focused on feasible representations of the complex dependency structure present in network data, by means of latent variables. Also, in the last years statisticians started to focus on more complex types of network data, as for example networks arising from different measurements or measurements repeated trough time on the same sample of units. Such data are called multidimensional network data, and their study is a flourishing area in network analysis.
The purpose of this chapter is to introduce the reader to the main concepts in network analysis and, in particular, in latent variable models for network data. First, basic notation that will be used throughout the dissertation will be introduced, together with a description of some well known and widely used exploratory and summary statistics to characterize observed networks. Then, we will introduce some of the first approaches proposed in the literature to network data, that have paved the way to modern network analysis. Last, we will discuss latent space models for network data, their principal forms and main assumptions. The final part of this chapter will give a brief summary of the dissertation structure.

1.1 Single and multidimensional networks

Any relational data observed for a group of units form a network. As networks can be represented by means of graph theory, units in the data are often referred to as *nodes*. Any relation that exists between any two nodes is represented by an *edge* linking the pair (referred to as a *dyad*). More formally, we may represent a network

as the realization of a random graph G:

$$G = (V, E),$$

where V is the set of the nodes and E that of the edges. We will denote the cardinality of V by n, which is the number of nodes, indexed by $i, j = 1, \ldots, n$. A general edge may be denoted by e_{ij}.
Depending on the type of relation being recorded, the set of edges may be characterized in different ways. One major distinction is whether the edges are valued or not:

- **Binary** edges appear when the relation that is being measured can either be present or not. Therefore, an edge is present between two nodes if the relation is present, otherwise it is not. A typical example of binary edges could be that of a network where a group of people, the nodes, are asked to state whether or not they consider other units in the sample as friend; a declaration of friendship would correspond to an edge. A network having binary edges is called *binary network* and may be represented via an adjacency matrix $\mathbf{Y}$, with dimension $n \times n$ and general element

$$y_{i,j} = \begin{cases} 1 & \text{if an edge exists between node } i \text{ and } j, \\ 0 & \text{else.} \end{cases}$$

- **Weighted** edges, instead, arise when the relation being recorded is valued. In this context, what is of interest is the strength of the relation between the nodes, represented by the values associated with the edges. Depending on what is being measured, the weights may be discrete or continuous, positive, negative or both. A typical example would be that of telephone networks. Suppose that, for the same group of people in the previous example, we recorded the number of calls between any two of them in a given period of time. Here, the strength of the relation between two nodes is represented by the number of phone calls they made: the higher, the grater. Of course, if a dyad (i, j) does not have any phone call, the weight associated with the corresponding edge is 0. A network with weighted edges is called *weighted* network.

The presence of an edge between a dyad (a couple of nodes), or the potential weight associated with it, does not tell the whole story. Indeed, it may be relevant to check whether the presence of an edge between nodes (i, j) is accompanied with the presence of an edge between (j, i). Indeed, edges may have a direction, as not all relations are reciprocal, and we can distinguish them in:

- **Undirected** edges, when the relation is reciprocal. For example, if one was recording kinships among a group of people in a village, the fact that person i is related with j implies that also j is related with i. This reciprocity is called *symmetry* and network exhibiting such characteristic are called *symmetric* networks.
  Similar to *symmetry* is the *reflexivity* property. This indicates a node that interacts with itself, that is, $y_{ii} = 1$ in binary networks or a non null weight

associated with e_{ii} in weighted networks. Such a property may be observed in some biological networks, for example in protein-protein interaction networks, but is quite rare in social network data.

- **Directed** edges represent relations which potentially have a direction. For example, a friendship network may have directed edges, as the fact that node i thinks that j is a friend does not imply the opposite. A network with directed edges is called *asymmetrical*.
  For asymmetrical networks, it may be of interest to consider the number of mutual edges between dyads, that is $\sum_{i<j} y_{ij}y_{ji}$. The presence of an high quantity of mutual links in asymmetrical networks would denote that most of the dyads exhibit some reciprocity, on the opposite, a low mutuality would denote a low value, or the absence, of reciprocity.

From this moment on, in this introduction, we will focus on binary networks, as the present thesis develops model for such type of network data. For a richer review of weighted networks, see for example Wasserman and Faust, 1994.
Note that, potentially, a network with n nodes may have a maximum number of edges equal to n^2, if the network is *reflexive*, or a maximum of $n(n-1)$ edges if it is not. A non-reflexive network can be either

$$\begin{cases} \textbf{dense or complete} & \text{if} \quad \sum_{i=1}^{n}\sum_{j\leq i}^{n} y_{ij} = n(n-1), \\ \textbf{sparse} & \text{if} \quad \sum_{i=1}^{n}\sum_{j\leq i}^{n} y_{ij} < n(n-1). \end{cases}$$

However, complete networks are quite hard to encounter in practice and, therefore, are usually of low interest. In general, a binary network may be called dense if almost all the entries of the corresponding adjacency matrix are 1 and sparse when, on the opposite, almost all of them are 0. An index of sparsity is the *density*, which computes the ratio between the number of observed edges and the maximum possible number of edges in a network; values of this index close to 0 denote sparsity, while values close to 1 indicate a dense network.
At a more local level, the number of edges corresponding to each node is a first indicator of how nodes interact with each other; this behaviour may be captured by looking at the nodes's*degree*. The degree $d(i)$ of node i basically counts how many edges are incident with it. In symmetric networks, it may be defined equivalently either looking at the columns or at the rows of the network adjacency matrix $\mathbf{Y}$:

$$d(i) = \sum_{j=1}^{n} y_{ij} = \sum_{j=1}^{n} y_{ji}.$$

Instead, in asymmetric matrix, rows and columns convey different informations regarding the behaviour of a node. By convention, we will state that a given row i collects edges going out of node i, edges that node i is *sending*, while column i collects edges entering node i, which the node is *receiving*. Then, we might define the out-degree $d_{\text{out}}(i)$ and the in-degree $d_{\text{in}}(i)$ of node i as:

$$d_{\text{out}}(i) = \sum_{j=1}^{n} y_{ij}, \quad d_{\text{in}}(i) = \sum_{j=1}^{n} y_{ji}.$$

Different network data may have different degree distributions $d_{\text{in}}(i), d_{\text{out}}(i)$, $i = 1, \ldots, n$, which can be used to characterize the network. For example, some networks may have homogeneous degree distributions, with all degrees having similar values. Other networks may have a small group of nodes with an high degree and the rest with only a few connections.

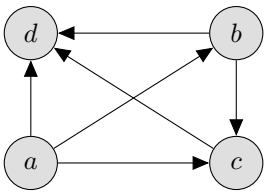


Figure 1.1. Graphical example of asymmetrical binary network, with 4 nodes and 5 edges.

Figure 1.1 provides an example of a network, which has $n = 4$ nodes and 6 directed edges. In this example, the in-degrees and out-degrees distributions are mostly inhomogeneous, with $d_{\text{out}}(a) = 3$, $d_{\text{out}}(b) = 2$, $d_{\text{out}}(c) = 1$, $d_{\text{out}}(d) = 0$ and $d_{\text{in}}(a) = 0$, $d_{\text{in}}(b) = 1$, $d_{\text{in}}(c) = 2$, $d_{\text{in}}(d) = 3$. Node a is the most active of the 4, with an out-degree of 3, but it is not popular at all, as the in-degree is null. The concepts of *popularity* and of *activity*, referred to the propensity of nodes to receive/send many links, will be further discussed in the next section and in Chapter 2, as a way to characterize the individual nodes behaviours.
Notice that one may not only look at dyadic relations, but at higher order ones as well; triangles, for example, are configurations arising when a group of three nodes (i, j, l) are all interacting with one another. The number and position of the edges may also be used to define other summary statistics, as for example *centrality* statistics, denoting the importance of each node in a network, in terms of its relevance in determining the observed structure. For different definitions of centrality, network statistics and other basic characterization of network data we refer to Salter-Townshend et al., 2012 and Wasserman and Faust, 1994.

1.1.1 Multidimensional networks

The type of network data we have defined in the previous section is the most simple one, where a single relation is recorded among a same group of nodes. However, sometimes it may be the case that multiple relations are recorded among the same group of nodes. Then, the corresponding network is referred to as *multidimensional network* or *multiplex*. A binary multidimensional network may be represented by a collection of adjacency matrices $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}\}$, with $k = 1, \ldots, K$. The index k denotes the k^{th} relation recorded on the n nodes, represented by the k^{th} network, or view. The k^{th} adjacency matrix has general entry

$$y_{i,j}^{(k)} = \begin{cases} 1 & \text{if an edge exists between node } i \text{ and } j \text{ in the } k^{\text{th}} \text{ network,} \\ 0 & \text{if not.} \end{cases}$$

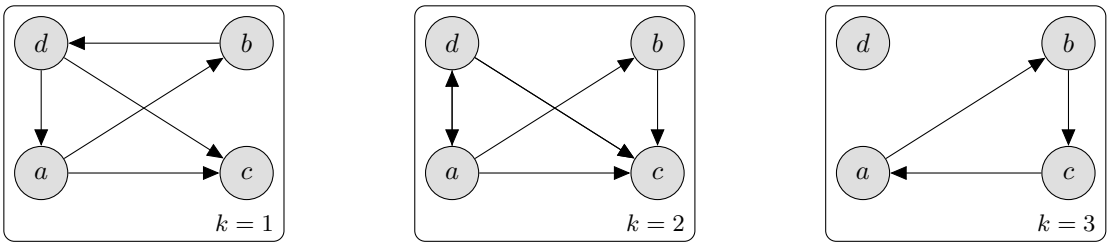


Figure 1.2. Graphical example of an asymmetrical binary multidimensional network, with 4 nodes and 3 views.

The summary statistics defined in the previous section for single network data may be computed for the views of a multiplex as well. Also, the same characterization in terms of types of edges and relations, binary, weighted, etc., still holds. Figure 1.2 represents, as an example, an asymmetrical binary multidimensional network, with 4 nodes and 3 views. These 3 networks could correspond, for example, to data collected among 4 people working together, who were asked whether they were friends ($k = 1$), if they worked well together ($k = 2$) and if they considered another person reliable ($k = 3$). As it can be seen from the picture, different relations correspond to different configurations in the views. Thus, multidimensional networks convey a variable information regarding the nodes of interest. The different views, when analysed all together, may bring interesting insights on the relational structure among the nodes, not captured by the analyses of the single networks alone. Indeed, in the previous example, if the research question was *how do the co-workers interact?*, considering jointly the different relations would give a richer picture on the interactions. Also, when the number of observed networks increases, analysing the multiplex as a single object may be a more parsimonious approach, when compared to K separate single network analyses.
A particular case of multidimensional network is represented by *dynamic* networks, which are defined by the recording of the same relation trough time, over the same set of nodes. One example could be that of a class of students, followed from the first year of studies until the last one, recording friendship relationships each year. As students spend more time together, new friendships may begin and old ones may end; analysing such an evolution may be of interest.

Until now, we have introduced some basics concepts and definitions regarding single and multidimensional network data. In the following section, we will discuss some of the main tasks arising when attempting to model such data. Further, we will briefly present the classical models developed for network data, their purposes, characteristics and limitations.

1.2 Models for network data

As mentioned before, network data record relations among a group of units, the nodes, and, for modelling purposes, an observed network may be assumed to

be the realization of a random graph. Hence, to better understand networks, one may aim at reconstructing the process that lead to their observed realizations. In other words, one may aim at modelling the edge probabilities p_{ij} underlying a given network $\mathbf{Y}$, that is, the probabilities of observing an edge between nodes (i, j), $i, j = 1, \ldots, n$. Do edges probabilities depend on the dyad or are they constant over the whole network? Is there a node that has a higher probability than others to send links? What do these probabilities depend on? These questions, and many others, drove the definition of different stochastic models for network data, with different aims and hypotheses. As in other branches of statistics, there is no "true" model to analyse network data; the choice among competing models mainly depends on the purposes of the analysis, and, of course, the data at hand. A factor that may determine the choice of a particular model is its computational complexity. Network data are high dimensional by nature, as, for example, even a "small", non reflexive, network with "only" $n = 50$ nodes, has potentially $50(50-1) = 2450$ edges; the computational burden increases quadratically with the number of nodes, which sometimes makes complex models unfeasible. Also, with multidimensional networks the computational complexity is additionally increased by the extra dimension given by the number of views. On the other hand, another task arising with network data is that such complex objects may be the realisations of highly dependent random variables. Therefore, simple models may prove unable to represent network data, and more sophisticated representations should be considered. This trade-off between faithful representation and feasible estimation is a key issue when dealing with network data. Below, we will provide a synthetic summary of the main classes of models developed in network analysis to model edge probabilities and briefly discuss them. We may consider:

- **Erdös-Rényi model**. Proposed by the homonymous authors in 1959 (Erdős and Rényi, 1959), this first model is the simplest of all, as it assumes independent and constant edge probabilities across the network. Although its simplicity makes it computationally feasible to fit and various theoretical results regarding it are available, its over simplistic assumptions make it often inappropriate to describe real world network data. However the Erdös-Rényi model is still a reference model in network analysis and sometimes it may be useful to compare different representations from more complex models.

- **p1 model**. Introduced by Holland and Leinhardt, 1981, this model still assumes independence between edges, but removes the assumption of constant edge probabilities. Indeed, it postulates that edge probabilities are function of some summary statistics (number of nodes, out-/in-degrees and number of mutual links between dyads) and four types of parameters. Such parameters capture the base rate for edge probabilities, the sender/receiver propensities of the nodes, and the level of mutuality in the network. The **p1** model introduces the concepts of sender/receiver propensities of the nodes, in this context denoted by productivity/attractiveness. Such characterizations have been later recovered and adapted in different other models, often proving to be quite useful to feasibly represent nodes individual behaviour in asymmetrical networks.

- **p2 model**. This model is an extension of the p1 model, proposed by Duijn, Snijders, and Zijlstra, 2004, where sender and receiver effects are modelled as additive random effects and covariates may be included in the specification of the model for the edge probabilities. The inclusion of covariates is a relevant step forward in network analysis, as it allows to explain an observed network using node-specific external information, other than with the summary statistics alone. Covariates account for *homophily* by attributes, that is, nodes with similar attributes may be more likely to link. However, the selection of covariates relevant to the analysis may prove to be a difficult task; covariates alone may not be capable to capture the whole dependence structure underlying network data.

- p^* **model** or **ERGM** (exponential Random Graph Models). This class of models represent an extension of **Markov Graphs**, proposed by Frank and Strauss, 1986, where it is assumed that possible edges between disjoint pairs of nodes are assumed to be independent, conditional on the rest of the graph. Thus, Markov Graphs introduce the idea of conditional independence between the edges, overcoming the independence hypothesis of previous models. ERGM builds on Markov Graphs and postulate that edge probabilities are of an exponential family form, and are also parametric functions of some summary statistics. A new feature of p^* models is based on the introduction of a number of triangles statistic when modelling edge probabilities, as a way to account for *transitivity*. Transitivity is a characteristic of many real network data, which may be expressed as "a friend of mine is a friend of mine". Higher order statistics can be considered as well when modelling edge probabilities, increasing the complexity of the model. Some basic references can be found in Wasserman and Pattison, 1996 and Robins et al., 2006.

- **Latent variable models**.This class of models aim at explaining the observed connections in a network by means of some unobserved, latent variable. In general, this broad class of models studies edge probabilities via an additional layer of modelling, that is the layer of the latent variable(s); edges are assumed to be independent conditional on some parametric function of the latent variable(s). As in Markov Graphs, the conditional independence assumption allows to better model the complex structure of network data, while the latent variable(s) introduces flexibility in the analysis. The two main sub-classes that have been proposed in the literature are:

  - **Stochastic Block models**. This class of models is an extension of blockmodelling, a non-stochastic approach for clustering of network data. Block models assume that a network can be decomposed into a set of G clusters, $C_1, \ldots, C_G$. Interactions within and between such clusters can be explained by a set of blocks, $B_{11}, B_{12}, \ldots, B_{GG}$. In block models, nodes are assigned to the same cluster if they are equivalent with respect to a given criterion. Usually, such criterion is determined by some external source of information, as for example sex in a student network or party membership in political network data. As node labels may not always be available, stochastic block models (Holland, Laskey, and Leinhardt,

1983; Nowicki and Snijders, 2001; Snijders and Nowicki, 1997) treat them as latent variables. In its basic formulation, two nodes belonging to the same cluster are considered stochastically equivalent. Edge probabilities for the dyads (i, j), $i, j = 1, \ldots, n$, are assumed to be conditionally independent given the cluster membership of the nodes. Stochastic block models are a valid approach to parsimoniously cluster network data, as $n \times n$ edge probabilities can be summarized by G^2 blocks. However, using such blocks prevents from accounting node-specific variability in modelling edge probabilities. An extension of such models that introduces more flexibility in the modelling of edge probabilities is the **Mixed Membership Stochastic Block model** by Airoldi et al., 2008. In this framework, the membership of a node may change with respect to the nodes it is interacting with.

– **Latent Space models**. This class of models, introduced by Hoff, Raftery, and Handcock, 2002, assume that nodes have latent positions in a $p-$dimensional Euclidean space. Edge probabilities p_{ij} depend on the latent positions of node i and j, $i, j = 1, \ldots, n$, and edges are assumed to be independent conditionally on such positions. More specifically, edge probabilities depend on some function of the latent coordinates, which is normally assumed to be the Euclidean distance; however, different specifications may be considered as well. We will discuss the principal ones in the next section.
As this thesis mainly focuses on latent space models, in the following section we provide a brief description of the main latent space models developed for single and multidimensional network data.

1.3 Latent space models

Latent space models, introduced by Hoff, Raftery, and Handcock, 2002, postulate that edges are independent conditional on some function of the node-specific unknown positions (coordinates), $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$, generally assumed to be normally distributed, in an unobserved $p-$dimensional Euclidean space. The basic model assumes that such a function is the Euclidean distance function, $d(\boldsymbol{z}_i, \boldsymbol{z}_j) = d_{ij}$: nodes close in the latent space have an high probability to link in the network. Thus, edge probability is an inverse function of the distance between the nodes in the latent space. This assumption easily allows to account for reciprocity, as the distance between node i and j is symmetric. Further, it allows to deal with transitivity, as if node i and j are close and j is close to node l, then also i and l may be close. The density of a network is modelled via an intercept parameter which regulates the effect of the distances in the latent space on the edge probabilities. Also, in most formulations of the latent space model, edge probabilities are assumed to be a logistic function of the latent positions and some other parameters. Edge covariates can be included in the specification too, therefore also homophily by attributes can be accounted for. Alternatives to the Euclidean distance have been proposed in the literature; in particular the most relevant ones are:

- *Squared Euclidean distance.* Proposed by Gollini and Murphy, 2016, this

function provides with the same latent space representation of the network as that of the Euclidean distance. However, the squared Euclidean distance, d_{ij}^2, allows to heavily penalize those edge probabilities corresponding to great distances in the latent space and less those corresponding to small distances. Such feature may prove to be useful when performing inference on the latent coordinates, as it allows to better discriminate between close and far positions.

- *Projection model.* This latent space model was proposed by Hoff, Raftery, and Handcock, 2002 as an alternative to the distance model. It postulates that edge probabilities p_{ij}, $i, j = 1, \dots, n$, are a function of the ratio of the inner product between the latent coordinates $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ to the norm of either $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. When the interest is in modelling the edge receiving propensity of each node, it is assumed that $p_{ij} \propto \boldsymbol{z}_i'\boldsymbol{z}_j/\|\boldsymbol{z}_j\|$, instead, when the interest is in modelling the edge sending propensity, $p_{ij} \propto \boldsymbol{z}_i'\boldsymbol{z}_j/\|\boldsymbol{z}_i\|$. While distance based latent space models produce symmetric edge probability matrices, the projection model does not, as node-specific effects $\|\boldsymbol{z}_i\|$ regulate the impact of the latent distances in the edge probabilities. Thus, this model accounts for something similar to sender or receiver effects, but it does not allow to separate the interpretation of such effects from that of the inner product of the latent coordinates. The inner product, which is symmetric, may account for reciprocity in network data.

- *Multiplicative model.* The multiplicative latent space model, introduced by Hoff, 2005, assumes that edge probability are a function of the inner product of the latent coordinates, as well as of other network parameters. As the inner product between coordinates $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ depends on their distances from the origin of the space, a latent space representation arising from a multiplicative model has different interpretation from that of the Euclidean distance model. As an example, we report in Figure 1.3 a latent space representation, with $p = 2$, of a network with $n = 4$ nodes. By graphical inspection, we may derive that z_1 and z_2 are exactly as close as z_3 and z_4; indeed, $d(1, 2) = d(3, 4) = 1.41$. In Euclidean distance latent space model we may interpret this equal closeness as an equal degree of similarity between the dyads $(1, 2)$ and $(3, 4)$. However, such an interpretation would not be correct in the multiplicative model as $z_1'z_2 = 2.75 \neq z_3'z_4 = 5.75$. Latent coordinates z_3 and z_4 are further from the origin then z_1 and z_2, and thus the dyads $(1, 2)$ and $(3, 4)$ interact differently. Hence, when the aim is at representing node similarities in the latent space, distance models seem to be providing with a more easily interpretable solution. Also, the coordinates of nodes 3 and 4, z_3 and z_4, are equidistant from that of nodes 1 and 2. Indeed, the Euclidean distances, $d(1, 3) = d(1, 4) = 1.58$ and $d(2, 3) = d(2, 4) = 2.92$, capture it and give the idea that node 3 and 4 interact in the same way with node 1 and node 2. This same interpretation does not hold with a multiplicative latent space model, as, for example, node 3 has an higher probability to interact with node 1 than node 4, as $z_1'z_3 = 5.00$ while $z_1'z_4 = 2.50$.
  A feature of the multiplicative model is that it guarantees, via the inner product, that, given any three nodes i, j, l with latent coordinates $\boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{z}_l$,

either $sign(\boldsymbol{z}_i'\boldsymbol{z}_j) = sign(\boldsymbol{z}_j'\boldsymbol{z}_l) = sign(\boldsymbol{z}_l'\boldsymbol{z}_i) = 1$, or $sign(\boldsymbol{z}_i'\boldsymbol{z}_j) = 1$ and $sign(\boldsymbol{z}_j'\boldsymbol{z}_l) = sign(\boldsymbol{z}_l'\boldsymbol{z}_i) = 0$, where 1 denotes positive sign and 0 negative sign. Such feature enforces a strong notion of transitivity in the latent space. Indeed, given any node i, any two nodes (j, l) interacting with it have to interact between themselves with a behaviour dictated by node i. For example, if node i "relates" with both nodes j and l, then nodes j and l should relate as well. The allowed relationships are of the form "a friend of mine is a friend of mine" and "an enemy of a friend of mine is my enemy". The configuration "an enemy of a friend of mine is a friend of mine" can not be represented. Instead, the distance model poses no explicit restriction on the relation between the distances between any three points, other than the usual triangular inequality, $d(i, j) \leq d(i, l) + d(l, j)$.

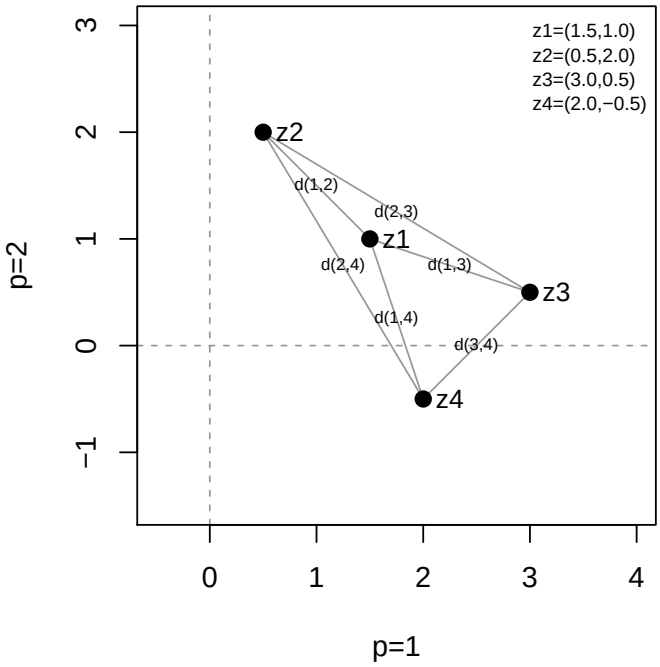


Figure 1.3. Two examples of a latent space representation of a network.

- Latent space models based on *ultrametrics*. Schweinberger and Snijders, 2003 consider a latent space build on an ultrametric. The ultrametric imposes stronger relations between the nodes than those imposed by the Euclidean distance. In particular, given a function $d(\cdot)$ and two nodes i, j, we have that $d(i, j) \leq max[d(i, l), d(l, j)] \leq d(i, l) + d(l, j)$, for any other node l. The first, more restrictive, bound is encountered when the function is an ultrametric, the second when it is a metric (as it is the case with the Euclidean distance). In particular, the ultrametric condition implies that all the triangles are isosceles; there could not be all different values of the ultrametric function computed between any dyads, and this corresponds to at least two equal edge probabilities p_{ij}, p_{lr}.

Due to the complex structure of network data, latent space models for networks have been always estimated within a Bayesian framework by adopting a Gaussian representation for the latent coordinates. The two main approaches are standard MCMC algorithms, which usually provide faithful estimates but have potentially long computational times, and variational Bayes algorithm, whose estimation is quite fast but often suffers from numerical instability problems. The trade-off between feasible computational times and reliability of the estimates is crucial and should always be considered when choosing the inferential approach to be used for model estimation.

1.3.1 Latent space models for multidimensional networks

A couple of decades ago, analysing multidimensional network data corresponded to find a way to aggregate the views into a single one, and then analyse the "collapsed" network with standard models available for single networks. Some well known datasets in single networks literature, as for example the "Zachary karate club" dataset, are the result of such an "aggregation". After these first attempts in multiplexes modelling, models specifically designed to deal with multiple views have been developed. In particular, in the context of latent space models, while the majority of the models have been designed specifically for single networks, several latent space models tailored to multidimensional networks have been proposed. Among such models, a distinction has to be made between latent space models for multidimensional network data and those for dynamic network data. Indeed, latent space models for dynamic networks assume temporal dependence between the networks and aim at capturing the potential evolution of the latent space trough time. An example is the modelling framework developed by Sarkar and Moore, 2005, later extended by Sewell and Chen, 2015, where it is assumed that node-specific latent coordinates evolve trough time, that is through the subsequent networks, following a first order Markov process. The dynamic component in the data is addressed, and the latent spaces serve to capture and visualize changes of the interactions between the nodes through time. The modelling framework proposed by Sewell and Chen, 2015 is quite flexible and allows also to model, with small adjustments, weighted network data or clustering tendencies in dynamic network data. Up to our knowledge, latent space models for dynamic networks with "longer memory" have not yet been proposed. For a more detailed descriptions of latent space models for dynamic network data, we refer to Kim et al., 2018.
Differently from dynamic networks, views in multiplex data can not be modelled by using a predetermined structure; no consequentiality can be assumed between the networks, and neither any pre-specified association rule. Thus, different aims can be pursued when modelling multidimensional network data. We may summarize such aims into, at least, three different questions:

1. How do the nodes interact at a single view level?

2. Are the node-specific behaviours in the networks somehow associated?

3. Is there any common tendency in the node-specific behaviours across the views?

Depending on the data at hand and the modelling purposes, we may be interested in only one of the above questions, two of them or all three. Single views node interactions have usually been modelled via single-network latent spaces. Basic models as the one proposed by Sweet, Thomas, and Junker, 2013 do not consider any "higher level" structure beyond the single-network level, but other authors proposed different ways to address the global node behaviours. One example is that of the model by Salter-Townshend and McCormick, 2017, which aims at capturing the associations between single network latent spaces, corresponding to different views in the multiplex, by means of association parameters defined for each couple of networks. Such a model can be quite useful when the purpose is to synthetically capture which views are similar or dissimilar. However, when the number of views is high, this approach does not provide an easily interpretable insight on the overall structure underlying the observed multidimensional network. A different approach has been proposed by Gollini and Murphy, 2016, who proposes to combine the information coming from different networks assuming that these have been generated by a *joint* latent space. Such a framework permits to have a look both at node-specific coordinates in the single networks and in the global (joint) representation. For a more detailed description of the model by Gollini and Murphy, 2016, we refer to the corresponding paper and to Appendix D. Later, single-network latent spaces are summarized by Durante, Dunson, and Vogelstein, 2017, which assumes that the K different views have been generated by a mixture of $H \leq K$ latent spaces. Motivated by a $K = 42$ views brain multiplex, such framework may be useful when the number of networks is quite high. Indeed, as networks are "units" belonging to the H mixture components, few network-units correspond to a component, and this could be a too small sample size for a mixture model. All of the above models have been derived in the context of binary multidimensional network data.
The following section will give an overview of the remaining chapters and appendixes.

1.4 Chapter summaries

Chapter 2 presents an extension of the latent space model by Hoff, Raftery, and Handcock, 2002 to model multidimensional network data. The proposed model structure aims to capture shared similarities between the nodes in a single latent space, when different networks are assumed to be different realizations of a single underlying phenomenon. The model is used to describe the potential affinities between competing Countries in the Eurovision Song Contest during the period 1998-2015, with the aim to verify whether the observed votes exchange was influenced by such affinities.
The contents of Chapter 2 have been developed with Prof. Marco Alfò and Prof. Thomas Brendan Murphy, and are reported in a paper which has been accepted for publication in the *Annals of Applied Statistics*, see D'Angelo, Murphy, and Alfò, 2018. Also, the proposed model structure has been implemented in the *R* package *spaceNet*, available on CRAN (`https://CRAN.R-project.org/package=spaceNet`).
Chapter 3 introduces a novel class of latent space models for multidimensional network data analysis, which allows for node-specific effects. A parsimonious representation of the multiplex may be obtained, as a single latent space is employed to describe

the overall shared similarities between the nodes, while asymmetrical node features are captured by the node-specific effects. Various types of node-specific effects may be specified, in order to model different degrees of heterogeneity in nodes behaviours, both within and between networks.
The contents of Chapter 3 have been developed with Prof. Marco Alfò and Prof. Thomas Brendan Murphy, and are reported in a paper which is currently under first revision in *Social Networks*, see D'Angelo, Alfò, and Murphy, 2018.
Chapter 4 presents a latent space model for network and multidimensional network data, which enables clustering of the nodes in the latent space. Clusters in the latent space correspond to communities of nodes in the multiplex, with clustering being a phenomenon which often arise in many observed network data, especially in social networks. Clustering structures are modelled using an infinite mixture distribution framework, which allows to perform joint inference on the number of clusters and the cluster parameters. The performance of the proposed method is illustrated through an application to Vickers multiplex data (Vickers and Chan, 1981), which represent different social relations among a group of students.
Chapter 4 is a pre-print and its content have been developed with Dr. Michael Fop and Prof. Marco Alfò.
All the models developed in Chapters 2-4 have been developed within a Bayesian hierarchical framework, with inference on model parameters being carried out via MCMC algorithms.
Appendices A-C report the supplementary materials for Chapters 2-4.
Last, Appendix D presents a different approach to estimate the *lsjm* model by Gollini and Murphy, 2016, where inference is carried out via an MCMC algorithm, instead of a Variational algorithm. In this Appendix, we prove that using a different procedure for model estimation we may obtain higher quality estimates. Also, we propose a comparison with the model developed in Chapter 2. Such model may be seen as a parsimonious version of the model by Gollini and Murphy, 2016, where the common latent space is used to represent the *joint* latent space of the *lsjm* model.

Chapter 2

Latent Space Models for Multidimensional Networks

2.1 Introduction

The Eurovision Song Contest is a popular TV show, held since 1956, that takes place every year with participants from the countries member of the European Broadcasting Union. The competition has undergone several modifications through years and the number of participants has increased, together with the popularity of the show. Since its beginning, countries had to express their preferences for the competing songs through a voting system; representatives vote only for the songs that meet their tastes. Despite that, many issues of bias in the voting system have been raised during the years (Yair, 1995). In the press and the literature, it has often been claimed that votes are not only the expression of preferences for the songs, but for the performing countries themselves. Therefore, it has been claimed that the exchange of votes is not random but rather it is determined by some kind of similarity: the more two countries are close according to an unknown proximity measure, the more they will tend to vote for each other.
The exchange of votes in the Eurovision contest can be represented by means of a network, where the countries represent the nodes and the votes are recorded as edges. More specifically, within each annual edition of the Contest, the data may be represented in the form of an adjacency matrix $\mathbf{Y}$, with generic element $y_{ij} = 1$ if a representative of country i votes for a song by a performer from the j^{th} country and 0 otherwise, where $i, j = 1, \ldots, n$ indexes countries. Network data can be represented by means of graph theory. More formally, a network is thought to be the realization of a graph $G(N, E)$, where N denotes the set of nodes and E the set of edges. The number of observed nodes and edges will be denoted, respectively, by $|N| = n$ and $|E| = e$. Generally, the law generating the observed networks is unknown and several different models have been proposed to describe such complex structures. Erdős and Rényi, 1959 and Erdős and Rényi, 1960 modelled arch formation in a network as arising from a random process: each dyad (i, j) is independent and the probability of forming a link is constant over the network. This first model was generalized, both relaxing the assumption of constant edge probability over the network and the assumption of independence of the dyads. Holland and Leinhardt, 1981 with models

p_1 and p_2 kept the assumption of independence among the dyads but increased the number of parameters describing edge probabilities, to take into account the attractiveness of a node (the highest the value the highest the probability for this node to be connected with others) and the mutuality (the propensity of forming symmetric relations). The independence assumption on the dyads was then relaxed via the introduction of *Markov graphs* by Frank and Strauss, 1986, attempting to model triangular relations in a network. Later on p^* models or ERGMs (Exponential Random Graph models) have extended the work done by Frank and Strauss, 1986 introducing differnet summary statistics, see for example Krivitsky et al., 2009 and Robins et al., 2007. A different approach is the so-called *stochastic block model*, which attempts to decompose the nodes in different sub-groups, see Holland, Laskey, and Leinhardt, 1983, Airoldi et al., 2008. In its basic formulation, nodes within a group have the same probability of forming edges, while this probability changes among groups. Hoff, Raftery, and Handcock, 2002 added an extra layer of dependence: the observed edge formation process is assumed to be a function of nodes' coordinates in a (low-dimensional) latent space. Two different specifications are considered, the *distance model*, where the latent space is euclidean, and the *projection model*, where it is bilinear. The model by Hoff, Raftery, and Handcock, 2002 has been extended to perform clustering on the latent nodes' coordinates by Handcock, Raftery, and Tantrum, 2007. A generalization of the projection model by Hoff, Raftery, and Handcock, 2002 is the multiplicative latent space model by Hoff, 2005, designed to capture certain types of third-order dependence patterns in the network. Another approach that makes use of latent variables has been proposed by Snijders and Nowicki, 1997. This model is based on the stochastic block model by Holland, Laskey, and Leinhardt, 1983, where latent variables are introduced in the determination of the nodes' group memberships. A more exhaustive review of models for statistical network analysis can be found in Goldenberg et al., 2010, Salter-Townshend et al., 2012 and Murphy, 2015.

The models presented above refer to single networks, that is, in the present context, to the modelling of one single edition of the Eurovision Song Contest or a summary of several editions. If a group of editions of the Contest is considered, many replications of the adjacency matrices, representing the preferences expressed by countries towards others, are available. Therefore, the data can be described by a multidimensional network (or multiplex), $\mathbf{Y} = (\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)})$, which may be thought of as the realization of a collection of graphs $G = (G^{(1)}, \ldots, G^{(K)})$, where $k = 1, \ldots, K$ indexes editions. The generic graph $G^{(k)} = (N, E^{(k)})$ has the same set of nodes N as the others $(K-1)$ graphs in the collection (the participants to the group of editions), but potentially different set of edges $E^{(k)}$ (the preferences expressed in each edition). Hence, a multidimensional network describes different (independent) realizations of a relation among the same group of nodes. Different models have been developed to deal with this kind of data. Fienberg, Meyer, and Wasserman, 1985 adapted a log linear model to the context of multiplex data. Greene and Cunningham, 2013 proposed to summarize the information coming from all the different networks (views) aggregating them into a single one. Sweet, Thomas, and Junker, 2013 proposed a Hierarchical Latent Space model, which generalizes network latent space models to a collection of networks. The joint multiplex distribution factorizes into single network distributions, which are modelled independently and

inference is carried out via MCMC. Gollini and Murphy, 2016 extended the latent space model in Hoff, Raftery, and Handcock, 2002 to multiplex data, assuming that the edge probabilities are function of a single latent variable. To estimate the joint latent space coordinates, they propose to use a variational Bayes algorithm and decompose the posterior distribution, fitting a different latent space to each network. Then, the separate estimates are employed to recover the joint latent space. The multiplicative latent space model was extended by Hoff, 2015 to the context of multidimensional networks. In that case, each network in the multiplex is modelled with its own latent space, independently from the others. Salter-Townshend and McCormick, 2017 proposed a method to jointly model the structure within a network and the correlation among networks via a Multivariate Bernoulli model. Another approach developed to describe the (marginal) correlation among different networks in a multiplex has been proposed by Butts and Carley, 2005. Hoff, 2011 proposed to model multiplex data as multiway arrays and applied low-rank factorization to infer the underlying structure. Durante, Dunson, and Vogelstein, 2017 proposed a Bayesian non-parametric approach to latent space modelling, where clustering is performed on the latent space dimensions in order to discriminate the most relevant ones for each view.
The present work aims at recovering the similarities among countries, modelling the exchange of votes during several editions of the Eurovision Song Contest. We adopt a framework similar to that of Gollini and Murphy, 2016 and we consider the projection of the countries into a common latent space. Similarities among countries are then expressed in terms of distances in this latent space. We introduce network-specific coefficient parameters to weight the relevance of the latent space in the determination of edge probabilities in each network. We consider the editions that took place after the introduction of the televoting system and focus on the period 1998-2015. Further, we consider geographical and cultural covariates in the analysis.
The paper is organised as follows. Section 2.2 summarizes the history of the Eurovision Song Contest together with the principal works on the subject (section 2.2.1) and presents the analysed data (section 2.2.2). Latent space models for network data are introduced in section 2.3 and the proposed model is outlined in section 2.3.1. Model estimation is discussed in section 2.4. Further issues are discussed in section 2.5, such as model identifiability (section 2.5.1), missing data (section 2.5.2) and the introduction of edge-specific covariates (section 2.5.3). The application is presented in section 2.6 and the results are discussed in section 2.7. A large scale simulation study is outlined in section 2.8, where also the main findings are reported. Section 2.9 presents the results of a comparison between the proposed model and the *lsjm* by (Gollini and Murphy, 2016). We conclude with some discussion in section 2.10.

2.2 The Eurovision Song Contest

2.2.1 History of the contest and previous works on the subject

The Eurovision Song Contest, held since 1956, is a TV singing competition where the participant countries are members of the EBU (European Broadcasting Union). Despite its name, the European Broadcasting Union includes both European and

non European countries. Indeed, Eurovision's fame has spread all over the world during the last years and it has been broadcast from South America to Australia. It is the non-sportive TV program with the largest audience in the world and one of the oldest ones (Lynch, 2015).
From its first edition, where only seven countries competed, there have been several changes in the number of participants, the voting system and the structure of the competition. Due to the increasing popularity of the program, many countries have been included in the contest. The current structure of the contest consists of two preliminary stages used to select the finalists, followed by the final stage for the title. The voting system has been modified several times, in the voting procedure and the grading scheme. In the early years of the competition, a jury elected the winning song. Later, the system has been supported by televoting [1], introduced in 1998 in all the competing countries. As for the grading scheme, it is positional[2] since 1962, but the method used to rank the countries has been modified across the different editions. From 1975 to 2015, each country had to express its top ten preferences ranking them from the most to the least favourite using the following scores: 12, 10, 8, 7, 6, 5, 4, 3, 2, 1. Each country had to vote exactly ten others, could not vote for itself and each grade could be used only once. At the end, the country receiving the highest overall score would have won the competition. A restriction has been imposed on the lyrics in the past, as the participants were required to perform a song written in their national language. However, this rule was definitively abolished after 1998.
Every year, both the singer and the song representing a country change, making each edition of the Eurovision independent from the previous one. Indeed, the structure of the competition is built in such a way that the past results will not influence the future performances. Countries should vote only according to their tastes and, as musical evaluation has no objective criteria, the voting results should not depend on the countries themselves, but only on the songs. However, this claim was often doubted, especially after the introduction of televoting. Several issues have been raised on the voting system, which was said to be biased. The first paper investigating the presence of bias in the voting system is Yair, 1995. This work considers voting relations among 22 of the 24 countries competing in the period 1975-1992 and claims that, according to their voting preferences, they can be clustered in three regional blocks: Mediterranean, Western and Northern. Countries tend to vote for others from the same block, hence following a non-objective (*non-democratic*) behaviour. However, the paper does not provide an in-depth statistical evaluation of the results. The author supports the theory that the geographic location of a country may influence its voting behaviour. This assumption has been further investigated by Fenn et al., 2006; in this work, the dynamic evolution of votes exchanged in the competition 1992 to 2003 has been analysed, with the aim at looking for sub-groups of countries. The sub-groups found are not fully explained by

[1] Televoting is a voting method conducted by telephone. The organizers of the event provide the audience with telephone numbers associated with the different participants. The rankings are then determined by the number of calls/SMS that each contestant receives.

[2] Positional voting is a ranked voting system where a list of candidates has to be ordered by voters. Rankings of different voters are converted into points and cumulated in scores, associated with each contestant. The one receiving the highest final score wins.

the geographical positions of the countries. Clerides and Stengos, 2006 developed an econometric framework to analyse the data and arrived to similar conclusions, in the sense that the authors do not find any "strategic" vote exchange in the period 1981-2005. Saavedraa, Efstathioua, and Reed-Tsochasb, 2007 investigated the structural properties of the dynamic network for the period 1984-2003 via q-analysis and found that clustering arises mainly between countries closed to each other in a geographical sense. Spierdijk and Vellekoop, 2006 applied multilevel models to look for the influence of geographic and cultural factors in the vote exchange from 1975 to 2003 and found that these do not explain the behaviour of all the competing countries. Ginsburgh and Noury, 2008 claim that having a similar culture may influence the votes expressed by a country. Cultural proximity, as well as geographic proximity and migration flows in the period 1998-2012 have been investigated as sources of bias by Blangiardo and Baio, 2014. The authors discovered the presence of a mild positive bias among few couples of countries but no evidence of a negative bias overall. Mantzaris, Rein, and Hopkins, 2018 analysed the editions 1975-2005 searching for couples of countries exhibiting preferential voting. They investigate the hypothesis of random allocation in the votes, and found some evidence that geographic proximity is influential up to an extent; however, many countries do not tend to vote according to such a rule.
The aforementioned works show that there has been a growing interest in the structure underlying votes exchange in the Eurovision Song Contest during the past twenty years. The authors investigated the influence of social, geographical, cultural and political factors on the mechanism forming preferences and agree that, at least to some extent, these components may be relevant. However, none of the factors above is able to explain satisfactorily the votes exchanged in the competition for the last years.

2.2.2 Data

The assumption that the exchange of votes is driven by similarities among countries may be reasonable. However, similarities among countries might not coincide with social, geographical, cultural and political factors that can be explicitly measured. In fact, there could be some unobservable (latent) factors influencing such a process. The aim of this work is at recovering the underlying latent similarities among countries. The idea is that the more two countries are similar, the more they will tend to vote for each other. To recover recurrent voting patterns, we need to examine a collection of editions. In particular, we focus on years subsequent to the introduction of the televoting system, 1998-2015. We assume that the televoting preferences reflect the preferences of the whole population, and that they are more representative when compared to the jury opinions alone. We do not consider years 2016 and 2017 because the voting system was again modified in that period. In fact, in the period from 1998 to 2015, votes given by the jury and by televoting were jointly considered: the final top ten for each country was determined looking at the intersection of the most voted songs by the two sources. From 2016 onwards, the final preferences expressed by each country are given by the union of the ten favourites of the jury and the ten favourites of the televoting. That is, in the last two years of the competition, each country could elicit more than ten preferences.

In the period 1998-2015, the countries were allowed to sing in any language and most of the songs were in English. The period we consider is homogeneous both with respect to the voting system and (for the large part) to the language used in the performance. Subsequent editions of the Eurovision are assumed not to depend on one another, as the singer and the song performed change every year without any pre-determined criteria. This assumption allows us to consider the exchange of votes in different editions of the contest as replications of the same phenomenon, which is the expression of musical appreciation between couples of countries. These different replications of preferences among countries will then be used to recover the similarities. Indeed, the basic assumption is that the more two countries are similar, the more they tend to vote for one another through the editions.
In each edition, the votes exchanged among the countries can be described by a network, where nodes represent the countries. As we are interested in the exchange of votes and not in the final ranking, an edge going from country i to country j will denote that i has j in its top ten. Vice versa, the edge will be from j to i if the latter has been voted by the j. Indeed, recall that although the grading scheme is positional, each country can express a limited amount of preferences $r = 10$ in the analysed period. Rosen, 1972 proved that when sampling r units from a population of size[3] n, if $r << n$, the first r units are independent with respect to the extraction order. As our interest lies in modelling the exchange of votes and not in determining the winner in each edition, we can treat the networks as binary, without much loss of information. An alternative approach, designed for contexts where the ranking scale is equally spaced, may be found in Hoff, 2015.
The resulting network is then directed, acyclic (as countries can not vote for themselves) and un-weighted. If we consider a group of editions for the contest, we will have a collection of networks, defined on the same group of countries (see paragraph 2.5.2), and this object is indeed a multidimensional network. The following section introduces a general modelling framework for such data.

2.3 Latent space models for multidimensional networks

Latent space models have been introduced by Hoff, Raftery, and Handcock, 2002 with the aim at reducing the complexity typical to the dependence structure in network data. This purpose is achieved via geometric projection of the nodes onto a low dimensional space. The probability of observing an edge between two nodes is assumed to be a function of the unknown nodes' coordinates in the latent space. Conditionally on the set of latent positions, the observed binary indicators are assumed to be independent across networks. Hoff, Raftery, and Handcock, 2002 distinguish between distance and projection latent space models, depending on the choice of the function that summarizes the latent coordinates. The distance model assumes that this is indeed a distance function. Gollini and Murphy, 2016 extended the distance model described by Hoff, Raftery, and Handcock, 2002 to the case of multiplex data, with the introduction of a so called Latent Space Joint Model (*lsjm* in the following). In that context, the probability of an edge in a given network depends on a specific latent space and a network specific intercept. The latent spaces

[3] In the present context, $n = 48$ is the number of countries that country i can potentially vote.

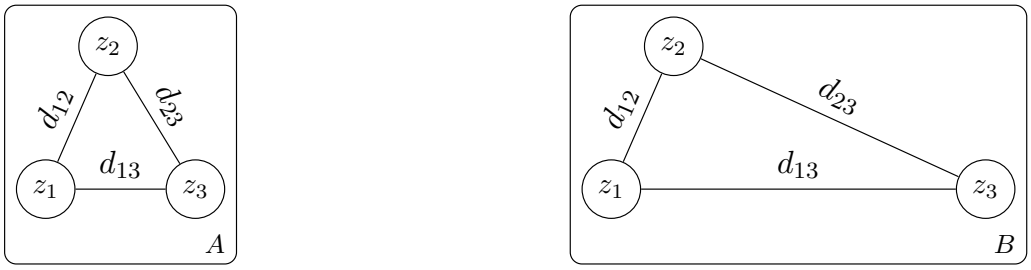


Figure 2.1. Latent space representations of a network with 3 nodes.

(one for each network) are thought to be realizations of a common latent space, which captures the average latent coordinates of the nodes and is behind all the networks. The variational approach used to estimate model parameters is fast but suffers from computational issues when the dimension of the multiplex is relatively high, either in the number of nodes n or in the number of networks K.
The present work builds on the model of Gollini and Murphy, 2016 with the aim at recovering the similarities among the countries participating in the Eurovision Song Contest. As our interest lies in recovering such similarities, rather than modelling observed ranks, the choice of a distance latent space model to reconstruct the network is quite appropriate. However, distances correspond to symmetric relations, which is a characteristic of similarity measures. Figure 2.1 shows an example for the latent space representation of a 3 nodes network. Node z_3 in space A has been moved in space B, so that $d_{13}^{(A)} < d_{13}^{(B)}$ and $d_{23}^{(A)} < d_{23}^{(B)}$. In our model, this correspond to a higher probability to observe a link between node 1 and 3 (or node 1 and 2) in space A when compared to space B. The distances in the latent space are scaled by a network specific coefficient, to weight the influence of the latent space in the determination of the votes for a given edition of the Contest. The lowest the value of this coefficient, the more the structure of edge probabilities resembles a random graph in this edition. This could lead to reject the claim that the votes patterns for the Eurovision contest are biased by some pre-existing preferences among countries.
The latent space is taken to have dimension $p = 2$, to allow for a graphical visualization and for a comparison of the estimated latent coordinates with the geographical positions (latitude and longitude) for the analysed countries. In this way we may be able to tell whether the latent configuration obtained resembles the geographical one and, if so, we may conclude that the position of a country on the map is indeed of some relevance in the competition. However, the choice of p is still an open problem in the literature and for other applications the choice of $p = 2$ could be suboptimal.
The model, formulated in section 2.3.1, also allows for the introduction of edge-specific covariates. In the present application, we use cultural covariates, such as the presence of a common language among a couple of participants (countries), to see whether these contribute to the formation of preferences. The multidimensional network is defined on all the countries that took part at least once in the Eurovision in the period 1998-2015, see section 2.5.2 for details.
The set collection of adjacency matrices will be denoted by $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}\}$. The generic element of each matrix, $y_{ij}^{(k)}$, in the collection is binary, with $y_{ij}^{(k)} = 1$ if there is an edge from node i to node j in the k_{th}

network, $y_{ij}^{(k)} = 0$ else. Indexes $i, j = 1, \dots, n$ are used to denote nodes in the network (countries) and index $k = 1, \dots, K$ refers to the K different networks in the multiplex.

2.3.1 The proposed model

Given the assumptions made in section 2.3 and following Hoff, Raftery, and Handcock, 2002 and Gollini and Murphy, 2016, the probability of observing an edge between node i and node j in the k^{th} network of the multiplex is given by:

$$Pr\Big(y_{ij}^{(k)} = 1 \mid \boldsymbol{\Omega}^{(k)}, d(\boldsymbol{z}_i, \boldsymbol{z}_j)\Big) = \frac{\exp\{\alpha^{(k)} - \beta^{(k)} d(\boldsymbol{z}_i, \boldsymbol{z}_j)\}}{1 + \exp\{\alpha^{(k)} - \beta^{(k)} d(\boldsymbol{z}_i, \boldsymbol{z}_j)\}} = p_{ij}^{(k)} \tag{2.1}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(K)}) = (\alpha^{(1)}, \dots, \alpha^{(K)}, \beta^{(1)}, \dots, \beta^{(K)})$ is the set of model parameters.
Following Gollini and Murphy, 2016, the function $d(\cdot, \cdot)$ is taken to be the squared Euclidean distance, that is $d(\boldsymbol{z}_i, \boldsymbol{z}_j) = \sum_{l=1}^{p} (z_{il} - z_{jl})^2 = d_{ij}$. The distance matrix will be denoted by $\mathbf{D}$. This choice of the distance function allows to penalize more heavily the probability of an edge linking two nodes that are far apart in the latent space when compared to one linking two closer nodes. Therefore, the latent space part of the model will push towards a non-random structure for the matrix of edge probabilities.
The edge-probability matrices are symmetric, that is $p_{ij}^{(k)} = p_{ji}^{(k)}$. As already mentioned, this choice is driven by the fact that we are interested in estimating similarities between the countries, and, therefore, we need to impose a symmetric relation between dyads (i, j) and (j, i). If we decide to also model explicitely node-specific characteristics, by means of row and column effects, we could define non-symmetric edge probabilitiy matrices. Hoff, 2015 showed that, if we compute the sample variance and correlation of the row and column means deviations from the overall mean in the observed adjacency matrices, we can empirically evaluate the row and column effects. However, note that, in the Eurovision networks, the out-degree is fixed by construction and there is no variability in the row means, and no correlation between the row and column effects are present. Also, the variability of the column effects is very low (0.02 on average). Therefore we decide not to consider row and column-specific effects and have symmetric probability matrices. Last, note that our model does not imply that a network generated from it is undirected. Indeed, symmetric probability matrices can, and usually do, generate directed networks, as the edge-draws are independently conducted.
Each view is associated with a couple of network-specific parameters: $\beta^{(k)}$ and $\alpha^{(k)}$, with $k = 1, \dots, K$. Differently from Gollini and Murphy, 2016, we introduce the scaling coefficient $\beta^{(k)}$ to weight the influence of latent space for the k^{th} network on the determination of edge probabilities. This parametrization is particularly suited for the Eurovision application, as it helps to address the eventual presence of bias in the exchange of votes. Indeed, if in the k^{th} network $\beta^{(k)} \approx 0$, edge probabilities do not depend on the latent structure, and edges form randomly. On the contrary, when $\beta^{(k)} > 0$, the latent structure will impact the edges formation. If these coefficients are estimated to be non-null for all the networks, or most of them, the latent space

has a constant influence in determining the structure of the observed multiplex. According to the assumption that the probability of observing an edge decreases with growing distances, the constraint $\beta^{(k)} \geq 0$ must be imposed. As the coefficient is bounded and can not take negative values, it follows that the lowest the value of $\beta^{(k)}$, the closer the structure of the network will be to a random graph. That is, if $\beta^{(k)} = 0$, we have:

$$p_{ij}^{(k)} = \frac{\exp\{\alpha^{(k)}\}}{1 + \exp\{\alpha^{(k)}\}} = p_{RG}^{(k)},$$

and the model for the k^{th} network reduces to a random graph (Erdős and Rényi, 1959) with edge probability $p_{RG}^{(k)}$. Thus, the coefficient $\beta^{(k)}$, when it is different from zero, can only decrease the edge probability values with increasing distances. In other words, edge probabilities are bounded from above by $p_{RG}^{(k)}$. To counterbalance the effect of the coefficient, the intercept parameter $\alpha^{(k)}$ is bounded as well so that the graph corresponding to $p_{RG}^{(k)}$ is not disconnected. Indeed, according to the properties of random graphs (Erdős and Rényi, 1960), if $p_{RG}^{(k)} > \frac{(1-\epsilon)\log(n)}{n}$, the graph will almost surely be connected. Taking $\epsilon = 0$, as $n \to \infty$, this property can be expressed in terms of $\alpha^{(k)}$ as:

$$\alpha^{(k)} > \log\left(\frac{\log(n)}{n - \log(n)}\right) = LB(\alpha^{(k)}) = LB(\alpha).$$

Thus, the lower bound $LB(\alpha)$ is independent of $k = 1, \ldots, K$, as the node set is constant across the multidimensional network. Defining a lower bound prevents from assigning large negative values to the intercept parameters. Indeed, if $\alpha^{(k)}$ is too low, the effect of the latent distances would be dominated by the intercept parameter, even if this effect is relevant ($\beta^{(k)} > 0$). That is, large negative values of $\alpha^{(k)}$ in equation 2.1 would correspond to edge probabilities tending to 0 and numerically undistinguishable. In such case, two distinct matrices of edge probabilities having elements $p_{ij}^{(k)} \approx 0$ would lead numerically to the same likelihood.

2.4 Parameter estimation

2.4.1 Likelihood and posterior

Given the model for the edge probabilities defined in equation (2.1), the likelihood function for the model is a product of $Kn(n-1)$ terms:

$$L(\mathbf{\Omega}, \mathbf{D} \mid \mathbf{Y}) = \prod_{k=1}^{K} \prod_{i=1}^{n} \prod_{j \neq i} \left(p_{ij}^{(k)}\right)^{y_{ij}^{(k)}} \left(1 - p_{ij}^{(k)}\right)^{1 - y_{ij}^{(k)}}, \tag{2.2}$$

and the corresponding log-likelihood is:

$$\begin{aligned} \ell(\Omega, D \mid \mathbf{Y}) &= \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} \ell_{ij}^{(k)} \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} y_{ij}^{(k)} (\alpha^{(k)} - \beta^{(k)} d_{ij}) - \log\left(1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\right) \end{aligned} \tag{2.3}$$

As the matrices of edge probabilities are symmetric for all the analysed networks, one could equivalently consider only their upper or lower triangular part, and the number of terms to be considered in the product for the likelihood reduces to $K\binom{n}{2}$. Similarly to Gollini and Murphy, 2016 and Handcock, Raftery, and Tantrum, 2007, we adopt a Bayesian approach to estimate the model. As in Gollini and Murphy, 2016, the latent coordinates are assumed to be independent random variables distributed according to a standard p-variate Gaussian distribution: $\boldsymbol{z}_i \sim MVN_p(\mathbf{0}, \mathbf{I})$. In the present context, the dimension of the multivariate Gaussian is fixed to $p = 2$ (see section 2.3).
The parameter space for the intercepts and the coefficients is bounded, as described in paragraph 2.3.1. For this reason, the prior distributions for these parameters are described by truncated Gaussian distributions. As no a priori information is available on their relationship, they are assumed to be independent, both within and across the networks:$\alpha^{(k)} \sim N_{\left[LB(\alpha),\infty\right]}\left(\mu_\alpha, \sigma_\alpha^2\right)$ and $\beta^{(k)} \sim N_{\left[0,\infty\right]}\left(\mu_\beta, \sigma_\beta^2\right)$.

The unknown $\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2$ play the role of nuisance parameters. Indeed, their value is of no interest but their specification is relevant, as they determine the solutions for the parameters of interest, $\alpha^{(k)}$ and $\beta^{(k)}$. Given their relevant role in the model, we decided to estimate these parameters, to avoid subjective specifications of their values. For this purpose, an extra layer is introduced in the model, as described in Figure 2.2, leading to a hierarchical structure. The prior distributions specified for the nuisance parameters are: $\mu_\alpha|\sigma_\alpha^2 \sim N_{\left[LB(\alpha),\infty\right]}\left(m_\alpha, \tau_\alpha\sigma_\alpha^2\right)$, $\sigma_\alpha^2 \sim \text{Inv}\chi^2_{\nu_\alpha}$, $\mu_\beta|\sigma_\beta^2 \sim N_{\left[0,\infty\right]}\left(m_\beta, \tau_\beta\sigma_\beta^2\right)$ and $\sigma_\beta^2 \sim \text{Inv}\chi^2_{\nu_\beta}$.
The distributions for the nuisance parameters depend on a set of hyperparameters $\eta = (\nu_\alpha, \nu_\beta, \tau_\alpha, \tau_\beta)$, that have to be specified. However, their choice is not as influential as the nuisance parameters to get estimates of $\alpha^{(k)}$ and $\beta^{(k)}$; in the following, we will present some criteria for the determination of η that were found to work well in practice. The posterior distribution is therefore defined by:

$$\begin{aligned} P(\alpha, \beta, \mathbf{z}, \mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2|\mathbf{Y}) =& L(\alpha, \beta, \mathbf{z}|\mathbf{Y})\pi(\mathbf{z})\pi(\alpha|\mu_\alpha, \sigma_\alpha^2)\pi(\mu_\alpha|\sigma_\alpha^2, \tau_\alpha) \\ & \pi(\sigma_\alpha^2|\nu_\alpha)\pi(\beta|\mu_\beta, \sigma_\beta^2)\pi(\mu_\beta|\sigma_\beta^2, \tau_\beta)\pi(\sigma_\beta^2|\nu_\beta) \end{aligned} \tag{2.4}$$

The posterior distributions for the parameters $\alpha^{(k)}, \beta^{(k)}$ and for the latent coordinates are not available in closed form. To obtain parameter estimates, proposal

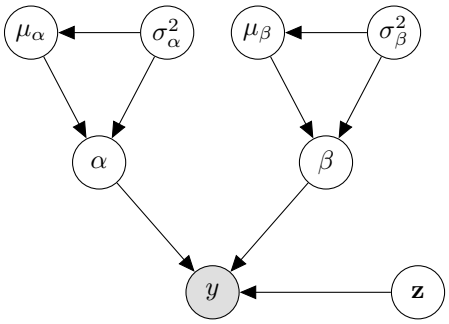


Figure 2.2. Hierarchy structure of the model.

distributions have been developed and are presented in the supplementary material, together with the distributions of the nuisance parameters.

2.4.2 The algorithm for parameter estimation

Estimation of model parameters is carried out using a Markov Chain Monte Carlo based approach. A detailed specification of the full conditional and the proposal distributions can be found in the supplementary material. Within each iteration of the chain, the nuisance parameters are updated from the corresponding full conditional; updated estimates for the intercepts, the coefficients and the latent coordinates are then proposed. The updates for the intercept $\alpha^{(k)}$ and the scaling coefficient $\beta^{(k)}$ in each network are jointly carried out, as it was empirically found that they may be correlated. The joint updating scheme helps improving the speed of convergence, while the latent coordinates are updated separately and sequentially. Indeed, it could be the case that the current estimates for a subset of the latent positions have already converged, while the remaining $\mathbf{z}_i$'s are still far from the "true" values. Jointly updating the $\mathbf{z}$ as block would not respond to the need to adjust just the *mislocated* coordinates; separate updating has been found to be a better strategy. After the set of latent coordinates has been updated at a given iteration of the algorithm, it has to be compared with the set of estimates obtained at the previous iteration. Since the likelihood in equation (2.2) considers the distances between the latent coordinates, it is invariant to rotation or translation of the latent positions $\mathbf{z}_i$. Therefore, it has to be ensured that the current set is not a "rigid" transformation of the previous ones, to prevent from non-optimal stationary solutions for the latent coordinates. To achieve such aim, Procrustes (Dryden and Mardia, 1999) correlation is computed and the current configuration is discharged if the value is above a certain threshold, fixed to be 0.85. The choice of this value, that ranges in $[0, 1]$ is arbitrary and should reflect the presence of high correlation. Values of the threshold above 0.80 have been found to work well in practice. The supplementary material reports the pseudo-code describing the estimation procedure.
Before starting the algorithm, the set of hyperparameters η needs to be defined. The degrees of freedom of the Inverse Chi-squared prior distribution for the variance parameters σ^2_α and σ^2_β are fixed to $\nu_\alpha = \nu_\beta = 3$, as values in the range $[2, 6]$ have been tried and it has been found that the different specifications do not have substantial impact on the parameter estimates. The variance-scale hyperparameters are set to be $\tau_\alpha = \tau_\beta = \frac{K-1}{K}$, so that the means of the proposal distributions for μ_α and μ_β reduce to the corrected sample means (see the supplementary material).
Starting values for the distances are taken to be the geodesic distances between the nodes in a randomly chosen network of the multiplex. From these distances, and for a fixed value for p, starting values for the latent coordinates are computed via multidimensional scaling as in Hoff, Raftery, and Handcock, 2002 and the starting values for the distances are obtained taking the squared Euclidean distances between the starting values for the latent coordinates. These are then used to model, via logistic regression, the adjacency matrices. The corresponding estimates for intercepts and scaling coefficients are taken to be the starting values for α and β. If these starting values fall outside the bounds specified for the parameters, they are replaced by these bounds. The nuisance parameters μ_α and μ_β are initialized as the

sample means of the initial estimates for α and β, respectively. In a similar fashion, σ_α^2 and σ_β^2 are initialised as the sample variances of α and β.

2.5 Further issues

2.5.1 Identifiability

As it can be easily noticed, the likelihood in equation (2.2) is invariant to linear transformations of the scaling coefficients. Indeed, for some constant c,

$$\alpha^{(k)} - \beta^{(k)} d_{ij} = \alpha^{(k)} - \frac{\beta^{(k)}}{c}(d_{ij} c) = \alpha^{(k)} - \beta^{(k)\star} d_{ij}^{\star}, \quad k = 1, \ldots, K.$$

For this reason, the coefficient for the reference network is fixed to $\beta^{(r)} = 1$ (index r denotes the reference network). The role of β parameters is to scale the distances and, therefore, the corresponding values are meaningful only when compared with each other given the reference. Thus, we have no loss of information in fixing $\beta^{(r)} = 1$. A further identifiability issue is:

$$\alpha^{(r)} - d_{ij} = (\alpha^{(r)} + c) - (d_{ij} + c) = \alpha^{(r)\star} - d_{ij}^{\star}.$$

To overcome this further issue, also the intercept $\alpha^{(r)}$ for the reference network needs to be fixed. As the intercept term defines an upper bound for the edge probabilities in the corresponding network, the value chosen for $\alpha^{(r)}$ should not underestimate this bound. We propose to fix it accordingly to the observed density in the network. More specifically, let us consider the expected value for the edge probability in network r:

$$\bar{p}^{(r)} = E\Big[\sum_{i=1}^{n}\sum_{j\neq i} P\Big(y_{ij}^{(r)} = 1 \mid \Omega^{(r)}, \mathbf{D}\Big)\Big] = E\left[\sum_{i=1}^{n}\sum_{j\neq i} \frac{\exp\{\alpha^{(r)} - d_{ij}\}}{1 + \exp\{\alpha^{(r)} - d_{ij}\}}\right].$$

A naive empirical approximation to this value, which is not available in closed form, is given by

$$\bar{p}^{(r)} \simeq E\left[\sum_{i=1}^{n}\sum_{j\neq i} \frac{\exp\{\alpha^{(r)} - 2\}}{1 + \exp\{\alpha^{(r)} - 2\}}\right] = \frac{\exp\{\alpha^{(r)} - 2\}}{1 + \exp\{\alpha^{(r)} - 2\}},$$

where the distances have been replaced by the constant 2 as this value is the mean empirical distance among coordinates simulated form a standard Gaussian distribution. This leads to:

$$\hat{\alpha}^{(r)} = \log\Big(\frac{\hat{\bar{p}}^{(r)}}{1 - \hat{\bar{p}}^{(r)}}\Big) + 2,$$

where $\bar{p}^{(r)}$ is $\hat{\bar{p}}^{(r)} = \sum_{i=1}^{n}\sum_{j\neq i} y_{ij}^{(r)}/(n(n-1))$. Thus, $\hat{\alpha}^{(r)}$, or any number greater than $\hat{\alpha}^{(r)}$, can be used as a fixed value for the intercept in the reference network.
As to the choice of the reference network, we may consider a network that is of particular interest. Alternatively, if there is no reason to prefer a network, a randomly chosen one can be selected. In the present work, for ease of interpretation of the results, the first (in terms of time) network of the multiplex has been set as the reference network.

2.5.2 The issue of "non participating" countries

Due to the increasing popularity that Eurovision gained over the years, many countries have requested to participate in the contest and have been accepted. With increasing number of participants, preliminary stages had to be introduced to select a smaller sub-group of countries accessing to the final, where they compete for the title. The winner of the previous edition enters the final straightforwardly, while the remaining countries have to compete in the qualifications round. Therefore, the selection of the finalists in the k^{th} edition does not depend on the results in the previous edition, apart from the specific case mentioned before. With the introduction of semi-finals, countries that do not make it to the last stage are allowed to vote for their favourite ten participants in the final. Their participation is passive, as they can vote but can not receive votes. Further, several countries have abandoned the competition for years, for a series of reasons. The pre-selection process, the presence of passive countries and the drop outs imply that the set of participants in two consecutive editions may not be exactly the same. However, the phenomena mentioned above are structural and, for that reason, we decided not to treat them as a missing values problem.
To model non participant (absent) countries, we define the set of nodes N in a more general way, as the set of countries that have voted at least once in the considered period. We can rewrite the log-likelihood as:

$$\ell(\mathbf{\Omega}, \mathbf{D} \mid \mathbf{Y}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} h_{ij}^{(k)} \ell_{ij}^{(k)}, \tag{2.5}$$

where $h_{ij}^{(k)}$ is an indicator variable, with $h_{ij}^{(k)} = 1$ if the i^{th} node was present in the k^{th} edition and could have voted for node j; while $h_{ij}^{(k)} = 0$ implies that the i^{th} node was not allowed to vote for node j in the k^{th} edition. Let us denote as $\mathbf{H}^{(k)}$ the binary matrix indicating whether or not a country was present in the k^{th} edition. The rows of the $\mathbf{H}^{(k)}$ matrix denote whether the corresponding countries may vote at the k^{th} occasion; more formally, if $\sum_{j=1}^{n} h_{ij} = 0$, then the i^{th} node was absent from that specific edition of the contest. Instead, the columns of $\mathbf{H}^{(k)}$ refer to the possibility of being voted for the corresponding countries. That is, if $h_{ji}^{(k)} = 1$ the i^{th} node has been voted in the k^{th} edition.

2.5.3 Covariates

Edge-specific covariates can be considered in the application. In the current application, all the used covariates do not depend on the specific network of the multiplex. From an exploratory analysis (see Figure 2.4) we could see that the association between the covariates and the different adjacency matrices was slightly varying over time. Therefore, in the current empirical application, we have considered constant effects, in the spirit of model parsimony. Of course, one could have had assumed that the effects of the covariates may vary over time. The implementation of such model would be straightforward.
Each covariate is stored in a $n \times n$ matrix, that will be denoted by $\mathbf{X}_f$, where $f = 1, \ldots, F$ is the index for the set of F covariates. To maintain the characterization

of the intercept term, the effect of the covariates is taken to be inversely related to edge probabilities (see section 2.6). Therefore, the effect associated to each covariate will be characterized in a similar fashion to the scaling coefficients $\beta^{(k)}$, $k = 1, \ldots, K$ (see the supplementary material). That is, the edge probability in equation (2.1) is modified to the following:

$$P\Big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, d_{ij}, \lambda, \mathbf{x}_{ij}\Big) = \frac{\exp\{\alpha^{(k)} - \beta^{(k)} d_{ij} - \sum_{l=1}^{F} \lambda_l x_{ijl}\}}{1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij} - \sum_{l=1}^{F} \lambda_l x_{ijl}\}} \tag{2.6}$$

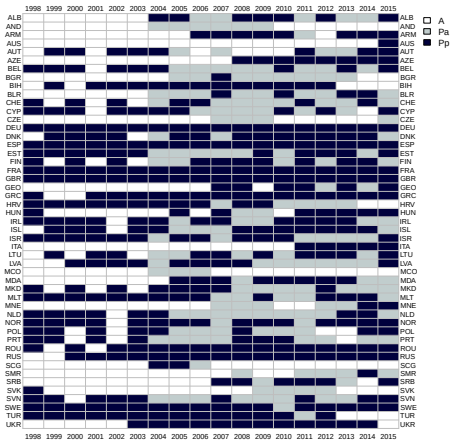
The proposal distribution to update the λ_l's is derived in the supplementary material, where we also suggest how to modify the proposal distributions for the other parameters when considering covariates.

2.6 The Eurovision song contest data

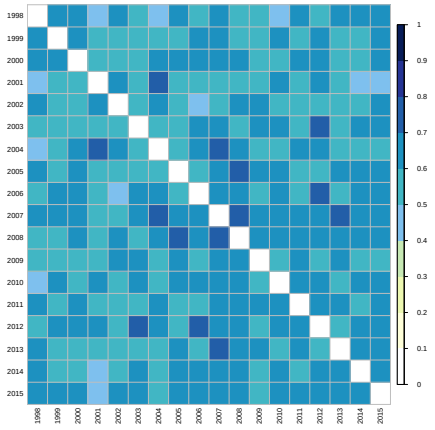
We considered 18 different editions of the Eurovison Song Contest, from 1998 to 2015. In this period, two major changes have occurred in the structure of the program, due to the growing number of countries willing to participate in the show. First, in 2004, a semi-final stage was added to select participants. In 2008, after the 50th anniversary of the competition, the event was rebuilt and two semi final stages were introduced. Countries that participate to the semi-finals are entitled to vote in the final, even if they have not qualified. Of course, the songs that do not go to the final can not compete for the title and can not receive any vote in the final. The voting structure in the final induced by the introduction of qualifying stages is modelled by the auxiliary variables $h_{ij}^{(k)}$, defined in section 2.5.2. During the period 1998-2015, a total of $n = 49$ countries took part to the competition. After 2004, on average, 14 countries were completely absent(not voting nor competing). A list of the 49 countries and their ISO3 codes is given in the supplementary material. Figures 2.3 describe some features of the 18 networks. In particular, in Figure 2.3(a) we give an overview of countries' participation per year, distinguishing the role that each country had in a given edition: absent (A), present but can not be voted (Pa) or fully present (Pp). It is easy to see from the plot that some countries, such as the UK or France, have been constantly present to the competition, while others had only made some sporadic appearances. Monaco for example competed from 2004 to 2006, but never made it to the final. Figure 2.3(b) reports the values for the association in the exchanging of votes between two different editions, measured by the index:

$$A_{(k,l)} = \frac{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = h_{ij}^{(l)} y_{ij}^{(l)})}{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = h_{ij}^{(l)} y_{ij}^{(l)}) + \sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} \neq h_{ij}^{(l)} y_{ij}^{(l)})}.$$

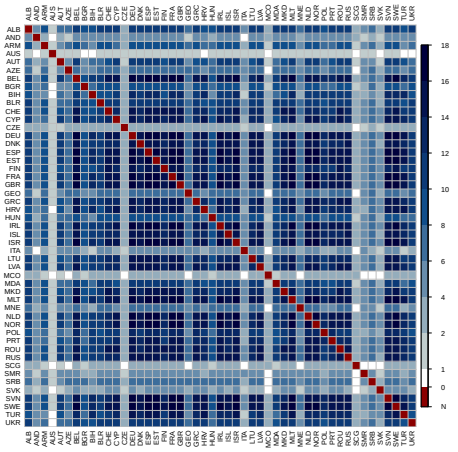
The index above is limited between 0 and 1 and the values observed for the data range from 0.4 to 0.8. However, there seems to be evidence that countries tend to repeat their patterns of votes through the analysed editions. The plots in 2.3(c) and 2.3(d) represent the number of joint participations for each couple of countries in the period 1998-2015 and the average number of votes they have exchanged while competing together. The matrix in 2.3(d) is not symmetric and the i^{th} row shows



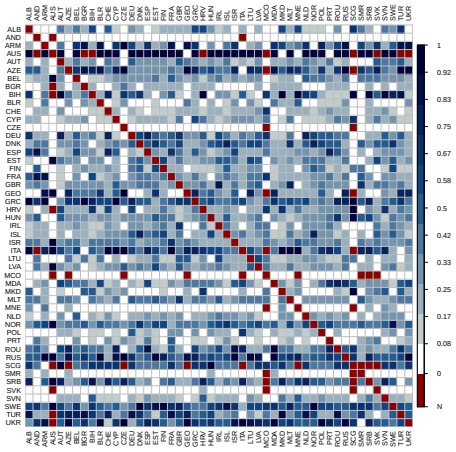
(a) Countries participation by year.



(b) Association between adjacency matrices.



(c) Number of times two countries have jointly participated to the competition, with null values in the diagonal (N).

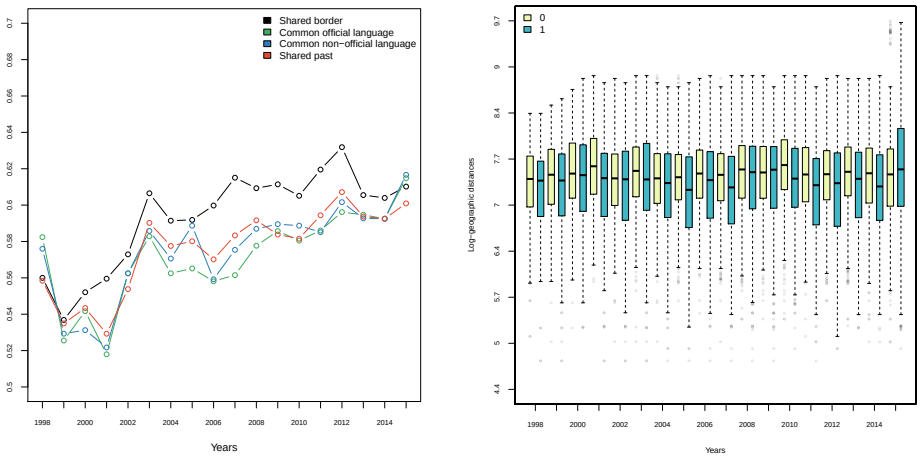


(d) Relatives frequencies of votes exchanged between couples of countries between 1998 and 2015, $\frac{\sum_k y_{ij}^{(k)}}{\sum_k h_{ij}^{(k)}}$. "N" denotes two countries that never attended the contest together.

Figure 2.3. Eurovision data: some exploratory statistics.

the average number of votes that country i gave to others. Instead, the j^{th} column reports the average number of votes that country j has received from the other participants. The last plot shows that many couples consistently voted/avoided to vote for the same group of countries, regardless of the edition.

At a second stage, covariates have been included in the analysis, similarly to what has been done by Blangiardo and Baio, 2014, Spierdijk and Vellekoop, 2006, in order to see whether the vote exchanges in the period could be, at least partially, explained by "cultural" factors. The covariates we considered are listed below:



(a) Association values for the couples adjacency matrices-covariates.

(b) Log-gegraphic distances vs observed arches, by year.

Figure 2.4. Covariates.

1. The log geographic distance between two countries ($\mathbf{X}_1$). These distances were computed using the coordinates of the centroids of each country, obtained from https://developers.google.com/public-data/docs/canonical/countries_csv; the centroids have been estimated considering latitude and longitude of the main cities of each country.

2. The presence of a border common to a couple of countries ($\mathbf{X}_2$). To maintain the characterization of the intercept, this information is coded as a binary variable that takes value 0 if there is a common border and 1 otherwise.

3. The presence of a common official language ($\mathbf{X}_3$). This information is coded as a binary variable that takes value 0 if they share the official language and 1 otherwise.

4. The fact that two countries share a major language, defined as a language spoken at least by 9% of the population ($\mathbf{X}_4$). This information is coded as a binary variable that takes value 0 if two countries share a major language and 1 if not.

5. The presence of a common past "history" shared by two countries ($\mathbf{X}_5$), they were colonized by the same country, they belonged to the same country, etc. This information is coded as a binary variable that takes value 0 if two countries share a common past and 1 otherwise.

Figure 2.4 describes the association between the covariates and the adjacency matrices. The plot in 2.4(a) displays the association between the set of binary covariates $\mathbf{X}_2$ to $\mathbf{X}_5$, measured by the index:

$$A_{(\mathbf{Y}^{(k)},\mathbf{X}_l)} = \frac{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = 1 - x_{l,i,j})}{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = 1 - x_{l,i,j}) + \sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} \neq 1 - x_{l,i,j})}.$$

The set of covariates which seems to be most relevant, when compared to the others, is the one indicating the sharing of a border ($\mathbf{X}_2$). The values for the different associations are quite constant over time, slightly increasing with the introduction of the semi-final stage. This leads to assume that the influence of the covariates on the edge probabilities is quite constant with editions. Figure 2.4(b) reports the boxplots for the couple (adjacency matrix, $\log \mathbf{X}_1$). We can not find, at least visually, no evidence of association between the distances and the presence of an arch in the adjacency matrix (that is, a vote). However, for each year, if we look at the median geographic distance for the block where an arch is present, we observe that this is usually lower than the one of the complementary block. Regressing the adjacency matrices on the log-geographic distances gives a negative estimate for every edition. That supports the claim that the geographic distances are indeed negatively correlated with the propensity to vote for a country.
Last, two sub-periods, 1998-2007 and 2008-2015, will be analysed separately, to check for large changes in the latent space position of a country according to the analysed years. That is, the analysis of the two sub-periods would give an idea on the stability of the average coordinates in the latent space recovered for the fill interval 1998-2015.
The set of covariates $\mathbf{X}_2 - \mathbf{X}_5$ have been collected from the CEPII database, http://www.cepii.fr/CEPII/en/welcome.asp, while the analysed data are available at http://eschome.net/.

2.7 Results

The following models have been considered in the analysis

1. *Model 1*: covariates not included;
2. *Model 2*: covariates $\mathbf{X}_1 - \mathbf{X}_5$ included;
3. *Model 3*: log-geographic distance included ($\mathbf{X}_1$);
4. *Model 4*: information on shared borders included ($\mathbf{X}_2$);
5. *Model 5*: no latent space ($\beta^{(k)} = 0 \ \forall k$) and covariate $\mathbf{X}_2$ included;
6. *Model 6*: random graph model (no covariates and $\beta^{(k)} = 0 \ \forall k$).

Models *5* and *6* have been estimated to test an additional layer in the modelling structure (the latent space). Each model was estimated running the MCMC algorithm for 50000 iterations, with a burn in of 5000 iterations. The intercept parameter in the first network was fixed to $\alpha^{(1)} = 0$, as $\hat{p}^{(1)} \approx 0.11$ [4] and $\beta^{(1)} = 1$. The estimated models have been compared using the Deviance Information Criterion (DIC) (see Spiegelhalter et al., 2002):

$$DIC = D(\hat{\Theta}) + 2(\bar{D}(\Theta) - D(\hat{\Theta}))$$

[4]The intercept in the reference network is fixed as

$$\hat{\alpha}^{(1)} = \log\Big(\frac{0.11}{1 - 0.11}\Big) + 2 \simeq -0.09 \approx 0$$

.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
DIC	19584.12	19494.74	19461.52	**19412.12**	23340.84	23899.62

Table 2.1. DIC values for fitted models.

Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$	Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$
1998	0	-	1	-	2007	1.15	0.15	0.92	0.14
1999	0.72	0.18	0.36	0.14	2008	1.01	0.16	0.92	0.15
2000	0.89	0.18	0.67	0.15	2009	0.66	0.17	0.49	0.14
2001	0.58	0.17	0.22	0.12	2010	0.69	0.15	0.54	0.12
2002	0.77	0.19	0.47	0.15	2011	0.84	0.17	0.70	0.15
2003	0.81	0.16	0.80	0.16	2012	0.79	0.15	0.75	0.14
2004	0.82	0.18	0.59	0.16	2013	0.76	0.16	0.73	0.13
2005	0.86	0.16	0.65	0.13	2014	0.57	0.16	0.48	0.13
2006	0.76	0.17	0.50	0.16	2015	0.91	0.16	1.01	0.17

Table 2.2. Model parameters: estimated averages and standard deviations, 1998-2015.

where $D(\Theta) = -2 \log L(\Theta)$ and $\Theta = (\mathbf{\Omega}, \mathbf{D}, \lambda)$. The deviance term $D(\hat{\Theta})$ is computed with the posterior estimates, while $\bar{D}(\Theta)$ is the mean deviance of the posterior distribution. The best model is the one with the lowest value of the DIC, see Table 2.1. According to such criteria, *Model 4* is found to be the best one, including both the latent space and the information on shared borders. The hypothesis of a random mechanism determining the exchange of votes can then be discarded in favour of a more complex solution. In fact, similarities among countries, described by distances in a latent space, play a substantial role in the formation process for observed preferences. Indeed, under *Model 4*, the scaling coefficient estimates associated to the latent space distances are quite high in each edition, see Table 2.2.

The covariate $\mathbf{X}_2$ seems to be the only one relevant in the analysis. Indeed, the estimated coefficients associated with the other edge covariates in *model 2* were all close to 0, which supports the model choice via the deviance information criterion. The Procrustes correlations among the latent space estimated under *model 1* and *model 2* is quite high, namely 0.975. That is because the matrix of covariates $\mathbf{X}_2$ acts like a fixed effect on the edge probabilities, with a decreasing effect when no common border is present between two countries. Therefore, the introduction of the set of covariates $\mathbf{X}_2$ does not seem to have a direct effect on the latent distances between the nodes. The mean of the posterior estimate for the effect λ associated with the border covariate is 0.60 with a standard deviation of 0.10.

Figures 2.5, 2.6 and 2.7 show the estimates obtained under *model 2* for the latent positions, the distances and the posterior distribution for the parameters of interest. Figure 2.5*a* reports the posterior means for the country latent coordinates (reported with their ISO3 codes, see the supplementary material) together with the corresponding standard deviations. Note that the model does not necessarily place in the center of the latent space those countries that have been most successful throughout the editions. Indeed, for example, Sweden and Denmark won the largest number of titles in the analysed period, respectively 3 and 2 titles. However, Denmark is not in the origin, but it is rather placed close to a group of countries from northern Europe. If we look at a specific country, its neighbours on the latent space are

the countries that were estimated to be more similar in *tastes*, expressed in terms of voting exchange patterns. The latent space presents a number of denser zones, that partly resemble northern Europe, eastern Europe and North Eastern Europe. However, these subgroups are not completely faithful to the geographic locations of countries, as, for example, Spain is closer to Romania than to Portugal or France. We should notice that Romanians defines one of the major immigrant group in Spain, as well as Latvians is one of the largest immigrant group in Ireland. Therefore, some of the *geographical misplacements* within the sub-groups in the latent space may be also explained in terms of migration flows. However, the message is that geographical locations can not fully explain the observed votes exchange. Indeed, Figures 2.6 and 2.8 show the presence of large differences when we compare estimated and geographical distances. In particular, for a given country i, the rows of the matrices in Figure 2.8 represent the intersection between its r nearest neighbours in the latent space (we denote this set as $LN_{i,r}$) and its r closer neighbours in terms of geographical distances ($GN_{i,r}$); we consider the values $r = 1, 2, 3, 5, 10, 15$. Given the number of neighbours r, the average[5] and the maximum number of common neighbour countries is reported in Table 2.3. The table confirms what was already visible from Figure 2.8: there is weak association between the coordinates in the latent space and the geographical ones. A similar comparison can be made with the information on the shared border. Given a country i, let us define r_i^* the number of bordering countries, LN_{i,r_i^*} the r^* nearest neighbours in the latent space and CN_{i,r_i^*} the set of bordering countries of node i. The average number of geographical bordering countries that are also neighbouring in the latent space[6] is 0.11, where $\hat{r}^* \approx 4$. This low association between bordering countries and closest countries in the latent space confirms that $\mathbf{X}_2$ is only partially relevant to the description of votes' exchange in the contest. Figure 2.9 reports the matrix of the intersections between the sets of neighbours LN_{i,r_i^*} and CN_{i,r_i^*}, for 1998-2015. Bulgaria, Lithuania and Serbia and Montenegro are the countries that tend to vote more their bordering countries. Indeed, their closest countries in the latent space are often countries with which they share a border ($|LN_{i,r_i^*} \cap CN_{i,r_i^*}|/r_i^* \geq 0.5$). In general, there is no strong association between the presence of a border and the closeness in the latent space. The estimated values for the network intercept parameters are quite similar for the different networks corresponding to the editions in the period 1998-2015 (Figure 2.7). Indeed, the voting rule (in the Eurovision song contest) for that period required that participating countries vote for exactly 10 others, which implied a fixed outdegree for each node in the corresponding networks. The observed densities are then quite similar and this is reflected in similar estimates for the $\alpha^{(k)}$ parameters, which define the upper bound for the edge probabilities in a given network.
The estimated values for the posterior means of the network-specific scaling pa-

[5] the average number of common neighbours is given by:

$$\frac{\sum_{i=1}^{n} |LN_{i,r} \cap GN_{i,r}|}{rn}$$

[6] this average is given by:

$$\frac{\sum_{i=1}^{n} |LN_{i,r_i^*} \cap CN_{i,r_i^*}|}{r_i^* n}$$

	$r=1$	$r=2$	$r=3$	$r=5$	$r=10$	$r=15$
average number	0.06	0.11	0.14	0.21	0.37	0.47
maximum number	1	1	2	3	8	12

Table 2.3. Average and maximum number of intersections of the set of the closest latent positions and the closest geographical positions.

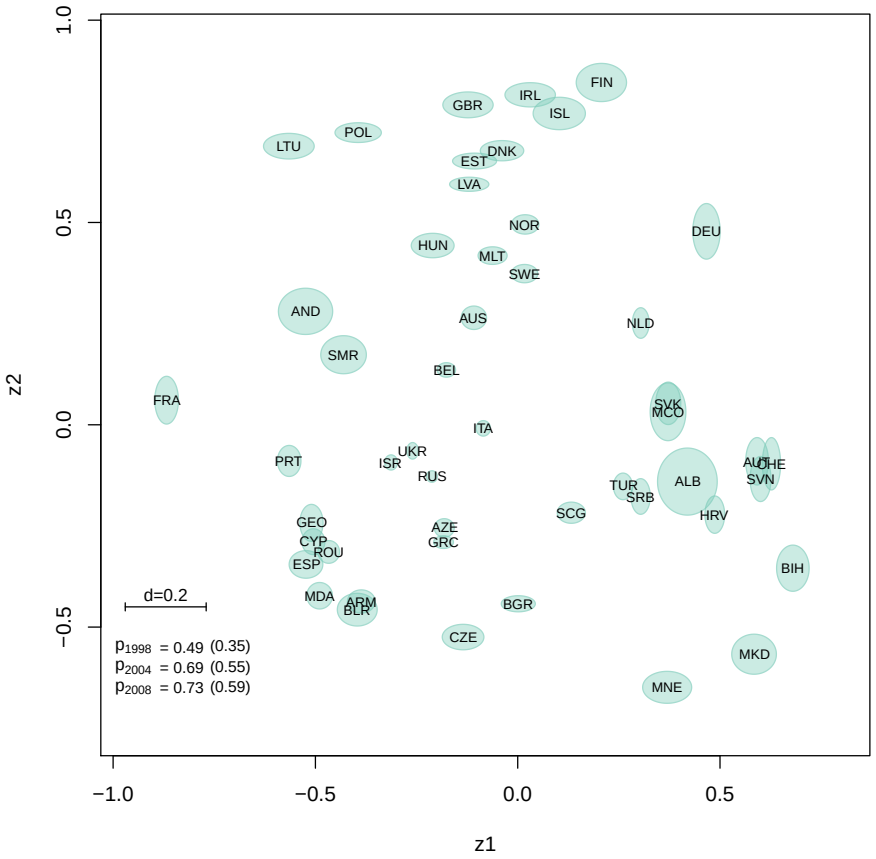


Figure 2.5. Estimated latent positions 1998-2015. The legend reports the probabilities corresponding to a distance of 0.2 in the latent space, years 1998, 2004 and 2008. The values refer to the case of $x_{2,ij} = 0$, within the brackets are reported the values for $x_{2,ij} = 1$.

rameters $\beta^{(k)}$ range from 0.22 in year 2001, to 1.01 in year 2015. As none of these parameters is estimated to be 0, the latent space is found to always play a role in the formation of observed networks. However, its influence depends on the dimension/magnitude of the scaling parameter $\beta^{(k)}$. In 2001 this role is quite limited, as it is in 1999. In the last network the influence of the latent space is the greatest ($\beta^{(18)} = 1.01$), and it is similar to the one in the first edition ($\beta^{(1)} = 1$). We have estimated the same model with different networks set as reference, and no substantial changes were observed in the pattern of the estimated scaling coefficients. Also, the estimated latent spaces were highly correlated with the one presented here. The supplementary material reports the results for the analysis of the two

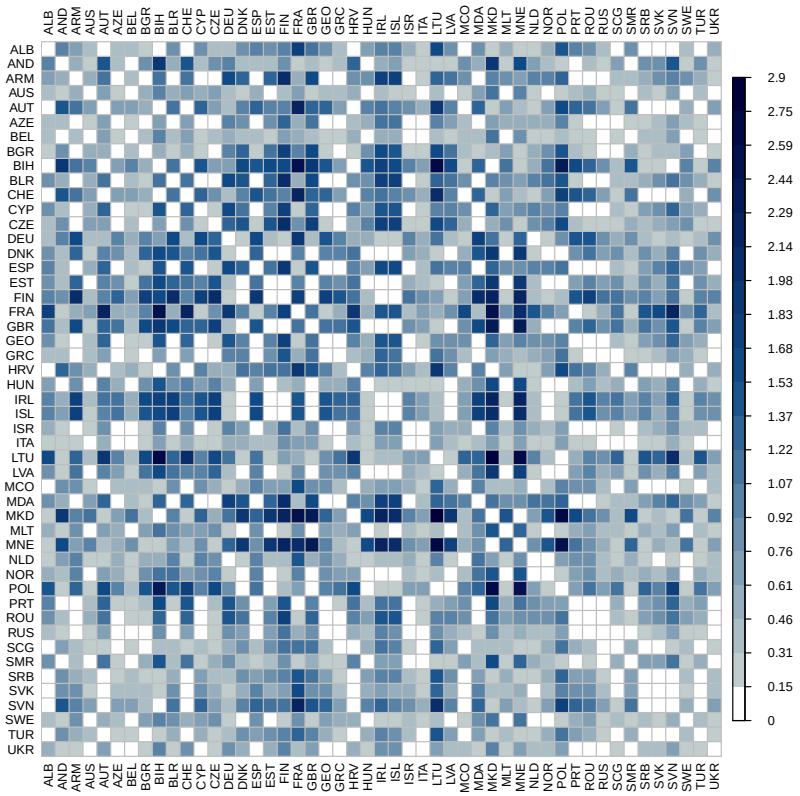


Figure 2.6. Estimated distances between couple of countries, 1998-2015.

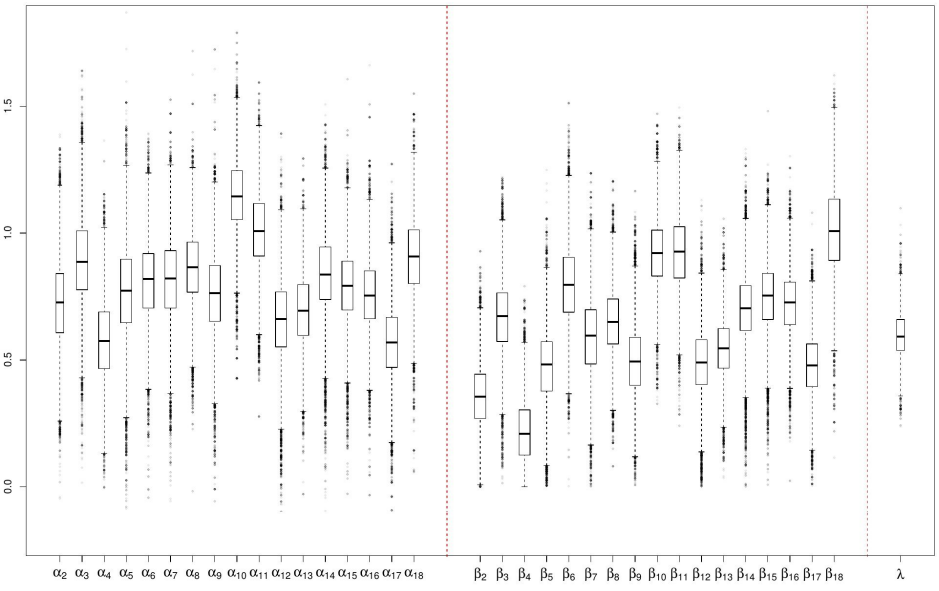
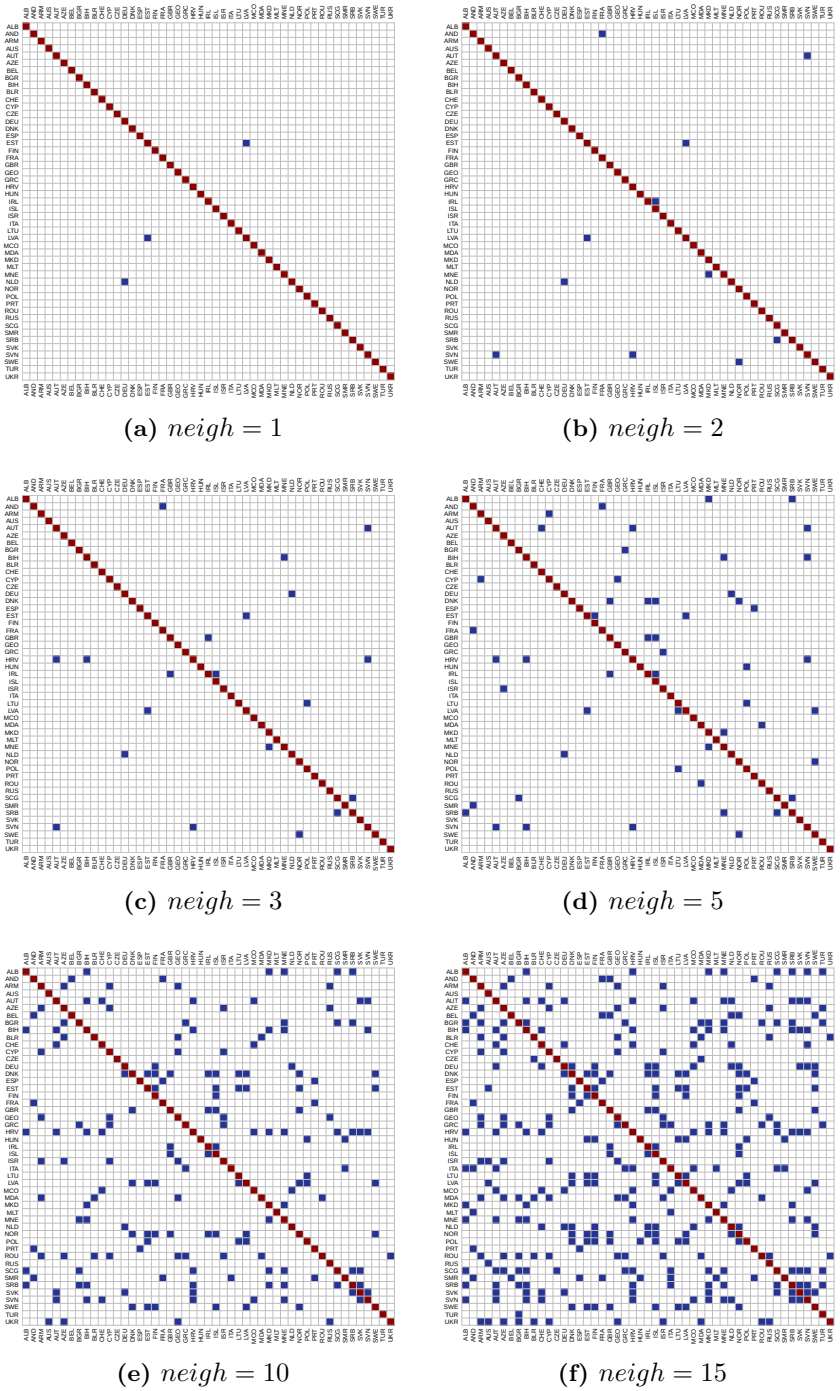


Figure 2.7. Boxplots for model parameter estimates and the coefficient for $\mathbf{X}_2$, 1998-2015.

sub-periods 1998-2007 and 2008-2015. The model considered for the sub-periods does not include any covariates, as the interest lays primarily in recovering the latent



(a) $neigh = 1$ **(b)** $neigh = 2$

(c) $neigh = 3$ **(d)** $neigh = 5$

(e) $neigh = 10$ **(f)** $neigh = 15$

Figure 2.8. Intersections of the set of neighbours $LN_{i,r}$ and $GN_{i,r}$, 1998-2015

coordinates. The findings confirmed a weak correspondence between the estimated latent position of a country and its actual latitude and longitude coordinates on the globe. A direct comparison of the latent space estimated for 1998-2015 to that estimated for 1998-2007 and 2008-2015 is not available, as the number of countries

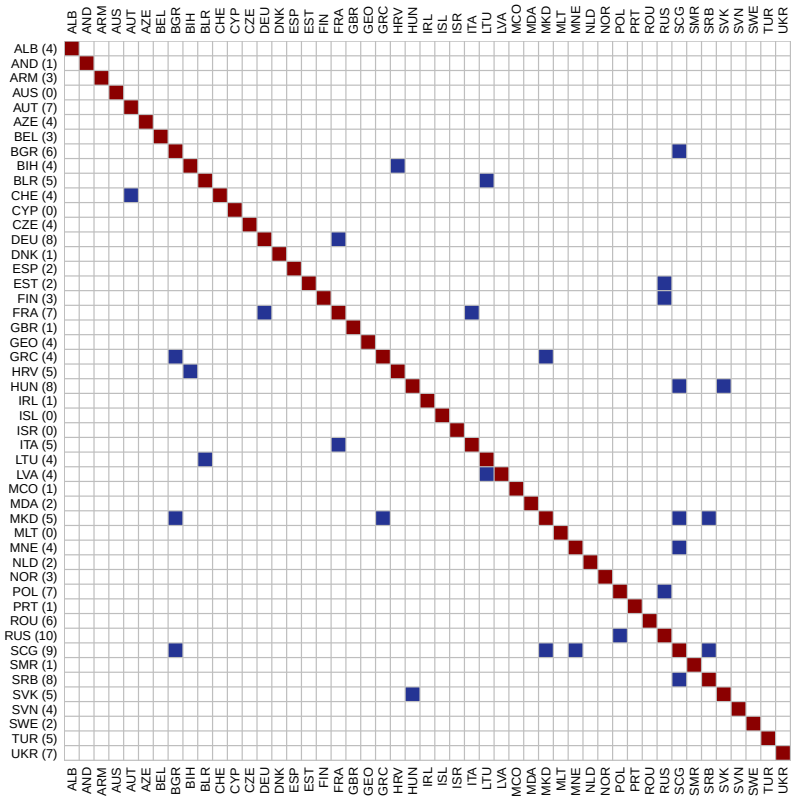


Figure 2.9. Intersection of the set of neighbours LN_{i,r_i^*} and CN_{i,r_i^*}, 1998-2015. On the left column, in brackets, are reported the values of r_i^*.

is different. Indeed, not all the participants in 1998-2015 were also participant in both of the sub-periods. Just to give an example, Italy rejoined the competition in 2011, after being absent in the period 1998-2007. The multidimensional networks for the sub-periods can not include the same set of nodes of the period 1998-2015. Indeed, countries completely absent from the competition, in a given sub-period, correspond to isolated nodes, and the pairwise distances from present countries are not be identifiable, as they are potentially infinite. Also, removing present countries to match the node set of the two sub-periods is not a valid option, as it would alter the voting structure. Although some of the distances vary with respect to those in the longer period 1998-2015, the sub-groups observed in figure 2.5 are still present. For example, Northern Europe countries tend always to be closer to each other, as well as Eastern Europe countries do.

2.8 A simulation study

A simulation study has been considered to test the proposed model. In particular, simulations have been exploited to assess the large-sample behaviour the parameters estimates, when the dimension K of the multiplex is large, and to verify the robustness in the latent coordinates estimates with respect to different underlying distributions. In all the different scenarios, the reference was taken to be the first network of the

multiplex and the reference parameters have been fixed to $\beta^{(1)} = 1$ and $\alpha^{(1)} = 0$ (as in the application). The intercepts and the scaling coefficients have been simulated from their prior distributions (see section 2.4.1), with $\sigma^2_\alpha = \sigma^2_\beta = 1$, $\mu_\alpha = \mu_\beta = 0$. Four simulation scenarios have been defined, divided in two blocks:

- **Block I** This block has been built to test the large-sample behaviour and the robustness of parameter estimates when the latent coordinates distribution is far from Gaussianity. Indeed, not all the observed data might be well described by latent Gaussian coordinates. Within each scenario in the block, we have considered 4 types of multidimensional networks, with a relatively small K but an increasing number of nodes:

  1. $n = 25$ and $K = 3$,
  2. $n = 50$ and $K = 3$,
  3. $n = 50$ and $K = 5$,
  4. $n = 100$ and $K = 3$.

  The values for the scaling parameters and the intercepts are constants in scenarios I-III, conditionally on the type of multiplex considered.

  - **Scenario I**: the latent coordinates have been simulated from a bivariate normal distribution (the prior distribution used in the model).
  - **Scenario II**: the latent coordinates have been simulated from a mixture of bivariate normal distributions, where the number of components was set to $G \approx n/7$. The mean vector for each group has been simulated from a standard bivariate normal distribution and the covariance matrices are diagonal with elements randomly sampled in the interval $(0.1, 1)$. This scenario corresponds to the case of data representing different kind of relations among separate groups of nodes/communities. Indeed, the probability for node i in group c to link with node j will be higher if $j \in c$ as well.
  - **Scenario III**: the latent coordinates have been simulated from a standard bivariate Hotelling's T^2 distribution with 4 degree of freedom. This scenario allows for some nodes to be located far from the center in the latent space. Thus, this case reflects the presence of inactive/semi-inactive nodes in the network, that tend to interact poorly with the rest of the network.

- **Block II** This block has been built to test the large-sample properties of the parameter estimates when the number of networks K is large. That is, the case of the application considered in the present application.

  - **Scenario IV**: the latent coordinates are simulated according to the bivariate Gaussian distribution specified in section 2.4.1. Multiplexes with different size have been simulated:

    1. $K = 10$ and $n = 50$,
    2. $K = 20$ and $n = 50$,

3. $K = 30$ and $n = 50$.

In all the considered scenarios, we treated both the case where all the nodes are present in each network (denoted by P) and the case where some of the nodes are absent in some networks (denoted by A) (see section 2.5.2). In the second case, the missing data process resembles the one observed in the Eurovision data with respect to the average number of absent nodes per network and the number of absences for each node. The reference network was taken to be the first of the multiplex and the corresponding parameters have been fixed to $\beta^{(1)} = 1$ and $\alpha^{(1)} = 0$ (the values in the application). The latent coordinates, the intercepts and the scaling parameters have been simulated from their prior distributions (see section 2.4.1), with $\sigma_\alpha^2 = \sigma_\beta^2 = 1$, $\mu_\alpha = \mu_\beta = 0$.

2.8.1 Results

To estimate the model on the simulated data (**Block I** and **Block II**), we fixed $\nu_\alpha = \nu_\beta = 3$, $\tau_\alpha = \tau_\beta = (K-1)/K$, $\alpha^{(1)} = 0$ and $\beta^{(1)} = 1$ (see section 2.4.2 for details). Each model was estimated 10 times, performing 40000 MCMC iterations and discarding the first 5000. The parameter estimates were consistent with the simulated values in all different scenarios and the estimates for the latent coordinates have been found to be robust to misspecifications of the corresponding distribution. In the supplementary material we present in details the results for the different scenarios. Boxplots and tables with mean and standard deviations for the parameter estimates are presented, as well as mean and standard deviation of the procrustes correlation between the estimated and the simulated latent space coordinates. Overall, the proposed method returns reliable estimates for the true parameter values. The simulated values fall within the 95% credible interval built on the posterior distributions, with a couple of exceptions that occur when the number of networks increases. However, in these cases, the true magnitude of the value is always recovered, as the estimates are still quite close to the actual simulated values. The simulated latent spaces are always recovered with high correlation, even in the stressed scenarios, I and II. There is no substantial difference in estimates for cases P and A. Thus, the absence of some of the nodes in some networks of the multiplex does not impact the estimation of the latent position estimates.

2.9 Comparison with the *lsjm* model

The lsjm model by Gollini and Murphy, 2016 considers a mean latent space, which originates the network-specific latent coordinates. To compare the two models, we have simulated different types of multiplex, fixing $\beta^{(k)} = 1$ for $k = 1, \ldots, K$, that is, according to the *lsjm*. The simulated multidimensional networks come from the following scenarios:

1. A sparse multiplex with $n = 50$, $K = 3$ and $\alpha = (-0.66, -0.70, -0.54)$,
2. A multiplex with $n = 70$, $K = 2$ and $\alpha = (-0.73, -1.12)$,
3. A small, denser multiplex with $n = 25$, $K = 3$ and $\alpha = (0, 1.02, 0.28)$.

Simulating from the model presented in this work (which will be referred to as *lsmmn*) when all the coefficients are fixed to 1 corresponds to simulating from the latent space joint model when all the network-specific latent spaces are the same. Therefore, in the present setting, the two models can be compared. Both the *lsjm* and the *lsmmn* models have been estimated 10 times for each of the simulated multidimensional networks. The *lsjm* is estimated using the R package *lvm4net* (https://cran.r-project.org/web/packages/lvm4net). In the supplementary material we report the results of such a comparison. As it can be derived by looking at the supplementary material, the model we propose outperforms the *lsjm* both in the quality of parameter estimates and in recovering the latent coordinates.

2.10 Discussion

In the present work, we have introduced a general and flexible model for the analysis of multidimensional networks (multiplexes). In particular, the model is defined to recover similarities among the nodes when the structure of the observed networks is complex and there is not a clear information on which external information can be used to explain the observed patterns. The model extends the latent space model by Hoff, Raftery, and Handcock, 2002 and the latent space joint model by Gollini and Murphy, 2016 with the introduction of network-specific scaling parameters representing the impact of the latent space on the edge-probabilities. When these coefficients are null, the model reduces to a random graph model for multidimensional networks. Moreover, missing data and edge-specific covariates are considered. A hierarchical Bayesian approach is employed to define the model and its estimation is carried out via MCMC. We have defined hyperprior distributions for the hyperparameters of the model, to avoid subjective specifications. The latent coordinates allow for an efficient visualization of the network, a well desired feature for large multidimensional networks.
The model has been applied to the votes exchanged among countries in a popular TV show, the Eurovision Song Contest, from 1998 to 2015. Cultural and geographical covariates have been included in the analysis and only the presence of a shared boarder between two countries was found to be relevant to explain observed voting patterns. The recovered similarities among the participants in the period 1998-2015 have been found to resemble only partially the corresponding geographical locations. Indeed, exploratory analysis displays a group structure among the nodes in the latent space which does not completely agree with geographical criteria. These findings sustain the claim of bias in the voting structure observed in the Eurovision, which, however, can not be attributed to geographical reasons alone.
In the simulation study we have applied the proposed model to a large varieties of multidimensional networks and successfully recovered the latent coordinates and the network-specific parameters. That has proved the ability of the model to recover the (latent) association structure among the nodes in a multiplex, also when the number of networks is large.
The latent space model for multivariate networks is implemented in the *R* package *spaceNet* and it is available on *CRAN* (https://CRAN.R-project.org/package=spaceNet).

Chapter 3

Modelling Heterogeneity in Latent Space Models for Multidimensional Networks

3.1 Introduction

Relational data can be, and often are, represented in the form of networks. In an observed network, dyadic relations of interest are coded as edges between nodes. When multiple relations are recorded among the same group of nodes, a multidimensional network (multiplex) is observed. Instead, if the same relation is observed through time on the same set of nodes, a dynamic network can be defined. Observed network data, either uni-dimensional or multidimensional, can exhibit different characteristics, and these may directly influence their structure. A frequently studied one is *transitivity*, which refers to whether the relation being represented in the network is, up to some extent, transitive. Roughly speaking, transitivity in social networks can described by the "a friend of my friend is my friend" phenomenon. A popular way to model such a feature is through latent space models, first introduced by Hoff, Raftery, and Handcock, 2002. The basic idea is to represent the nodes in a low-dimensional, unobserved, space, postulating that the probability of observing an edge in the network depend on the positions of the nodes in such a space. Different latent space approaches have been proposed in the literature, either based on metrics, see Hoff, Raftery, and Handcock, 2002, Hoff, 2005, or on ultrametrics, see Schweinberger and Snijders, 2003. Among metric latent spaces, models based on the Euclidean distance are likely the most widespread, as they produce easily interpretable representations of the networks while being flexible enough to describe a large variety of network data. Alternatives to distance-based latent space model are the projection, Hoff, Raftery, and Handcock, 2002, and the multiplicative latent space model, Hoff, 2005. Both postulate that the edge probabilities depend on the inner products of the latent coordinates. Such models address transitivity differently from distance models, as they also incorporate in the latent space representation of the nodes also the "direction" of the relation, see Hoff, Raftery, and Handcock, 2002. Extensions of the latent space models to multidimensional networks are those described by Gollini and Murphy, 2016,

Salter-Townshend and McCormick, 2017, Hoff, 2011 and D'Angelo, Murphy, and Alfò, 2018. Sewell and Chen, 2015 introduce a latent space model for dynamic binary networks, and later have extended it to include weighted dynamic networks, see Sewell and Chen, 2016. Durante and Dunson, 2016 developed a framework based on a dynamic latent space model to describe dynamic networks of face-to-face individual contacts.

Another interesting feature is that of *degree heterogeneity*, which refers to the propensity of some nodes to send/receive more edges than others. Holland and Leinhardt, 1981 proposed the so called "p_1" model, where node-specific sender and receiver effects are treated as random effects. Duijn, Snijders, and Zijlstra, 2004 developed the "p_2" model, an extension of the "p_1" model, where node-specific attributes are introduced in the form of covariates, together with the sender/receiver random effects. Other extensions of the basic model are that by Hoff, 2003 and Hoff, 2005, that bring together sender/receiver effects and latent space representations for a single network. Part of this framework was later extended by Krivitsky et al., 2009, to allow for clustering of the nodes in the latent space. In the context of dynamic networks, Sewell and Chen, 2015 model the overall sender/receiver effect in the networks, investigating whether activity (sending links) or popularity (receiving links) is more important when the edge formation process is considered.

Starting with latent space models, we develop a latent space approach based on the Euclidean distance to model transitivity and heterogeneity in multidimensional networks. Our aim is to extend the model by D'Angelo, Murphy, and Alfò, 2018 to account for degree heterogeneity while modelling different levels of complexity in multidimensional networks. Indeed, multidimensional networks data are complex in two directions: the number of nodes and the number of views. A model that aims at describing the interactions between the actors in such a high dimensional complex should explain the view-specific features, while being parsimonious with respect to the number of parameters. For this purpose, we model transitivity via a single latent space, common to the whole multiplex, assuming that the distances in such a space represent the overall association between the nodes. Heterogeneity across different view will be addressed introducing node-specific sender/receiver parameters, that will account both for intra(-) and inter(-) networks degree heterogeneity.

The paper is organized as follows: in Section 3.2 we introduce more formally the concept of multidimensional networks and define the proposed class of models for directed multiplexes. Section 3.3 provides with the details of the estimation procedure and discusses issues of model identifiability. Section 3.4 describes the proposed class of models for the particular case of undirected multidimensional networks. A simulation study is conducted is Section 3.5, to investigate the properties of the proposed class of models, and to study the performance of a novel heuristic procedure for model selection. Then, an application on FAO trade data is presented in Section 3.6. Last, we conclude with a discussion in Section 3.7.

3.2 The collection of latent space models for directed multiplexes

A binary multidimensional network (multiplex) $\mathbf{Y}$ is a complex object defined by a collection of networks (also known as views). These networks can be represented by $n \times n$ adjacency matrices $\{y^{(1)}, \ldots, y^{(K)}\} = \mathbf{Y}$, where the index $k = 1, \ldots, K$ indicates the different views. The entries in each matrix (k^{th} network) can take two values, $y_{ij}^{(k)} = 1$ when nodes (i, j) are joined by an edge, and $y_{ij}^{(k)} = 0$, when they are not. Views in a multiplex share the same set of nodes, whose cardinality is denoted by n. In the present context, nodes will be indexed by $i, j = 1, \ldots, n$. A multiplex can be either undirected, if $y_{ij}^{(k)} = y_{ji}^{(k)}$ or directed, if the different adjacency matrices are not symmetric, $y_{ij}^{(k)} \neq y_{ji}^{(k)}$ for at least one (i, j) couple. In general, many interesting real world multiplexes are directed, and different levels of "symmetry" can be observed in the adjacency matrices at hand. Notice that, even if a network is undirected, this does not imply that all the nodes have the same number of connections, that is, the same degree. Modelling the degree, or, for directed networks, the out-degree, $\sum_{j \neq i}^{n} y_{ij}^{(k)}$, and the in-degree, $\sum_{j \neq i}^{n} y_{ji}^{(k)}$, is a task that might be of interest in many empirical applications. Indeed, it can help recover the most influential, or popular, nodes in a network, or the more active ones. Further, different views might exhibit different levels of heterogeneity in the node-specific degree distribution. In this sense, multidimensional networks can be heterogeneous in two directions: within and between the views. In the present work, we introduce a class of latent space models that address transitivity and view-specific heterogeneity in the analysis of multiplex data. In this section, we introduce the more general latent space framework for directed multidimensional networks. Section 3.4 discusses their restriction to the particular case of undirected multidimensional networks.
Latent space models based on Euclidean distances are based on the assumption that each node is located into an unobserved p-dimensional Euclidean space; according to model specification, the probability of observing an edge between the dyad (i, j), conditionally on the latent coordinates $\boldsymbol{z}_i$, $i = 1, \ldots, n$, of these nodes, does not depend on the other nodes positions, see Hoff, Raftery, and Handcock, 2002. We hold these assumptions in our model, together with a further one, as we assume the probability of a connection between a dyad also depends on its node-specific propensities to send/receive links. In multidimensional networks, the propensities may vary with the views, as an actor could be quite popular in a network, without receiving many edges in another one. Different levels of heterogeneity in edge probabilities may depend on different levels of heterogeneity in the behaviour of the nodes in the different networks.
For this purpose, let $\theta_i^{(k)}$ and $\gamma_i^{(k)}$, $i = 1, \ldots, n$ and $k = 1, \ldots, K$ represent the sender and the receiver parameters for the i^{th} node in the k^{th} network, respectively. These parameters are introduced in the model specification to describe the propensity of a given node to send/receive edges, respectively. Then, the probability $p_{ij}^{(k)}$ of a connection from node i to node j, in the k^{th} network, depends on the parameters

$(\theta_i^{(k)}, \gamma_j^{(k)})$, through the following model:

$$p_{ij}^{(k)} = P\Big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}, d_{ij}\Big) = \frac{\exp\{f(\alpha^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}) - \beta^{(k)} d_{ij}\}}{1 + \exp\{f(\alpha^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}) - \beta^{(k)} d_{ij}\}} \tag{3.1}$$

where $\boldsymbol{\alpha} = (\alpha^{(1)}, \ldots, \alpha^{(K)})$ and $\boldsymbol{\beta} = (\beta^{(1)}, \ldots, \beta^{(K)})$ are the sets of network-specific parameters, while d_{ij} is the squared Euclidean distance between node i and node j in the p-dimensional latent space.
According to the node-specific behaviours, we may define three different scenarios for each parameter:

- *null (N)*: $\theta_i^{(k)} = 0$ or $\gamma_i^{(k)} = 0$, $\forall\ i = 1, \ldots, n\ ,\ k = 1, \ldots, K$;
- *constant (C)*: $\theta_i^{(k)} = \theta_i$ or $\gamma_i^{(k)} = \gamma_i$, $\forall\ i = 1, \ldots, n,\ k = 1, \ldots, K$;
- *variable (V)*: $\theta_i^{(k)}$ or $\gamma_i^{(k)}$, $\forall\ i = 1, \ldots, n,\ k = 1, \ldots, K$.

Note that we assume the same type of effect (null, constant or variable) for all the views in the multiplex; hence, if, for example, nodes in the k^{th} network are assumed to have a constant receiver effect, all the other $(K-1)$ networks will have the same effect. While this assumption may seem a stringent one, we may observe that, in practice, assuming $\theta_i^{(k)}$ and $\gamma_i^{(k)}$ are variable we may have some nodes with null effects, others with constant effects and the remaining with variable effects. We further discuss this assumption in Section 3.7. Table 3.1 presents a schematic taxonomy of the 9 potentially different models arising from the different assumptions on the sender and receiver effects.
The impact of the sender/receiver effects on the edge probabilities can be made explicit by defining a collection of network-specific matrices $\Phi = (\Phi^{(1)}, \ldots, \Phi^{(K)})$, with generic element defined by

$$\left[\phi_{ij}^{(k)}\right] = \left[\frac{\theta_i^{(k)} + \gamma_j^{(k)}}{g}\right], \tag{3.2}$$

where the auxiliary variable g is defined as follows:

$$g = \begin{cases} 0 & \text{if both effects are absent,} \\ 1 & \text{if only one effect is present,} \\ 2 & \text{if both effects are present.} \end{cases} \tag{3.3}$$

			$\gamma_j^{(k)}$		
			0	γ_j	$\gamma_j^{(k)}$
			N	C	V
$\theta_i^{(k)}$	0	N	NN	NC	NV
	θ_i	C	CN	CC	CV
	$\theta_i^{(k)}$	V	VN	VC	VV

Table 3.1. The class of models defined by the different assumptions on the sender/receiver effects.

We can define the function $f(\cdot)$ as:

$$f(\alpha^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}) = f\Big(\alpha^{(k)}\phi_{ij}^{(k)}\Big) = \Big(1 - \mathbb{1}_g(0)\Big)\alpha^{(k)}\phi_{ij}^{(k)} + \mathbb{1}_g(0)\alpha^{(k)}. \tag{3.4}$$

where

$$\mathbb{1}_g(0) = \begin{cases} 1 & \text{if} \quad g = 0 \\ 0 & \text{if} \quad g \neq 0 \end{cases} .$$

Equations (3.1), (3.2) and (3.4) explicit the basic modelling assumption; within each view, the sender and the receiver effects jointly impact the view-specific intercept. Furthermore, if we assume that

$$\gamma_j^{(k)}, \theta_i^{(k)} \sim Unif(-1, 1), \quad i, j = 1, \ldots, n$$

we have that, differently from the additive sender and receiver effect specification, see Krivitsky et al., 2009, and as in standard latent space models, the intercepts $\alpha^{(k)}$ still correspond, on the logit scale, to the maximum value that edge probabilities in the networks may achieve. Thus, recalling the definition of $\phi_{ij}^{(k)}$ and of g in equations (3.2) and (3.3), we may notice that the sender and receiver effects can be considered as relative. Indeed, inactive (or unpopular) nodes will have a value of the effect close to -1, while active (or popular) nodes will have a value close to 1. Obviously, the combined effect $\phi_{ij}^{(k)}$ varies in the same range. Bounding these parameters in the interval $[-1, 1]$ allows to easily interpret the differences in levels of activity (or popularity) across nodes. We must also notice that if we allow the $\phi_{ij}^{(k)}$ to be negative and $\alpha^{(k)} < 0$, a fundamental problem arises. In fact, if we consider two dyads with a common node, say (i, j) and (i, l) with $\phi_{ij}^{(k)} < \phi_{il}^{(k)} < 0$, we obtain $p_{ij}^{(k)} > p_{il}^{(k)}$ and this would violate the assumption that sender and receiver effects are directly proportional to edge probabilities. For this reason, we bound the intercept to be non-negative, $\alpha^{(k)} \geq 0 = LB(\alpha)$, $k = 1, \ldots, K$. This constraint does not alter the interpretation of the intercept and other model parameters; fixing a lower bound for the intercept does not imply a lower bound for the edge probabilities, as the impact of the latent space might decrease the effect of the intercept and the sender/receiver effects.

From equation (3.4), we may also notice that when no sender/receiver effects are present, the edge probability in (3.1) reduces to the model specification by D'Angelo, Murphy, and Alfò, 2018 (scenario NN in Table 3.1). For the NN scenario, inference procedures have already been provided; therefore, in the next section, we will focus on those scenarios that include at least one effect. For these models, the edge probability formula presented in equation (3.1) can be rewritten as:

$$p_{ij}^{(k)} = \frac{\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij}\}}{1 + \exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij}\}} \tag{3.5}$$

3.3 Estimation

We propose a hierarchical Bayesian approach to parameter estimation for the latent space model proposed in section 3.2. The corresponding (log-)likelihood can

be derived from equation (3.5),

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \Phi, D \mid \mathbf{Y}) = \sum_{k=1}^{K} \sum_{\substack{i=1 \\ j \neq i}} \ell_{ij}^{(k)} = \sum_{k=1}^{K} \sum_{\substack{i=1 \\ j \neq i}} y_{ij}^{(k)} \log(p_{ij}^{(k)}) + (1 - y_{ij}^{(k)}) \log(1 - p_{ij}^{(k)}). \quad (3.6)$$

The prior distributions for model parameters can be specified as follows:

$$\beta^{(k)} \sim N_{(0,\infty)}\left(\mu_\beta, \sigma_\beta^2\right), \quad \alpha^{(k)} \sim N_{(0,\infty)}\left(\mu_\alpha, \sigma_\alpha^2\right),$$
$$\boldsymbol{z}_i \sim MVN_p\left(0, I\right), \quad \gamma_j^{(k)}, \theta_i^{(k)} \sim Unif(-1, 1).$$

Since $\mu_\beta, \mu_\alpha, \sigma_\beta^2, \sigma_\alpha^2$ are nuisance parameters whose specification could be relevant, we introduce an extra layer of dependence using the following (hyper) prior distributions:

$$\mu_r \mid \sigma_r^2 \sim N_{(0,\infty)}\left(m_r, \tau_r \sigma_r^2\right), \quad \sigma_r^2 \sim \text{Inv}\chi_{\nu_r}^2,$$

with $r = \{\alpha, \beta\}$. The hyperparameters $m_\alpha, m_\beta, \tau_\alpha, \tau_\beta, \nu_\alpha, \nu_\beta$ have to be specified by the user. The constraint $\beta^{(k)} \geq 0$, $k = 1, \ldots, K$ is imposed according to the assumption that edge probabilities are inversely proportional to the distance between nodes in the latent space. Figure 3.1 provides a representation of the hierarchical structure of the proposed model.

3.3.1 Identifiability

As discussed by D'Angelo, Murphy, and Alfò, 2018, to ensure parameter identifiability in the basic latent space model, one out of the K parameters $\alpha^{(k)}$ and $\beta^{(k)}$ must be fixed. The corresponding network is then referred to as the "reference" network. In this context, fixing these two parameters is not enough, as the multiplicative effect of $\phi_{ij}^{(k)}$ can still cause problems. To avoid such issues, if both effects are present, one sender and one receiver effect should be fixed in each view, and the corresponding nodes will be considered as "reference" nodes. However, due to the possible scenarios for each parameter, we have to give some more details. When the effect is variable, we propose to choose as reference, in each network, the node (i) with the highest observed out-degree (or in-degree) and fix $\theta_i^{(k)} = 1$ (respectively $\gamma_i^{(k)} = 1$). Instead,

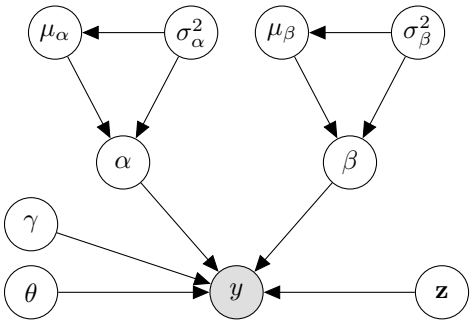


Figure 3.1. Hierarchy structure of the model.

when the effect is constant, we propose to select the node i with the highest observed mean out-degree (or mean in-degree) and fix $\theta_i = 1$ (respectively $\gamma_i = 1$). Fixing the sender/receiver parameter for one node to 1 does not change the interpretation of the model, as we are most interested in ordering the nodes with respect to the sender/receiver effects rather than in the parameter estimates.

3.3.2 MCMC algorithm

We propose a MCMC algorithm to estimate parameters for the latent space model proposed in section 3.2. The algorithm iterates over the model parameters, the latent coordinates and the nuisance parameters; since full conditionals are available in closed form only for the latter ones, we use a Metropolis-Hastings step to update the other estimates. The full conditional distributions for the nuisance parameters, the proposal distributions for the network-specific parameters and the latent coordinates are presented in appendix B, together with the proposal distributions for the sender/receiver parameters. The adopted procedure starts by simulating a new value for each nuisance parameter; then, it proposes a new value for the network-specific parameters $\alpha^{(k)}$ and $\beta^{(k)}$, with a joint MH step on each network. A further MH step sequentially proposes new latent coordinate values. As the likelihood in equation (3.6) is invariant to rotations and translations of the latent coordinates, when a new set of positions is defined, Procrustes transformation is employed to check whether this new set is just a simple transformation of the previous solution. If so, the proposed set is discarded in favour of the previous one. After that, sender/receiver parameters are updated sequentially on the nodes, but jointly over the different networks via an additional MH step. The joint update is performed to speed up the calculations, given the high number of model parameters. When all the effects are updated, a new solution of $\Phi^{(k)}$ is available.
The latent coordinates are initialized via multidimensional scaling on the average geodesic distances calculated over the different networks. Squared Euclidean distances are then computed on such starting latent positions, and the distances are used to perform a logistic regression of the adjacency matrices to get starting values for the intercept and the coefficient estimates $\alpha^{(k)}$ and $\beta^{(k)}$: Sender and receiver parameters are initialized in a non-informative way, by fixing them to 0.
Edge specific covariates, either constant or variable across the networks can be easily incorporated in the proposed framework, for the edge probabilities. Also, the model can be easily extended to deal with the presence of missing edges/nodes in the data. The full conditional and proposal distributions for model parameters are presented in Appendix B for the most general case of missing data and edge-specific covariates.

3.4 Undirected network

In the particular case of undirected networks, out-degrees and in-degrees are identical for each node, $\sum_i y_{ij}^{(k)} = \sum_i y_{ji}^{(k)}\ \forall i = 1, \dots, n$. Therefore, the framework proposed in section 3.2 can be easily modified to deal with undirected multidimensional networks. In such a context, we have to impose the constrain: $\theta_i^{(k)} = \gamma_i^{(k)} = \delta_i^{(k)}$, for all $i = 1, \dots, n$. According to such assumption, the edge probability equation (3.5)

can be rewritten as:

$$P\left(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, \delta_i^{(k)}, \delta_j^{(k)}, d_{ij}\right) = \frac{\exp\{\alpha^{(k)} \frac{\delta_i^{(k)} + \delta_j^{(k)}}{2} - \beta^{(k)} d_{ij}\}}{1 + \exp\{\alpha^{(k)} \frac{\delta_i^{(k)} + \delta_j^{(k)}}{2} - \beta^{(k)} d_{ij}\}} \tag{3.7}$$

The effect $\delta_i^{(k)}$, if not null, can either be variable across the different networks, $\delta_i^{(k)}$, or constant, $\delta_i^{(k)} = \delta_i$, $k = 1, \ldots, K$ and $i = 1, \ldots, n$. Appendix B provides the reader with the proposal distributions used to estimate these model parameters.

3.5 Simulations

We defined a simulation study to evaluate the performance of the proposed estimation procedure for the latent space models, where we examine the behaviour of the model parameter estimates of the model parameters for models NC, NV, CC, CV and VV. For each one of the simulated multiplexes, we fit the "true" model, that is, the model a given multiplex was simulated from. These five models are chosen to investigate the properties of the estimators when the true scenarios refer to different numbers of parameters. A first scenario of simulated multidimensional networks, **B1**, has dimensions $(n = 50, K = 5)$. A second larger one, **B2**, has dimensions similar to those of the multiplex that we analyse in the application, in section 3.6, namely $(n = 50, K = 10)$.
In both these simulation scenarios, we set $\alpha^{(1)} = 2$ and $\beta^{(1)} = 1$. The prior parameters are $\nu_\alpha = \nu_\beta = 3$, $m_\alpha = 2$, $m_\beta = 0$, $\tau_\alpha = \tau_\beta = (K-1)/K$ and $p = 2$. Small variations of these values have been also tried and did not affect the obtained results. Also, there are no missing edges in the simulated data. The MCMC algorithm run for 60000 iterations with a burn in of 15000.
Table 3.4 shows the average values of the distance correlation computed between the simulated and estimated edge probabilities, for all the different networks composing the multiplexes in scenarios **B1** and **B2**. As we may observe, the correlations are always high, proving that we are able to recover the edge probabilities with good quality, regardless of the true model, the view or the multiplex size. Also, last column of Table 3.4 reports the average values of the Procrustes transformation between the simulated and the estimated latent space coordinates. As for the edge probabilities, we see that the latent coordinates are appropriately recovered. To evaluate the estimates of the sender/receiver parameters in the different models, we compute, for each network, the Spearman correlation coefficient between the simulated and the estimated parameters. This coefficient is used as we are mainly interested in recovering the nodes ordering, with respect to the two estimated effects. Indeed, sender and receiver parameters vary in a relatively small interval, $(-1, 1)$, and the exact numerical values might not be of much interest. Table 3.3 reports, for each network in the simulated multiplexes, the average values of the Spearman correlation coefficient between the n simulated and estimated receiver parameters, $\gamma_i^{(k)}$. These values are always much greater than 0.5, with a couple of exceptions for some networks in the models with higher complexity, that is CV and VV. However, as we could see from Table 3.4, this does not impact the recovering of edge probabilities. Table 3.2 reports the average values of the Spearman correlation coefficient between

the $n = 50$ simulated and estimated sender parameters, $\theta_i^{(k)}$. The behaviour of these estimates complies with those of the receiver parameters previously discussed, proving that the two effects, when both presents, are estimated equally good.
We refer to Appendix B for the $\alpha^{(k)}$ and the $\beta^{(k)}$ estimates. The intercepts are recovered within a 95% credible interval, with two limited exceptions which occur when the simulated values are "extreme". However, the ordering between the different intercepts in a given multiplex is always recovered. Instead, the $\beta^{(k)}$ coefficient tends to be overestimated. Also in this case, the ordering of coefficients in a multiplex is correctly recovered. The overestimation of this coefficient may be likely caused by a corresponding underestimation of the latent distances, as the simulated and estimated products $\beta^{(k)}d_{ij}$ are always well recovered. Precise point estimates of all the parameters are quite hard to recover, due to the large number of parameters in the models. However, the aim of this class of latent space models is to describe different features of a multiplex by comparing nodes and networks. This intent is met, as we are always able to recover the corresponding orderings.

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
B1	CC	0.84	0.84	0.84	0.84	0.84	-	-	-	-	-
	CV	0.90	0.90	0.90	0.90	0.90	-	-	-	-	-
	VV	0.58	0.57	0.55	0.58	0.59	-	-	-	-	-
B2	CC	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
	CV	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
	VV	0.65	0.66	0.66	0.65	0.66	0.66	0.66	0.66	0.67	0.65

Table 3.2. Simulation study. Spearman correlation between the simulated and the estimated sender effects, by simulation scenario and true model structure.

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
B1	NC	0.84	0.84	0.84	0.84	0.84	-	-	-	-	-
	NV	0.85	0.80	0.86	0.92	0.91	-	-	-	-	-
	CC	0.84	0.84	0.84	0.84	0.84	-	-	-	-	-
	CV	0.80	0.86	0.81	0.73	0.6	-	-	-	-	-
	VV	0.67	0.61	0.50	0.64	0.75	-	-	-	-	-
B2	NC	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	NV	0.92	0.86	0.84	0.74	0.84	0.83	0.92	0.61	0.80	0.79
	CC	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
	CV	0.51	0.66	0.83	0.50	0.73	0.82	0.82	0.85	0.61	0.79
	VV	0.61	0.48	0.81	0.83	0.86	0.58	0.54	0.57	0.87	0.59

Table 3.3. Simulation study. Spearman correlation index between the simulated and the estimated receiver effects.

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	PC
B1	NC	0.86	0.84	0.86	0.87	0.85	-	-	-	-	-	0.94
	NV	0.87	0.87	0.87	0.85	0.86	-	-	-	-	-	0.92
	CC	0.89	0.89	0.90	0.91	0.87	-	-	-	-	-	0.96
	CV	0.84	0.81	0.83	0.84	0.85	-	-	-	-	-	0.90
	VV	0.83	0.81	0.80	0.81	0.81	-	-	-	-	-	0.93
B2	NC	0.89	0.84	0.83	0.83	0.81	0.92	0.84	0.81	0.81	0.82	0.85
	NV	0.83	0.80	0.79	0.76	0.81	0.78	0.88	0.71	0.75	0.75	0.93
	CC	0.83	0.87	0.91	0.84	0.90	0.91	0.91	0.84	0.86	0.91	0.90
	CV	0.75	0.86	0.87	0.76	0.86	0.88	0.87	0.86	0.83	0.87	0.91
	VV	0.73	0.72	0.82	0.85	0.88	0.79	0.79	0.79	0.86	0.84	0.90

Table 3.4. Simulation study. Distance correlation between the simulated and the estimated edge-probabilities. Last column (PC) shows the Procrustes correlation between the simulated and the estimated latent space coordinates.

3.5.1 An heuristic procedure for model selection

In Section 3.2 we have proposed a class of 9 models for multidimensional networks, based on 9 different alternatives for the sender/receiver effects in a model for multiplex data. In general, the issue of model selection can be addressed in two different ways. A first approach is that of an expert, that has some previous knowledge on the data and suggests a particular model should be used. A second, more common, approach, is that of choosing the "best" model by some selection criteria.
Note that, in the present context, the estimation of the nine models, on a specific observed multiplex may request some (computational) time, especially when the number of nodes is large.
Hence, it could be good to have some idea on which model to estimate on a priori basis. In particular, we propose to determine the model to fit on the basis of the correlations between the observed in-degrees and out-degrees of the networks in the observed multiplex. The general idea is that computing the correlations among networks out-degree/in-degree distributions could serve as a proxy of the heterogeneity within and between the views. Let us denote by $\mathbf{S} = [s_{ik}]$ the matrix of the observed out-degrees and by $\mathbf{R} = [r_{ik}]$ the matrix of the observed in-degrees; both matrices have dimension $n \times K$, where n is the number of nodes and K the number of networks. Then, the matrices $cor(S) = c_{s_k}$ and $cor(R) = c_{r_k}$, of dimension $K \times K$, contain the values of the correlation of the sender/receiver effects between views. Instead, the matrices $cor(S^T) = c_{s_i}$ and $cor(R^T) = c_{r_i}$, of dimension $n \times n$, include the correlations of the sender/receiver effects between nodes. Let us now define $\bar{c}_{s_k}$, $\bar{c}_{r_k}$, $\bar{c}_{s_i}$ and $\bar{c}_{r_i}$ the mean values among all the cells of the matrices introduced above and $sd(c_{s_k})$, $sd(c_{r_k})$, $sd(c_{s_i})$ and $sd(c_{r_i})$ the corresponding standard deviations. We propose to use such quantities to choose which type of model has to be estimated, given an observed multidimensional network, by the heuristic procedure described in Figure 3.2. Observed multiplexes with similar values of $\bar{c}_{s_k}$ and $\bar{c}_{r_k}$, and of $\bar{c}_{s_i}$ and $\bar{c}_{r_i}$, could have similar types of sender and receiver effects. On the contrary, a multiplex that exhibit conflicting values of sender/receiver correlation among views or among nodes might come from a model where the two effects are of different types.

In the latter case, the higher the discrepancy between sender and receiver "node" correlations, the higher the chance that the underlying model has two most different types of effect, that is, one null and one variable. Whether the most complex effect is the sender or the receiver one is decided by looking at $\bar{c}_{r_k}$ and $\bar{c}_{s_k}$, while the smallest of the two determines which effect should be the less complex. When the sender/receiver effects are selected to be of the same type, a low correlation among views and the presence of variability between the nodes correlations indicate that the underlying model is the most complex one, the VV model. In a similar way, the variability between the nodes correlations is used to discriminate between the NN and the CC models; a low variability, below a threshold ϵ_1, characterizes the NN model, as no sender/receiver effects modify the influence of the distances on the edge probabilities. As it can be seen from the scheme in Figure 3.2, this procedure depends on four thresholds: ϵ_1, ϵ_2, c_1 and c_2. We tested the performance of the proposed procedure via a large scale simulation study and the results are presented in Appendix B. In the simulation setting, we have run the procedure on two groups of 10000 multidimensional networks. The first multiplex has dimensions (50×10), while the second one has dimensions similar to those of the multiplex studied in the application in Section 3.6, (65×10). The thresholds have been fixed via cross validation and are $\epsilon_1 = 0.12$, $\epsilon_2 = 0.2$, $c_1 = 0.5$, $c_2 = 0.8$. The procedure proved to be fast and to give quite reliable results. Indeed, on average, the "true" model is chosen 90% of the times. Also, when the heuristic procedure fails to select the right model, it proposes a model that is "near" the "true" one. For example, when a multiplex of dimension (65×10) is simulated from a VV model, the procedure returns the VV model 90% of the times and the CC model the other 10%.

3.6 FAO trade data

The application deals with FAO food and agricultural trade data, measuring annual import/exports between countries. The data are available at the FAO website (FAOStat, 2013), and the most recent subset refers to 2013. Here we consider the fruit sub-market, in particular, fresh fruit, as fresh items are the most internationally traded. For illustrative reasons, we consider a restricted number of fruits, by choosing 10 out of the most commonly consumed and traded goods: "Grapes", "Watermelons", "Apples", "Oranges", "Pears", "Bananas", "Pineapples", "Tangerines, mandarins, clementines, satsuma", "Plantains" and "Grapefruit (inc. pomelos)". The original data register the volume of the trade, that is, the quantity traded among each couple of countries; however, we focus on the presence/absence of an import/export relation between couples of countries. Our aim is to verify whether close countries are more likely to trade, by comparing the estimated latent coordinates of the countries with the geographical ones. Also, we can address which countries are the most relevant in the exchange of fruits, via the estimated sender/receiver parameters, and whether their relevance is constant throughout the different markets. The original number of trading countries in the data is large, more than 200, but not all of them trade in all of the markets. Thus, to avoid the presence of isolated node-countries and to guarantee an easy and feasible representation of the results, we focus on a subgroup of 64 countries, reported in Table B.1, appendix B. To define such a sub-sample,

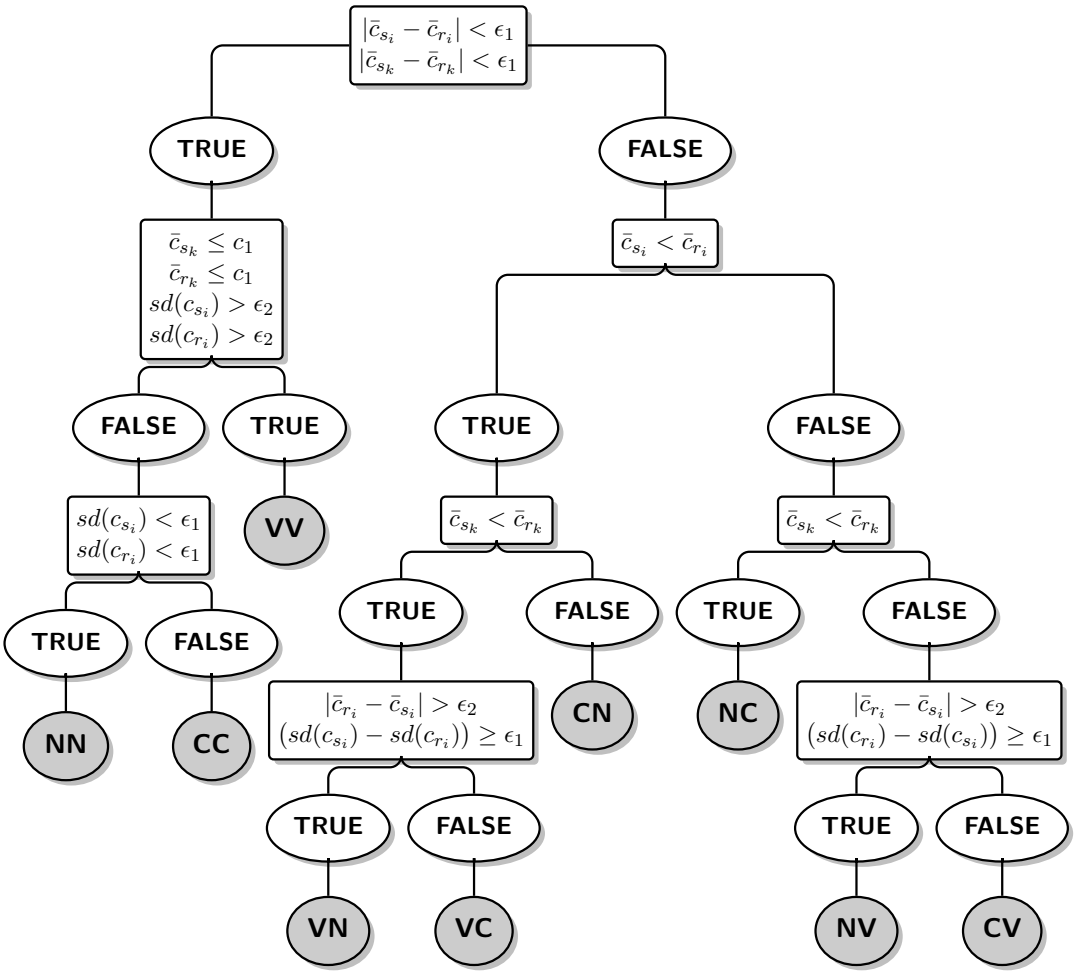


Figure 3.2. Synthetic representation of the heuristic procedure for model selection. The thresholds have been fixed via a cross validation exercise and are: $\epsilon_1 = 0.12$, $\epsilon_2 = 0.2$, $c_1 = 0.5$, $c_2 = 0.8$, respectively.

we have considered the median number of countries with which a country trades (equal to 7) and removed all the countries with a value under the median. We end up considering a multidimensional networks with the same number of $n = 64$ nodes and $K = 10$ networks.

The observed densities range from 0.10 (Plantains market) to 0.28 (Apples market), with a mean of 0.20. Also, the associations[1] between couples of adjacency matrices are quite high, ranging in between 0.8 and 0.9, suggesting that countries tend to import/export fruits from/to a relatively constant set of partners. The observed

[1]The association between any two adjacency matrices, $k, l = 1, \ldots, K$, is computed comparing the total number of concordant cells between the two matrices and the total number of cells:

$$As\big(\mathbf{Y}^{(k)}, \mathbf{Y}^{(k)}\big) = \frac{\sum_{i,j}^{n} \mathbf{I}\big(y_{ij}^{(k)} = y_{ij}^{(l)}\big)}{\sum_{i,j}^{n} \mathbf{I}\big(y_{ij}^{(k)} = y_{ij}^{(l)}\big) + \sum_{i,j}^{n} \mathbf{I}\big(y_{ij}^{(k)} \neq y_{ij}^{(l)}\big)}.$$

out-degrees and in-degrees present a strong association, as it can be seen by looking at Figure 3.3; thus, the data are a good candidate to test the proposed model.
The heuristic procedure described in Section 3.5.1 suggested to use the VC model. Indeed, the average Spearman correlation (Figure 3.3) computed among couples of observed out-degree distribution, 0.73, is lower than the one computed among couples of observed in-degree distributions, 0.80. However, we fit all the nine models proposed to the data and compute different model selection criteria, to further validate the model proposed by the heuristic procedure. The hyper parameters are fixed as in the simulation setting (Section 3.5), and $p = 2$ is the dimension of the latent space, both for plotting reasons and to compare the estimated with the geographical coordinates. AIC (Akaike, 1974), BIC (Schwarz, 1978) and PWAIC1 (Gelman et al., 2014) agreed with the heuristic procedure in the selection of the VC model. DIC instead opted for the VV model, the most complex one. Almost all of the selection criteria used and the heuristic procedure agree in proposing the VC model. Therefore, here we present only the results of this model for the FAO fruit trade multidimensional network.
Table 3.5 reports the top three exporting countries, by means of the estimated sender effects in the different networks. Top exporting are to be interpreted as those countries that tend to export fruit to a large group of trading partners, conditionally on the the latent distances to other nodes. Countries appearing as top exporting in any of the considered fruit markets never show up in the bottom three positions. Also, some countries tend to be top exporting in more than one market. For example, Netherlands (NLD) appears seven times in one of the top three positions. Indeed, Netherlands is amongst the world's biggest (re-)exporting countries for many fruits and vegetables (FreshPlaza, 2015b). Also, together with Belgium, it is one of the major trade hubs for fresh fruits, importing goods from developing countries and then reselling them (mostly) to the European market (CBI, 2015). Spain (ESP) is estimated to be among the top three exporting in six markets, and first exporter in three of them. Contrary to Netherlands, Spain directly grows most of the fruits it exports; just to give an example, Canary Islands are great pineapples producers. Also, Spain in 2017 became the world's largest watermelon exporting country (FreshPlaza, 2018). Italy is often estimated to be among the top three exporters, as it appears five times in the top three. Something similar happens with the estimated bottom three positions; some countries are estimated to be bottom exporters in more than one market. Maldives (MDV) and Kazakhstan (KAZ) appear in the bottom three positions 7 and 6 times, respectively. Indeed, both countries depend largely on imports for most fruit requirement (Aquastat, 2011; Food and Drink, 2017). As for the receiver effects, see Table 3.6, these are constant across the networks, and Germany (DEU) and Netherlands (NLD) are estimated to be first and second overall importing countries. Indeed, in 2014, the German market accounted for a third of the Dutch (re-)export of fresh fruit (CBI, 2015; FreshPlaza, 2015a). Also, Germany is estimated to be in the top three importers ranking for three markets, always jointly with Netherlands. The results are coherent with what we observed in the data. For example, in 2013, Germany is the country that imports from a higher number of trading partners, while Maldives never export. Also, Netherlands is the country that mostly export to an higher number of trading partners.
The strong similarity in the sender/receiver behaviour of Germany and Netherlands

is reflected in the estimates of their latent coordinates, see Figure 3.4; the coordinates of the two countries almost overlap and are quite close to Belgium, another big European (re-)exporter. The estimated latent coordinates do not resemble much the geographical ones; indeed, the Procrustes correlation between the two sets is not high (0.45). The estimated latent space, Figure 3.4, is characterized by a large number of Asiatic countries placed in the top-left of the space and the majority of the European countries in the bottom-right part. United States are placed in between Asiatic and Oceanic countries. Most of these countries are leading suppliers of fruits for the United States, thanks to established or pending free trade agreements (Johnson, 2016). The latent space is estimated to be always relevant in the determination of the edge probabilities, and to have similar effect on the different networks. Indeed, Table 3.7 shows that the estimated coefficients $\beta^{(k)}$ range in $(1, 1.60)$; the same Table displays the estimates of the intercepts in the different networks. All of the intercepts correspond to high edge probability values, from 0.971 in the first network, the lowest one, to 0.996 in the eight network, the highest one. Figure 3.5 represents the estimated probabilities in the fruit networks, given some values of the distances and of the $\phi_{ij}^{(k)}$ coefficients. In particular, the probabilities are computed for the first (0.53), second (1.36) and third (2.71) quartiles of the distances, and for $\phi_{ij}^{(k)} = (1, 0.5, 0, -0.5)$. The plot in Figure 3.5 shows that, even though the latent space is constant, quite different values of the edge probability correspond to the same estimated distances, depending on the joint sender/receiver effects and on the network-specific parameters, $\alpha^{(k)}$ and $\beta^{(k)}$.

Last, for illustrative purposes, we display the estimated sender/receiver effects of the countries for the Apples and Pineapples networks (Figure 3.6); as the receiver effects are constant (y-axis), the plot helps to visualize the different role of each country in the two markets. Two red lines divide the figure in four quadrants. The first and third quadrants contain countries whose estimated sender and receiver effects have same sign.The second (fourth) quadrant displays those countries that have negative (positive) sender effects and positive (negative) receiver effects. Nineteen countries change quadrant in the two markets (Apples and Pineapples). For example, Greece (GRC) and New Zeland (NZL) have both quite an high sender effect in the apple market, but this effect turns negative when trading Pineapples. In general, for most countries, the sender effect estimates tend to be lower in the Pineapples market. Malaysia (MYS) and India (IND) are the only two countries displaying the opposite behaviour, as their estimated sender effects are higher in the Pineapples market. Indeed, in 2018, the two countries where both ranked among the world top 20 Pineapples producers (Worldatlas, 2018). Nonetheless, none of the two is estimated to be among the top three Pineapples exporting countries in Table 3.5. However, recall that the sender/receiver estimates are to be interpreted conditionally on the latent distances and that India and Malaysia are located near the majority of the Asiatic countries in the latent space, the top-left panel of Figure 3.4. Looking at the original data, both countries, especially Malaysia, tend to export mainly to other Asiatic countries, to which they are close in the latent space. That is, an high probability of trading with Asiatic countries is already guaranteed by the closeness in the latent space. Then, the "residual" contribution given by the sender effect to the edge probability does not need to be really high. On the opposite, Thailand tends

to trade with a larger and more variable group of other countries and, therefore, the estimated sender effect in the Pineapples market is higher than those of Malaysia and India.

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
Top	1°	ITA	ESP	ITA	ZAF	ESP	NLD	THA	ESP	DEU	ZAF
	2°	ESP	NLD	FRA	ESP	ITA	THA	NLD	NLD	NLD	NLD
	3°	ZAF	ITA	NZL	EGY	NLD	DEU	DEU	ITA	IND	ESP
Bottom	62°	KWT	QAT	KAZ	RUS	NOR	MDV	LBN	QAT	MAR	QAT
	63°	RUS	MDV	MAR	KAZ	KAZ	KAZ	MDA	NOR	NOR	MDV
	64°	KAZ	HKG	MDV	MDV	MDV	MKD	KAZ	MDV	UKR	MKD

Table 3.5. Estimated sender effects: top and bottom three exporting countries.

1°	2°	3°	4°	5°	6°	...	59°	60°	61°	62°	63°	64°
DEU	NLD	BHR	SGP	DNK	POL	...	AUS	EGY	MAR	ZAF	PAK	IRN

Table 3.6. Estimated receiver effects: top and bottom six importers countries.

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
α	mean	3.50	5.10	5.46	5.45	5.29	4.91	5.13	5.50	5.07	5.37
	sd	-	0.21	0.22	0.22	0.22	0.20	0.21	0.25	0.23	0.23
β	mean	1.00	1.34	1.03	1.18	1.13	1.51	1.17	1.36	1.60	1.31
	sd	-	0.06	0.04	0.06	0.05	0.06	0.05	0.06	0.08	0.06

Table 3.7. Averages and standard deviations of the estimated posterior distributions for the intercept and the scale coefficient parameters in the fruit networks.

Figure 3.3. Fruit multiplex: Pairplots of the observed out- and in-degrees of the networks. The upper diagonal matrix represents the associations between the observed out-degrees in any couple of networks, while the lower diagonal refers to the association between the in-degrees. The values of the Spearman correlation indexes are reported.

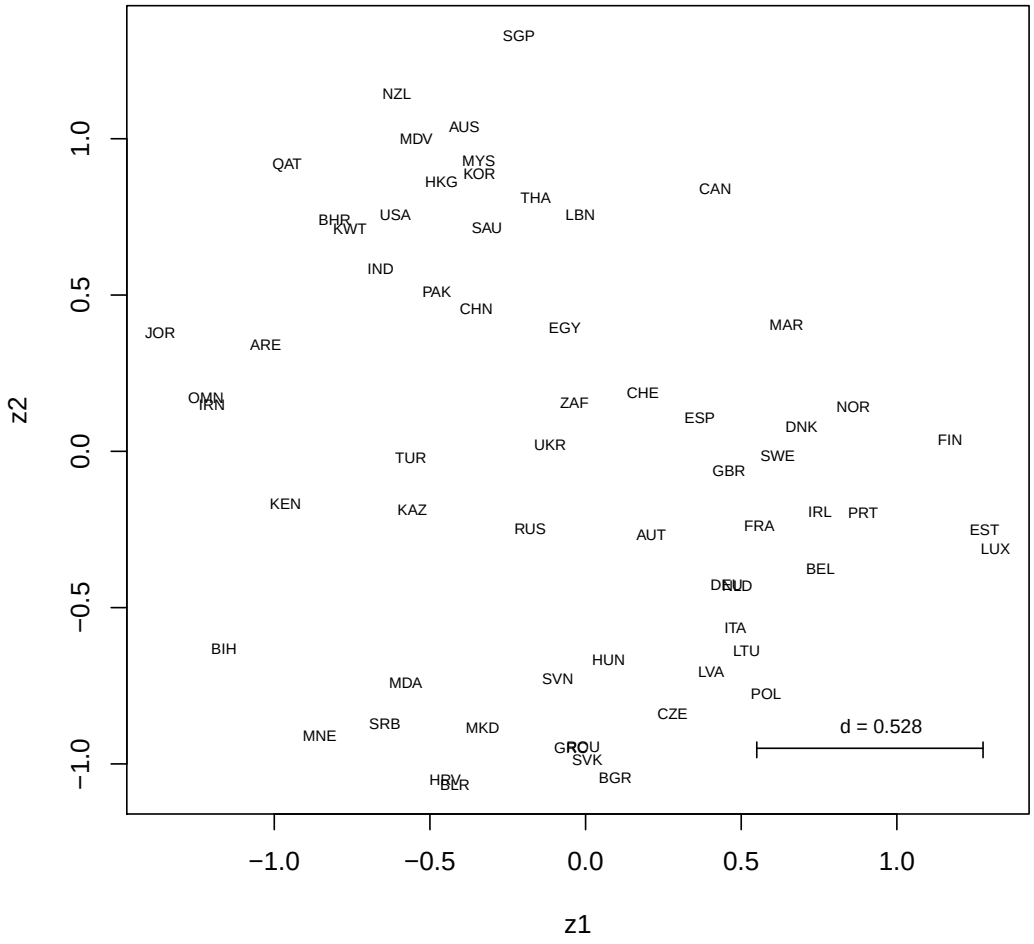


Figure 3.4. Fruit multiplex. Estimated latent coordinates for the countries. The segments represented at the bottom of the plot displays the value for the first quartile of the estimated distribution of the distances.

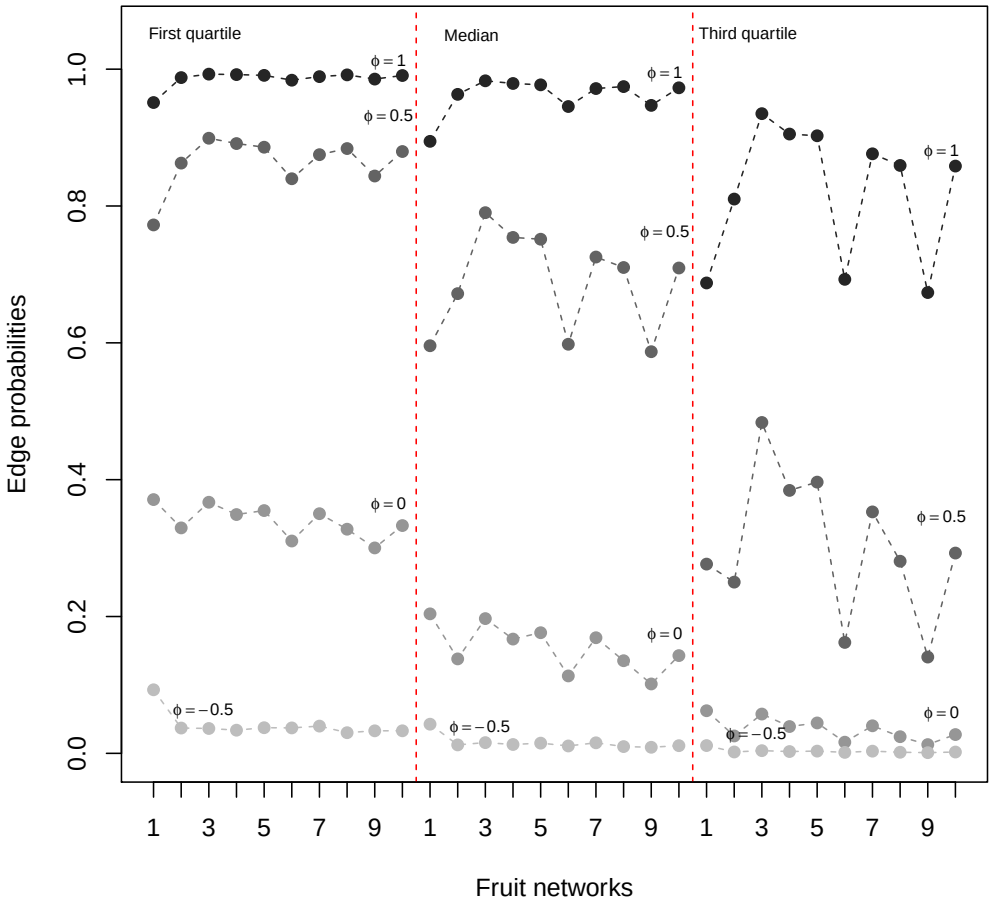


Figure 3.5. Fruit networks. Estimated probabilities in the multiplex, for different values of the estimated distances (first quartile, median and third quartile) and of the $\phi_{ij}^{(k)}$ coefficients.

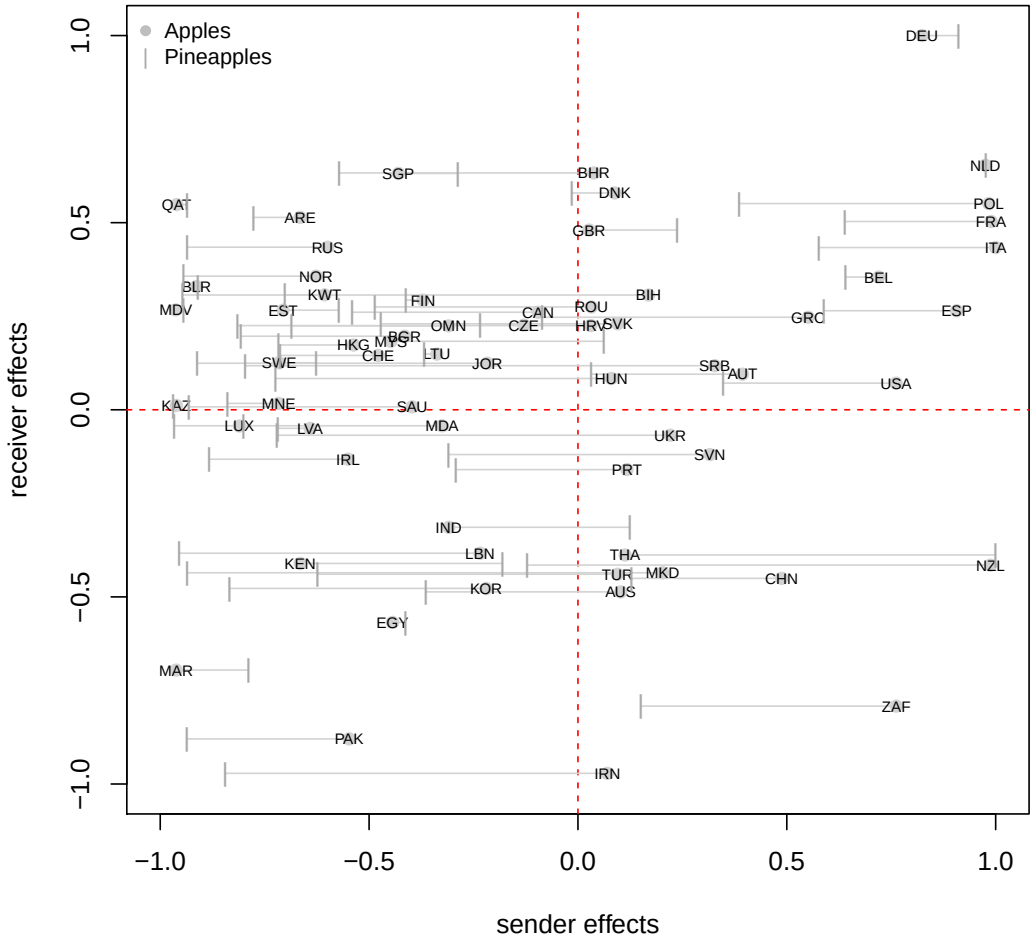


Figure 3.6. Apples (black) and Pineapples (grey) networks. Estimated sender and receiver effects of the countries.

3.7 Discussion

In the present work we have introduced a novel class of Euclidean distance latent space models for multidimensional network data. The models allow to represent transitivity in a parsimonious way, via a single latent space. Also, different levels of node-specific degree heterogeneity can be specified. In the spirit of model parsimony, we assume that the type of sender/receiver effect ("Null", "Constant" and "Variable") is constant across the views. An interesting relaxation of such hypothesis would be to have the type of effect varying with the networks. Indeed, for example, it may be that a subset K^* of the K views has no sender effect, but the remaining networks have constant sender effect. Finding such sub-groups of networks would then become a clustering problem, with extra complexity brought by the allocation of each network to the specific effect-sub-group and the estimation of the number of clusters, which, however, would be bounded in $(1, 3)$.
Also, we have proposed an heuristic procedure for model selection, that allows to choose an appropriate model for observed multiplex data without the need to estimate all the possible models. Thus, the procedure permits to bypass a classical model selection step. A preventive selection of the model may be convenient in many real data applications, as model estimation for network data can be quite (computationally) demanding. The performance of the proposed heuristic procedure and that of the latent space model have been tested in separate simulation studies and have proven to give quite good results.
An illustrative application to FAO trade data regarding different fruit trades has been presented, where our method was able to uncover trade patterns and shared similarities among different fruit markets. The data may be an interesting research problem per se, and an interesting extension of the proposed class of models could take into consideration weighted multiplexes, analysing import/export values or quantities. However, considering such weighted edges is non trivial, as the distributions of the exchanged quantities are both right skewed and zero-inflated.
The proposed models will be incorporated shortly in the R package *spaceNet*, already available on *CRAN*.

Chapter 4

Clustering Multidimensional Networks via Infinite Mixtures

4.1 Introduction

Network data describe relations and interconnections among n units. Interacting units are denoted as *nodes*, while connections are called *edges*. Well known examples are social network data, where friendship, approval, admiration, and other social relations are expressed between individuals and may be modelled to understand how people interact in certain contexts. Other than in social applications, networks may be observed in many fields, as for example biology or economics. In general, connections observed in network data are of difficult visualization and interpretability, due to the complex nature of networks themselves. Therefore, network analysis methods mainly focus on reconstructing and explaining the connections observed among the nodes. Such methods have to deal with the double task of faithfully modelling the relations between the units, which lead to the observed connections, and of providing with a feasible and interpretable summary of the data. Observed connections are modelled stochastically and each pair of nodes, also known as a dyad, is associated to an edge probability. Edge probabilities describe dyads connection propensity and can be modelled in many different ways, according to the data at hand or the purpose of the analysis. First attempts to model network data were built on the assumption of independence between the edges. Edge probabilities were assumed to be constant, see (Erdős and Rényi, 1959), or simple functions of some network statistics as in (Duijn, Snijders, and Zijlstra, 2004; Holland and Leinhardt, 1981). These first models paved the way for a more in depth analysis of edge formation in network data, where the edges independence assumption was either reduced or removed and complexity was added to the specification of edge probabilities (Hoff, Raftery, and Handcock, 2002; Holland, Laskey, and Leinhardt, 1983; Nowicki and Snijders, 2001; Robins et al., 2006; Snijders and Nowicki, 1997; Wasserman and Pattison, 1996). Among these models, the class of latent variable models has gained a growing attention in the last years. Such models are particularly interesting and meaningful as they explain the observed interconnection structure in network data by means of latent variables, which capture the association between the nodes. First latent variable models are the latent space model by Hoff, Raftery, and Handcock, 2002 and the stochastic

block model by Holland, Laskey, and Leinhardt, 1983; Snijders and Nowicki, 1997. The latter may be thought as a latent class analysis model for network data and it is explicitly designed to model clustering of the nodes. Clustering is a feature often observed in many real world network data, as nodes may tend to connect more frequently within given sub-groups. The stochastic block model classifies the nodes into different sub-groups and provides a framework for modelling the between cluster interaction structure. Some recent developments on the topic are those by Signorelli and Wit, 2018, which analyses bill cosponsorships in the Italian Chamber of Deputies, Bouveyron, Latouche, and Zreik, 2018, which extends the stochastic block model to the analysis of textual data, and those by Bartolucci, Marino, and Pandolfi, 2018; Matias, Rebafka, and Villers, 2018, extending the model to dynamic (longitudinal) network data. Although stochastic block models may synthetically represent clusters, they fail to represent within cluster transitivity, which refers to how nodes locally interact. A flexible extension of the stochastic block model which addresses such issue is the mixed membership stochastic block model by Airoldi et al., 2008. In this framework, nodes may belong to different clusters, depending on whom they are interacting with. A different approach is that of model-based clustering for latent space models, introduced by Handcock, Raftery, and Tantrum, 2007. This framework is based on the class of latent space models, and directly account for transitivity in network data (Handcock, Raftery, and Tantrum, 2007; Hoff, Raftery, and Handcock, 2002). Edge probabilities are described as a function of node positions in an unobserved space, which is responsible for the observed structure in the network. Such latent positions arise from a mixture distribution, whose components correspond to clusters in the data, as in standard model-based clustering framework (Fraley and Raftery, 2002). The original model is estimated with a MCMC algorithm; Salter-Townshend and Murphy, 2013 re-implemented it with a variational Bayesian inference approach. An extension to the framework of latent position cluster model is that by Gormley and Murphy, 2010, which combines it with a mixture of experts framework. Fosdick et al., 2018 attempt to bridge stochastic block models and latent position cluster models, using the so-called *Latent Space Stochastic Blockmodel*. In this framework, within cluster probabilities are modelled via a latent space model, while between cluster interactions are expressed as in stochastic block models.

The number of clusters is often unknown and needs to be inferred from the data. So far, almost all latent variable-based clustering frameworks proposed for network data assumed a fixed number of clusters, and then compared models with different number of clusters via cross validation or model selection criteria. In this context, a different approach is that by Ryan, Wyse, and Friel, 2017, which allows to estimate the number of clusters in a network, by analytically integrating out cluster-specific parameters.

In model based clustering, the problem of selecting the number of clusters often coincides with that of selecting the number of mixture components. Several approaches have been proposed to tackle this problem. In a first approach, choosing the number of mixture components corresponds to selecting the most suitable model for some observed data. This choice is often based on model selection criteria, used to compare different models specified according to different number of components. Various criteria, derived under different modelling assumptions, have been proposed

in the literature; for a review see McLachlan and Rathnayake, 2014. In principle, this approach would require to estimate models with all possible number of components and then compare them. However, this is often computationally unfeasible, especially with increasing sample size[1]. In a Bayesian framework, reversible jump MCMC algorithms (Green, 1995) allow to make inference on the unknown number of components, as simulations from the component parameter posterior distributions are drawn from a space of varying dimensions (Zhang et al., 2004). However, trans-dimensionality makes reversible jump algorithms quite computationally intensive, for a solution see e.g. Petris and Tardella, 2003. Also, these algorithms strongly depend on the modelling framework; different models would require the implementation of different reversible jump MCMC algorithms. Another approach is that of overfitting mixture distributions by Malsiner-Walli, Frühwirth-Schnatter, and Grün, 2016. The authors suggest to estimate the number of mixture components specifying sparse hierarchical priors on the mixture weights and component parameters. This results in an overfitting mixture model where superfluous components are emptied during MCMC sampling. This approach presents a general and straightforward way to estimate the number of components, which is defined as the most frequent number of non-empty components visited during MCMC sampling. However, a maximum number of components, which will later be partially emptied, needs to be specified. Related to the overfitting mixture distribution approach is that of Dirichlet process mixture models (Frühwirth-Schnatter and Malsiner-Walli, 2018; Rasmussen, 2000). Here, the number of components is assumed to be potentially infinite, and the component parameters are taken to be the realizations of a Dirichlet process (Antoniak, 1974). Since in the observed sample the number of component is likely finite, joint inference on the number of components and the mixture distribution parameters is made possible by the Dirichlet process properties. Differently from overfitting mixture distributions, the maximum number of components does not need to be fixed and the number of inferred components may change with samples of different sizes.

In many situations, the same set of nodes can be characterized by multiple relations, or the same relation can be recorded over multiple time points. Such setting produces a more complex type of network data, referred to as *multidimensional networks*, or *multiplexes*. A multiplex is a collection of K $n \times n$ networks, or *views*, where multiple sets of edges are observed for the same group of nodes. Few works extended the latent position cluster model to such complex multidimensional network data. In the particular case of dynamic networks, where the multiplex refers to the same relation recorded at different time occasions, Sewell and Chen, 2017 extended the work by Handcock, Raftery, and Tantrum, 2007 to perform clustering in dynamic network data; also in this context, the number of cluster should be fixed a priori. As already noted, stochastic block models have already been extended to dynamic network data, see Matias, Rebafka, and Villers, 2018 or Yang et al., 2011. However, to the authors knowledge, no specific clustering approach for multidimensional network data has been proposed in the literature.

[1] In a network with n nodes, the sample size is n^2, which corresponds to the number of potentially observable edges. Hence, network data computational complexity grows quadratically with the number of nodes.

In the present work, we develop an *infinite latent position cluster model* for single and multidimensional network data, built within the class of latent space models. The proposed framework allows to jointly estimate cluster parameters and latent coordinates without previous specification of the number of clusters. Indeed, this is taken as a model parameter and inference is performed on it as well. Here, we model binary (multidimensional) networks, although the clustering method we propose can be extended to other types of multidimensional network data as well.
Let $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(K)}\}$ be a multiplex with K networks (views). Each element $\mathbf{Y}^{(k)}$ of $\mathbf{Y}$, $k = 1, \dots, K$, is an adjacency matrix of dimension $n \times n$, where n is the number of nodes in the multiplex. Recall that, in multidimensional network data, the number of nodes is constant across the views, while the observed edges may change with the view. Thus, the general element of $\mathbf{Y}^{(k)}$ is

$$y_{ij}^{(k)} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are connected in the } k^{\text{th}} \text{ view;} \\ 0 & \text{else.} \end{cases}$$

When $K = 1$, the multidimensional network reduces to a single network with n nodes.
As it is generally done in latent space models, we assume that each node has an unknown position in a latent $p-$dimensional Euclidean space. The probability of observing a connection between nodes i and j is a function of their latent coordinates. The presence/absence of a connection between node i and j is assumed to be independent of all other connections in the network, conditional on the latent coordinates of the two nodes. We build our model within the class of distance latent space models, which assumes that edge probabilities are a function of the pairwise distances between the nodes in the latent space. Our choice is driven by the intent of modelling similarities between nodes in network data, for which distances represent a good proxy (Hoff, Raftery, and Handcock, 2002). We further assume that the latent coordinates arise from an infinite mixture of Gaussian distributions, represented in terms of the Dirichlet process mixture model (Antoniak, 1974; Ferguson, 1973). The infinite mixture framework allows to treat the number of mixture components G, and consequently the number of clusters, as a model parameter, on which inference is performed. Also, component-specific parameters are estimated to characterize the clusters. The proposed model is estimated within a hierarchical Bayesian framework and inference is carried out using a MCMC algorithm. Section 4.2 details the infinite latent position cluster model. Section 4.3 illustrates the estimation procedure, while some practical issues are discussed in Section 4.4. A simulation study to illustrate the performances of the proposed model is presented in Section 4.5. In Section 4.6 we present an application of the model to the analysis of self-reported social interactions among 7^{th} grade students. We conclude with a final discussion in Section 4.7.

4.2 The model

Before presenting the infinite latent position cluster model, we briefly review the latent position cluster model by Handcock, Raftery, and Tantrum, 2007, of which the proposed model provides an extension.

4.2.1 Latent position cluster model

Model based clustering for social networks was first introduced by Handcock, Raftery, and Tantrum, 2007, in the context of single network data. This model was developed to account for clustering of the nodes which could not be explained by transitivity or homophily by attributes alone (Hoff, Raftery, and Handcock, 2002). The authors assume that nodes have latent coordinates in a p-dimensional Euclidean latent space: z_i, $i = 1, \ldots, n$. The latent coordinates are taken to be drawn from a mixture of G spherical Gaussian components,

$$\boldsymbol{z}_i \sim \sum_{g=1}^{G} \pi_g MVN_p(\boldsymbol{\mu}_g, \sigma_g^2 \mathbf{I}),$$

for $g = 1, \ldots, G$, and where π_g are the mixture weights, $\boldsymbol{\mu}_g$ the components means and σ_g^2 the components variances. The choice of spherical covariance matrices, $\sigma_g^2 \mathbf{I}$, was driven by the fact that in latent space models the likelihood is invariant to rotations of the latent space. Thus, the authors specified an independence model for the coordinate system. The choice of the number of components G is a difficult task in this context, as clustering is performed on unobserved quantities, the latent coordinates. The authors proposed to address the issue using the Bayesian Information Criterion (BIC). However, such an approach may result unfeasible when applying the method to many real world network data. Indeed, in principle, one should estimate G different models, one for each possible value of the number of components $G \leq n$. This could represent a severe issue, as the computational burden associated with the implementation of each model grows quadratically with the number of nodes n. In the next section, we will introduce a latent space approach for network data based on infinite mixture models (Rasmussen, 2000) to overcome such an issue.

4.2.2 Infinite latent position cluster model

As in Handcock, Raftery, and Tantrum, 2007, we assume that nodes of the network lay in a p-dimensional Euclidean latent space. To introduce a clustering structure, we assume that a node-specific latent coordinates are distributed according to a mixture of p-variate Gaussian distributions. Adopting an infinite mixture representation for the latent coordinates,

$$\boldsymbol{z}_i \sim \sum_{g=1}^{\infty} \pi_g MVN_p(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{4.1}$$

we can jointly estimate the unknown number of components G and the component-specific parameters, $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$, where $g = 1, \ldots, G$ denotes the number of components. Such an infinite mixture distribution framework is particularly suited for (multidimensional) network data, as the number of mixture components is bounded by the number of nodes (Antoniak, 1974). Hence, it is implicitly assumed that the number of mixture components, can change if new nodes "enter" a network. Such an assumption is reasonable with network data, as new actors may alter clustering in the data, depending on how they interact with pre-existing nodes.

The number of components is only potentially infinite. Indeed, given an observed $n \times n$ network $\mathbf{Y}$, the maximum possible number of components is $G_{\text{MAX}} = n$. Hence, the number of components is finite in real world applications. Each component is characterized by two parameters, the mean $\boldsymbol{\mu}_g$ and the covariance matrix $\boldsymbol{\Sigma}_g$, and a weight π_g[2], indicating the fraction of nodes belonging to the g^{th} component.
We adopt a distance latent space modelling approach, and assume that the probabilities of an edge associated to the dyad (i, j) are a function of the distance between couple of nodes i and j. These distances are denoted by d_{ij}, $i, j = 1, \ldots, n$ and they are defined as squared Euclidean distances, $d_{ij} = ||\mathbf{z}_i - \mathbf{z}_j||^2$, see D'Angelo, Murphy, and Alfò, 2018; Gollini and Murphy, 2016. The distance matrix, of dimension $n \times n$, is denoted by $\mathbf{D}$. However, although we develop our proposal and the estimation procedure considering the squared Euclidean distance, the proposed framework may be extended to incorporate different specifications for the distance function as well.
In the context of single network data, edge probabilities may be defined as

$$P\Big(y_{ij} = 1 \mid \alpha, \beta, d_{ij}\Big) = \frac{\exp\{\alpha - \beta d_{ij}\}}{1 + \exp\{\alpha - \beta d_{ij}\}}, \tag{4.2}$$

where α is an intercept capturing the overall connectivity level in the network and β is a scale coefficient which weights the influence of the latent space on the edge probabilities[3]. When $\beta \approx 0$, the latent space is practically irrelevant and the edge probabilities reduce to those generated by a random graph (D'Angelo, Murphy, and Alfò, 2018; Erdős and Rényi, 1959).
Following D'Angelo, Murphy, and Alfò, 2018, we extend the model defined in equation 4.2 to multidimensional network data by considering view-specific parameters:

$$p_{ij}^{(k)} = P\Big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, d_{ij}\Big) = \frac{\exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}}{1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}}. \tag{4.3}$$

We assume that a unique latent space representation captures the overall similarities between the nodes in the multiplex and that network-specific parameters $\alpha^{(k)}$ and $\beta^{(k)}$ can modify the effect of the latent space on network-specific edge probabilities[3]. These two parameters have the same interpretation as α and β in the model for a single network, see equation 4.2. In fact, when $K = 1$, the model in equation 4.3 reduces to that in equation 4.2. In the next section, we provide details on the estimation procedure for infinite latent position cluster model parameters. We develop the procedure in the more general case of multiplex data (equation 4.3), but the procedure is easily reduced to the specific case of unidimensional network data (equation 4.2).

[2]Recall that

$$\sum_{g=1}^{G} \pi_g = 1$$

[3]As in D'Angelo, Murphy, and Alfò, 2018, distance scale coefficients are bound by $\beta^{(k)} \geq 0$, $k = 1, \ldots, K$. Also, network intercepts are bound by:

$$\alpha^{(k)} \geq \log\Big(\frac{\log(n)}{n - \log(n)}\Big) = LB(\alpha).$$

4.3 Model parameters estimation

Based on the edge probability model defined in equation 4.3, we write the log-likelihood for model parameters as

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{D} \mid \mathbf{Y}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} y_{ij}^{(k)} (\alpha^{(k)} - \beta^{(k)} d_{ij}) - \log(\alpha^{(k)} - \beta^{(k)} d_{ij}), \tag{4.4}$$

where $\boldsymbol{\alpha} = (\alpha^{(1)}, \ldots, \alpha^{(K)})$ and $\boldsymbol{\beta} = (\beta^{(1)}, \ldots, \beta^{(K)})$.
We propose a hierarchical Bayesian approach to estimate model parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and latent distances $\mathbf{D}$, using a MCMC sampling scheme. Each parameter is assumed to have a prior distribution, whose parameters are unknown. We refer to these as nuisance parameters. Nuisance parameter values are of no interest but their specification could be relevant. Instead of fixing a priori such values, we give nuisance parameters hyper-prior distributions and estimate them as well. Hence, the parameters of interest depend on nuisance parameters posterior distributions, controlled by hyperparameters, and not on fixed nuisance parameter values. This extra variability gives more flexibility in the estimation procedure. The final parameter estimates should be less sensible to the specification of hyperparameters than to that of nuisance parameters. More in details, the logit parameters are modelled via the following prior/hyper-prior specifications:

$$\alpha^{(k)} \sim N_{[LB(\alpha),\infty]}(\mu_\alpha, \sigma_\alpha^2), \quad \text{where:} \quad \mu_\alpha \mid \sigma_\alpha^2 \sim N_{[LB(\alpha),\infty]}(m_\alpha, \tau_\alpha \sigma_\alpha^2), \quad \sigma_\alpha^2 \sim \text{Inv}\chi^2_{\nu_\alpha};$$

and

$$\beta^{(k)} \sim N_{[0,\infty]}(\mu_\beta, \sigma_\beta^2), \quad \text{where:} \quad \mu_\beta \mid \sigma_\beta^2 \sim N_{[LB(\beta),\infty]}(m_\beta, \tau_\beta \sigma_\beta^2), \quad \sigma_\beta^2 \sim \text{Inv}\chi^2_{\nu_\beta};$$

for $k = 1, \ldots, K$. The set of logit nuisance parameters is $\phi_1 = (\mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2)$ and that of hyperparameters is $\phi_2 = (m_\alpha, m_\beta, \tau_\alpha, \tau_\beta, \nu_\alpha, \nu_\beta)$. Since the estimation procedure for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and the corresponding nuisance parameters ϕ_1 do not directly depend on the latent coordinates, we proceed with estimation using the same proposal/full conditional distributions derived in D'Angelo, Murphy, and Alfò, 2018.
The estimation of the latent distances in equation 4.4 is not that straightforward. Each distance $d_{ij} = ||\boldsymbol{z}_i - \boldsymbol{z}_j||^2$ depends on the latent coordinates for node i and j, distributed a priori according to equation 4.1. Hence, distances depend on the clustering of the unknown node coordinates, with no information available regarding the component parameters or the number of mixture components, which is (potentially) infinite.
A tractable way to deal with infinite mixture models is to assume that they are the realization of a Dirichlet process (Antoniak, 1974; Ferguson, 1973; Hjort et al., 2010; Müller et al., 2015), a stochastic process whose realization is indeed a probability distribution. This process is characterized by a continuous base distribution S_0, the expected value of the process, and a concentration parameter $\psi \geq 0$. Rewriting equation 4.1 as a Dirichlet process leads to:

$$\boldsymbol{z}_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \sim MVN_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); \qquad \boldsymbol{\Omega}_i \sim S; \qquad S \sim DP(\psi, S_0); \tag{4.5}$$

for $i = 1, \dots, n$. $\boldsymbol{\Omega}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the set of parameters for the prior distribution of $\boldsymbol{z}_i$, distributed according to S. Even though S_0 is continuous, distributions drawn around it are almost surely discrete. The degree of discretization depends on the concentration parameter ψ; the lower the value of ψ, the lower the number of unique realizations. The discrete nature of the Dirichlet process is what makes it suitable to describe mixture models. Indeed, in a finite sample of size n, the number of unique realizations is finite and it can be denoted by G, $G \leq n$, indexed by $g = 1, \dots, G$. This implies that some of the units come from a common component, that is some of the latent coordinates z_i share the same prior distribution parameters Ω_g, $g = 1, \dots, G$. The set of parameters $\boldsymbol{\Omega}_g = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, $g = 1, \dots, G$, corresponds to the set of component parameters, and the number of finite realizations G is indeed the number of mixture components. Differently from Handcock, Raftery, and Tantrum, 2007, see Section 4.2.1, we opt for a more flexible specification of the component covariance matrices. Indeed, we let the component variances vary across the latent dimensions,

$$\boldsymbol{\Sigma}_g = \begin{bmatrix} \sigma^2_{1g} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2_{pg} \end{bmatrix} \boldsymbol{I},$$

in order to capture heterogeneity of the nodes across the latent dimensions and the components. We decided to address the issue of the latent space rotation invariance in a different way, as it will be specified further on. The diagonal form of the component covariances allows us to specify S as a Normal-Inverse Gamma distribution, where

$$\boldsymbol{\Omega}_{rg} = (\mu_{rg}, \sigma^2_{rg}) \sim NIG(m_r, \tau_z \sigma^2_g, \nu_1, \nu_2), \quad r = 1, \dots, p.$$

We denote the mixture components hyperparameters with $\omega_g = \omega = (\boldsymbol{m}, \tau_z, \nu_1, \nu_2)$, for all $g = 1, \dots, G$. As for the logit parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, an extra layer of dependence is introduced for a flexible modelling of mixture components and latent coordinates. The hyperparameter $\boldsymbol{m} = (m_1, \dots, m_r, \dots, m_p)$ may be assumed to have standard multivariate Gaussian distribution, to allow for a more flexible estimation.
Given a sample of dimension n and a number of groups $G \leq n$, the sub-groups of latent coordinates arising from the same component are unknown. These may be modelled introducing a multinomial cluster label auxiliary variable $\boldsymbol{c} = (\boldsymbol{c}_1, \dots, \boldsymbol{c}_i, \dots, \boldsymbol{c}_n)$. The i^{th} entry is a G-dimensional binary vector $\boldsymbol{c}_i = (c_{i1}, \dots, c_{iG})$, whose elements are all 0 except from the g^{th} one, $c_{ig} = 1$, meaning that the i^{th} latent coordinate comes from the g^{th} component. Using the auxiliary variables $\boldsymbol{c}_i$, we may rewrite equation 4.1 as

$$(\boldsymbol{z}_i, \boldsymbol{c}_i) \mid_{\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g} \sim \prod_{g=1}^{G} \Big[\pi_g MVN_p(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \Big]^{c_{ig}}, \tag{4.6}$$

for $i = 1, \dots, n$. Then, given the mixture representation in equation 4.6, we may rewrite the mixture weights as

$$\pi_g = \sum_{i=1}^{n} \frac{c_{ig}}{n}, \qquad g = 1, \dots, G. \tag{4.7}$$

The unknown quantities of interest are now the actual number of components G, the component-specific parameters $\boldsymbol{\Omega}_g$, $g = 1, \dots, G$, and the cluster labels $\boldsymbol{c}_i$, $i = 1, \dots, n$. To estimate such quantities, we exploit the *Chinese restaurant* representation of the Dirichlet process (Aldous, 1985). Given a set of n units, this representation may be summarized as follow. Let us suppose we have a group of n individual entering in a restaurant, one at a time. The first individual enters and seats at the first table. Then, a second client arrives and may choose to sit with the first individual or by himself at a different table. Then the third individual arrives and faces the same choice: either seating at tables already occupied by other people or choose a new, empty one. The mechanism is repeated for each new person entering the restaurant, until the n^{th} one. It is straightforward to understand that n individuals may seat in a number of tables which is $G \leq n$. The restaurant tables are an analogy of the "formation" of mixture components in Dirichlet process Mixture models. In our context, clients are indeed the latent coordinates $\boldsymbol{z}_i$ and the table choice is coded by the auxiliary variable $\boldsymbol{c}_i$. The "propensity" with which clients choose to seat at those tables that are already occupied by other people is regulated by the concentration parameter ψ. The concentration parameter may be chosen a priori, if some information is available on the number of components G. However, as here the intent is to perform clustering on unobserved quantities, the latent coordinates, a subjective specification of the ψ parameter may be too informative. To incorporate the uncertainty on the concentration parameter in the modelling framework, we assume that ψ has a Gamma prior distribution, $\psi \sim \Gamma(\xi_1, \xi_2)$, see Müller et al., 2015. To summarize what has been introduced so far, we report in Figure 4.1 a schematic representation of the hierarchical structure in the proposed model.

Inference procedures for the Chinese restaurant representation of the Dirichlet process when the base distribution is a Normal-Inverse Gamma have been widely studied in the literature, see Hjort et al., 2010; Müller et al., 2015. In particular, here we adopt the proposal by Bush and MacEachern, 1996 to update the cluster labels $\boldsymbol{c}_i$ and that by Escobar and West, 1995 to update the concentration parameter ψ.

Given an observed multidimensional network $\mathbf{Y}$ with n nodes and K views, the estimation procedure for the latent space part of the model in equation 4.3 can be sketched as follows:

1. First, we fix the values for hyperparameters, p, ω and (ξ_1, ξ_2). We randomly initialize the other quantities: G, $\boldsymbol{c}$, $(\pi_1, \dots, \pi_G)$, $\boldsymbol{z}$, $(\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_g)$ and ψ.

2. Given the other current parameters, for $i = 1, \dots, n$, we update sequentially the latent coordinates $\boldsymbol{z}_i$ from their proposal distribution, see Appendix C. To account for rotation and translation invariance of the latent space, we compute the value of the Procrustes correlation between the old set of latent coordinates and the just updated one. If this value is high, the new set of coordinates is discarded.

3. Given the current parameters and latent coordinates, for $i = 1, \dots, n$, we update the cluster labels $\boldsymbol{c}_i$, from their full conditional distribution, see Appendix C.

4. Given the current parameters and latent coordinates, the number of current components G is updated computing the length of $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iG})$, together with the mixture weights π_g (see equation 4.7).

5. Given the current parameters and latent coordinates, mixture component parameters $\boldsymbol{\Omega}_g$ are updated using the full conditional distributions for $\boldsymbol{\mu}_g$ and σ_g^2, presented in Appendix C. Also, if the hyperparameter $\boldsymbol{m}$ is not fixed a priori, it is updated from the corresponding full conditional distribution (Appendix C).

In conclusion, we propose a Metropolis within Gibbs MCMC algorithm which iterates T times between steps $2-5$ of the above procedure and the update steps for the logit parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and the corresponding nuisance parameters. The algorithm is initialized via step 1 of the above procedure. Also, hyperparameters in ϕ_2 are fixed and starting values for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and ϕ_1 are specified. A simulation study to investigate the performance of the proposed estimation procedure is presented in Section 4.3. Some practical issues arise when estimating the latent position cluster model. Next session discusses these issues in details, and the approaches we use to address them.

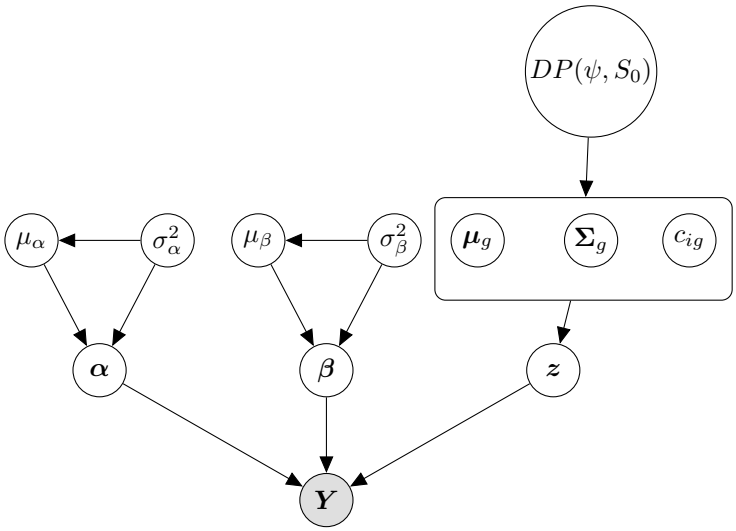


Figure 4.1. Hierarchy structure of the model.

4.4 Practical implementation details

The model proposed by equations 4.3 and 4.1 and the estimation procedure proposed in Section 4.3 present a couple of practical issues which need to be addressed.

4.4.1 Model parameters identifiability

The first issue regards the estimation of the logit model parameters. To guarantee identifiability, a network in a given multiplex has to be taken as reference network,

and the corresponding parameters $(\alpha^{(k)}, \beta^{(k)}) = (\alpha^{(\text{ref})}, \beta^{(\text{ref})})$ need to be fixed. We suggest to set $\beta^{(\text{ref})} = 1$. This constraint does not alter the interpretation of the scale coefficient parameters $\beta^{(k)}$, as their values are meaningful only when compared with each other. We propose to choose $\alpha^{(\text{ref})}$ as suggested in D'Angelo, Murphy, and Alfò, 2018:

$$\alpha^{(\text{ref})} \geq \log\left(\frac{\hat{\bar{p}}}{1-\hat{\bar{p}}}\right) + 2, \tag{4.8}$$

where $\hat{\bar{p}} = \sum_{i=1}^{n}\sum_{j\neq i} y_{ij}^{(\text{ref})}/(n(n-1))$ denotes the observed mean edge probability. The term 2 on the right side of equation 4.8 is the mean empirical distance among coordinates simulated from a standard Gaussian distribution. In the present model, we assume that the prior distribution for the latent coordinates is a mixture of Gaussian distributions with unknown number of components. Therefore, it is not possible to empirically estimate the average distance among coordinates. However, it is reasonable to expect that coordinates drawn from a mixture of Gaussian distributions will be, on average, further apart than coordinates drawn from a single Gaussian distribution. The "greater than or equal to" condition in equation 4.8 comes from this last consideration.

4.4.2 Posterior distributions post-processing

A second issue arises from the estimation of the number of mixture components G, see Section 4.3. At each iteration of the algorithm, the value of G is updated, possibly leading to $G^{(t)} \neq G^{(s)}$, for some value of the iteration index $t, s = 1, \ldots, T$. At the end of the estimation procedure, U unique values of G are proposed. We suggest to select the value associated with the highest posterior probability as the estimated number of components:

$$\hat{G} = \arg\max_{u} \frac{\sum_{t=b}^{T} \mathbf{I}[G^{(t)} = u]}{(T-b+1)}, \tag{4.9}$$

where b is the number of discarded iterations in the burn-in phase. Throughout the iterations, the multiplicity of the U unique G values explored by the algorithm leads to different dimensions for the estimates of the cluster labels $\boldsymbol{c}_i^{(t)}$, $i = 1, \ldots, n$, and the set of component parameters $(\boldsymbol{\Omega}_1^{(t)}, \ldots, \boldsymbol{\Omega}_g^{(t)}, \ldots, \boldsymbol{\Omega}_G^{(t)})$. Hence, the posterior distributions of such quantities need some post-processing procedure, to harmonize their dimensions. Below, we briefly illustrate the post-processing procedures we have adopted in this work.

Cluster labels posterior distribution

At the t^{th} iteration of the algorithm, the length of each $\boldsymbol{c}_i^{(t)}$ vector is $l(\boldsymbol{c})^{(t)} = G^{(t)}$ (Section 4.3). As the number of estimated components may vary throughout the iterations, we have that, if $G^{(t)} \neq G^{(s)}$, then $l(\boldsymbol{c})^{(t)} \neq l(\boldsymbol{c})^{(s)}$, for $s, t = 1, \ldots T$. Such length mismatches imply that the cluster label posterior distribution needs to be post processed in order to get estimates for this variable. To address the issue, we implement the method adopted by Carmona, Nieto-Barajas, and Canale, 2018 and originally proposed by Dahl, 2006. A collection of $n \times n$ co-occurrence

adjacency matrices, denoted by $\boldsymbol{C}^{(t)}$ is computed, $t = b, \ldots, T$. These matrices contain a 1 in position (i, j) if nodes i and j are allocated in the same component at the t^{th} iteration. Further, an average co-occurrence matrix $\bar{\boldsymbol{C}}$ representing the "average clustering" is computed, as the Monte Carlo average of all $\boldsymbol{C}^{(t)}$ adjacency matrices. The $\boldsymbol{C}^{(t)}$ matrix with minimum squared distance from $\bar{\boldsymbol{C}}$ is selected as the estimated co-occurrence matrix. Then, the corresponding t^{th} set of cluster labels $(\boldsymbol{c}_1^{(t)}, \ldots, \boldsymbol{c}_i^{(t)}, \ldots, \boldsymbol{c}_n^{(t)})$ is the set of cluster labels estimates $(\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_i, \ldots, \hat{\boldsymbol{c}}_n)$.

Component parameters posterior distribution

An issue similar to that occurring with cluster labels arises with the posterior distributions of the component parameters $\boldsymbol{\Omega}_g = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, $g = 1, \ldots, G$. Indeed, different iterations of the algorithm (Section 4.3) present a different number of component parameters $\boldsymbol{\Omega}^{(t)} = (\boldsymbol{\Omega}_1^{(t)}, \ldots, \boldsymbol{\Omega}_g^{(t)}, \ldots, \boldsymbol{\Omega}_G^{(t)})$, $t = 1, \ldots, T$. However, after having estimated the number of components $\hat{G}$ (equation 4.9), precisely $\hat{G}$ parameter estimates are needed to describe the components. To solve the problem, one naive approach may be that of taking in consideration only the set of T^* parameters $(\boldsymbol{\Omega}_1^{(t)}, \ldots, \boldsymbol{\Omega}_g^{(t)}, \ldots, \boldsymbol{\Omega}_G^{(t)})$ for which $G^{(t)} = \hat{G}$. The set would contain only those $\boldsymbol{\Omega}^{(t)}$ with matching dimensions, from which $\hat{G}$ Monte Carlo average estimates $\hat{\boldsymbol{\Omega}} = (\hat{\boldsymbol{\Omega}}_1, \ldots, \hat{\boldsymbol{\Omega}}_g, \ldots, \hat{\boldsymbol{\Omega}}_{\hat{G}})$ could be easily computed. However, such a naive approach would not take into consideration the information and the uncertainty brought by those $\boldsymbol{\Omega}^{(t)}$ for which $G^{(t)} \neq \hat{G}$. A more suitable approach, which is adopted in the present paper, has been proposed by Frühwirth-Schnatter, 2011. The author proposes a K-means based procedure to summarize the whole posterior distributions in $(\boldsymbol{\Omega}^b, \ldots, \boldsymbol{\Omega}^T)$. K-means clustering, with $\hat{G}$ clusters, is performed on the whole distribution of $(\boldsymbol{\mu}^b, \ldots, \boldsymbol{\mu}^t, \ldots, \boldsymbol{\mu}^T) = (\boldsymbol{\mu}_1^b, \ldots, \boldsymbol{\mu}_g^b, \ldots, \boldsymbol{\mu}_g^t, \ldots, \boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_G^T) = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m, \ldots, \boldsymbol{\mu}_M)$, and this delivers a classification index $\boldsymbol{I}_m$ for the $m = 1, \ldots, M$ posterior draws. The index is then used to allocate the M posterior draws $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m, \ldots, \boldsymbol{\mu}_M)$ to $\hat{G}$ posterior distributions, one for each component. Also, a permutation test is performed to order the draws and ensure a unique labelling. For further details we refer to Frühwirth-Schnatter, 2011. The same classification index $\boldsymbol{I}_m$, $m = 1, \ldots, M$ is used to reorder the posterior draws for the component covariance matrices: $(\boldsymbol{\Sigma}^b, \ldots, \boldsymbol{\Sigma}^t, \ldots, \boldsymbol{\Sigma}^T) = (\boldsymbol{\Sigma}_1^b, \ldots, \boldsymbol{\Sigma}_g^b, \ldots, \boldsymbol{\Sigma}_g^t, \ldots, \boldsymbol{\Sigma}_1^T, \ldots, \boldsymbol{\Sigma}_G^T) = (\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m, \ldots, \boldsymbol{\Sigma}_M)$. Last, estimates $\hat{\boldsymbol{\Omega}} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ are computed as Monte Carlo averages of the corresponding ordered posterior distributions.

4.5 Simulation study

We have designed a simulation study to evaluate the performance of the proposed infinite mixture model for multiplex data. Our aim is to evaluate the performance of our approach in recovering the latent coordinates, the number of clusters and the cluster allocation. Four simulation scenarios have been defined, in order to analyse the performance of the estimation procedure when different levels of latent space "complexity" are specified. In all four scenarios, we consider two settings with $(n = 25, K = 3)$ and $(n = 50, K = 5)$, respectively. For each scenario and each

dimensionality setting, we generate a bi-dimensional latent space in which positions are generated from a mixture of $G = (2, 3, 4)$ distributions. For each combination of scenario type, dimensionality setting and value of G, we replicate the experiment 10 times. First three scenarios are built generating the latent space according to the assumptions presented in Section 4.2.2. They differ with respect to cluster size and distance between mixture components. The fourth simulation scenario is built to analyze the results of the estimation procedure when the model for the latent space is incorrectly specified. More in details, the scenarios are structured as follows:

- Scenario I. In this first scenario, we aim to evaluate the estimation procedure when the latent coordinates and the clustering components have been generated according to the proposed model (equations 4.3 and 4.1) and the components are of approximately equal size:
  - when $G = 2$, $\pi = (0.5, 0.5)$. The component sizes n_g, $g = 1, 2$, are generated from a Multinomial distribution with parameters (n, π);
  - when $G = 3$, $\pi = (0.\bar{3}, 0.\bar{3}, 0.\bar{3})$. The component sizes n_g, $g = 1, 2, 3$, are generated from a Multinomial distribution with parameters (n, π);
  - when $G = 4$, $\pi = (0.25, 0.25, 0.25, 0.25)$. The component sizes n_g, $g = 1, 2, 3, 4$, are generated from a Multinomial distribution with parameters (n, π).

  Here the clusters are sufficiently separated since, on average, the simulated edge probability is in the interval $(0.40, 0.80)$ when two nodes belong to the same cluster, and is below 0.20 when two nodes belong to different clusters. This scenario corresponds to cohesive and equally proportioned groups, as it could be the case when recording vote exchanges in bipolar political systems, where the members of two main parties vote in agreement with their co-members and not with their opponents.

- Scenario II. As in the first scenario, we have generated the latent coordinates and the clustering components according to the model specified in equations 4.3 and 4.1, and the same ranges are verified for the within and the between cluster edge probabilities. However, here we assume that most of the nodes belong to a single, big component, and the rest is spread into smaller ones:
  - when $G = 2$, $\pi = (0.2, 0.8)$. The component sizes n_g, $g = 1, 2$, are generated from a Multinomial distribution with parameters (n, π);
  - when $G = 3$, $\pi = (0.8, 0.1, 0.1)$. The component sizes n_g, $g = 1, 2, 3$, are generated from a Multinomial distribution with parameters (n, π);
  - when $G = 4$ $\pi = (0.7, 0.1, 0.1, 0.1)$. The component sizes n_g, $g = 1, 2, 3, 4$, are generated from a Multinomial distribution with parameters (n, π).

  This scenario represents the case where few sub-groups of units are isolated from the vast majority of other nodes, at least partially. This could happen, for example, with data regarding relations among a group of students, where the majority of them interacts with one another, but a few students are "unsociable" and only relate to one or two others, as in the Vickers data discussed in Section 4.6.

- Scenario III. In this scenario, multidimensional networks have been simulated according to the model specified in equations 4.3 and 4.1. Weights are simulated from a Uniform distribution, $\pi \sim U(0.3, 0.8)$, and then normalized. Component sizes are generated from a Multinomial distribution with parameters (n, π). Here, when two nodes belong to the same cluster, the average edge probability is in the interval $(0.40, 0.80)$. On the other hand, the average edge probability for two nodes belonging to different clusters is higher than in Scenarios I and II, as it can take value up to 0.40. Such an increase in between clusters edge probabilities corresponds to partially overlapping clusters in the latent space. This situation may appear when some nodes interact quite a lot with nodes from another cluster (or clusters), even if belonging to the same cluster. Recalling the example given in Scenario I, this can happen when two parties in a bipolar system are not that much separated, as some politicians in a given party may agree with various proposals from the other party.

- Scenario IV. This scenario is built to evaluate the estimation procedure when the model for the latent space is incorrectly specified. In particular, we simulate the latent coordinates using a mixture of multivariate non-central Student t distributions, with 3 degrees of freedom. Component sizes are generated as in Scenario I. However, here the average edge probability for two nodes belonging to the same cluster is in the interval $(0.15, 0.5)$. Instead, the average edge probability is below 0.30 when two nodes belong to different clusters. The overlap between different components is greater than that in Scenario III; also, the shapes of the components are different, as the mixture component distributions are misspecified.
  Also, note that a Gaussian and a Student t clustering of the latent coordinates may return the same edge probabilities, as the edge probabilities ultimately depend on the distances between the latent positions. Nonetheless, we test our proposed method in such a misspecified scenario to evaluate up to what extent latent coordinates and cluster allocations can still be recovered.

For all four scenarios, we have run the MCMC procedure described in Section 4.3 for 60000 iterations, with a burn-in of 10000 iterations. We have set $p = 2$, $m_\alpha = m_\beta = 0$, $\nu_\alpha = \nu_\beta = 3$, $\nu_1 = n$, $\nu_2 = 1$, $\tau_\alpha = \tau_\beta = \tau_z = 1$. Small variations of these values have been tried but they did not affect substantially the simulation results. Also, a Jeffreys prior is adopted for the ψ parameter, with the corresponding hyperparameters set to $\xi_1 = \frac{1}{2}$ and $\xi_2 = 1$, see Bernardo and Giron, 1998 and Grazian and Robert, 2015.

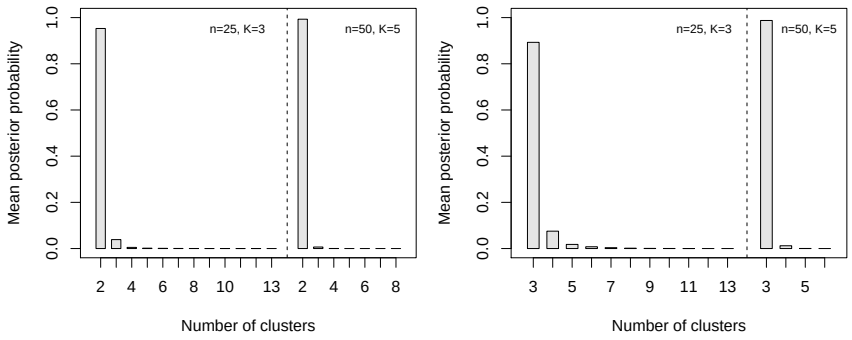
4.5.1 Simulation results

As our interest lies in recovering the latent space positions and the clustering structure, we will primarily focus on such aspects. However, we briefly mention that the "true" intercept values $\alpha^{(k)}$ are always included within a 99% credible interval, while the scale coefficients $\beta^{(k)}$ tend to be overestimated, due to underestimation of the latent distances. However, the products $\beta^{(k)} d_{ij}$ are well recovered, for $i, j = 1, \dots, n$ and $k = 1, \dots, K$. This could be due to the mixture structure which is quite more complex than a single Gaussian distribution. Figures 4.2-4.5 show the (mean)

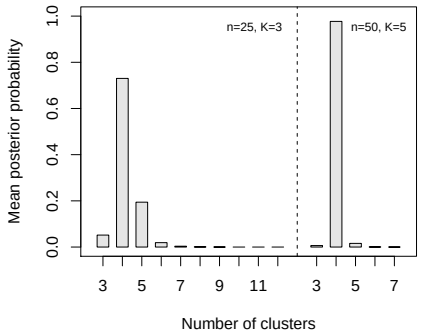
estimated posterior distributions for the number of clusters in the four simulation scenarios. In the first three scenarios, the "true" number of clusters is always recovered with a high posterior probability, with exception of Scenario II when the number of simulated clusters is $G = 4$. In this case, the estimation procedure gives no clear hint to distinguish between the solutions corresponding to $G = 3$ and $G = 4$ mixture components. This can be due to the fact that, when $G = 4$, the "small sized" groups in Scenario II have approximately size $n_g = 3$, when $n = 25$, and $n_g = 5$, when $n = 50$. When small groups are close, or are close to the "big" component, it may be hard to distinguish them and a couple of components may be merged into a single one by the estimation procedure. In scenario IV, the number of mixture components tend to be overestimated, and there is more uncertainty on the estimation of G, as more than one value is associated with quite high posterior probability. However, the number of components is only slightly overestimated, as the estimation procedure suggests values for G which are close to the simulated ones. This proves that, even if the model for the latent space is misspecified, our procedure can still recover rather well the number of clusters.
The results for the estimation of the latent coordinates and the cluster labels are summarized in Tables 4.1-4.4. In the first three groups of columns, the tables report the values of the mean and the standard deviation for the Procrustes correlation between simulated and estimated latent coordinates, for different number of clusters and multiplex dimensions. Such correlations are always quite high, regardless of the considered scenario. These results suggest that the proposed estimation method may be able to recover quite well the latent distances between the nodes, even when the mixture components are not Gaussian. The last three groups of columns display the values of the mean and standard deviation for the Adjusted Rand Index (ARI) (Hubert and Arabie, 1985), computed between the simulated and estimated partition of the nodes. Values of this index close to 1 correspond to a perfectly recovered cluster allocation. In Scenarios I, II, and III, the value of ARI is always greater than 0.85, suggesting that the method is able to reconstruct the latent space and infer the cluster membership of the nodes. Instead, from Table 4.4 we see that the values for the ARI in Scenario IV range between 0.3 and 0.48 and that lower ARI values are associated with higher number of simulated clusters. Even if the ARI values are much lower than those in other scenarios, they represent quite a good result considering that the model was incorrectly specified. Indeed, an ARI close to 0.40 still denotes a clustering solution for the nodes which may be considered of an acceptable quality.

	Procrustes correlation						Adjusted Rand Index					
	$G = 2$		$G = 3$		$G = 4$		$G = 2$		$G = 3$		$G = 4$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
$n = 25$, $K = 3$	0.96	0.01	0.95	0.02	0.75	0.18	0.99	0.03	0.98	0.04	0.98	0.02
$n = 50$, $K = 5$	0.95	0.01	0.95	0.01	0.88	0.15	1.00	0.00	1.00	0.00	0.97	0.01

Table 4.1. Scenario I. Procrustes correlation between estimated and simulated latent spaces and Adjusted Rand Index for the estimated-simulated cluster labels.



(a) $G = 2$ simulated clusters. **(b)** $G = 3$ simulated clusters.



(c) $G = 4$ simulated clusters.

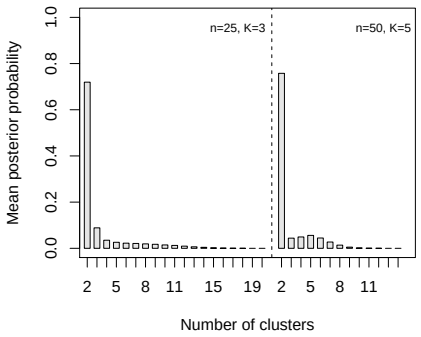
Figure 4.2. Scenario I. (Mean) estimated posterior distribution for the number of clusters G.

	Procrustes correlation						Adjusted Rand Index					
	$G = 2$		$G = 3$		$G = 4$		$G = 2$		$G = 3$		$G = 4$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
$n = 25$, $K = 3$	0.95	0.01	0.92	0.05	0.85	0.14	1.00	0.00	0.91	0.19	0.85	0.19
$n = 50$, $K = 5$	0.96	0.01	0.94	0.03	0.85	0.16	0.91	0.23	0.98	0.06	0.89	0.13

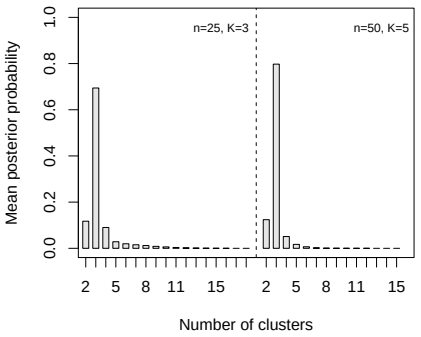
Table 4.2. Scenario II. Procrustes correlation between estimated and simulated latent spaces and Adjusted Rand Index for the estimated-simulated cluster labels.

	Procrustes correlation						Adjusted Rand Index					
	$G = 2$		$G = 3$		$G = 4$		$G = 2$		$G = 3$		$G = 4$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
$n = 25$, $K = 3$	0.95	0.01	0.97	0.01	0.95	0.02	0.96	0.12	0.99	0.02	0.91	0.12
$n = 50$, $K = 5$	0.95	0.02	0.97	0.01	0.97	0.02	0.91	0.11	0.98	0.01	0.85	0.14

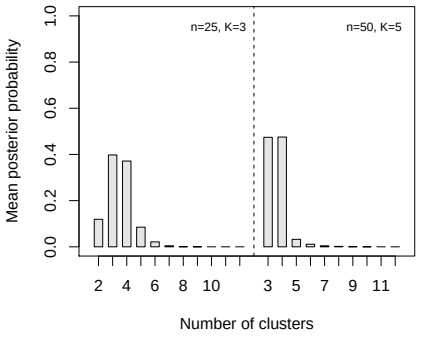
Table 4.3. Scenario III. Procrustes correlation between estimated and simulated latent spaces and Adjusted Rand Index for the estimated-simulated cluster labels.



(a) $G = 2$ simulated clusters.



(b) $G = 3$ simulated clusters.

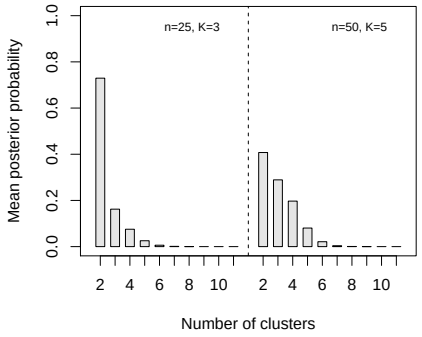


(c) $G = 4$ simulated clusters.

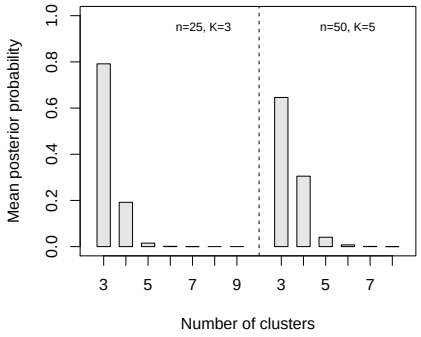
Figure 4.3. Scenario II. (Mean) estimated posterior distribution of the number for clusters G.

	Procrustes correlation						Adjusted Rand Index					
	$G = 2$		$G = 3$		$G = 4$		$G = 2$		$G = 3$		$G = 4$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
$n = 25$, $K = 3$	0.74	0.13	0.71	0.20	0.76	0.24	0.48	0.15	0.42	0.17	0.36	0.10
$n = 50$, $K = 5$	0.87	0.12	0.69	0.14	0.76	0.13	0.46	0.31	0.40	0.18	0.30	0.16

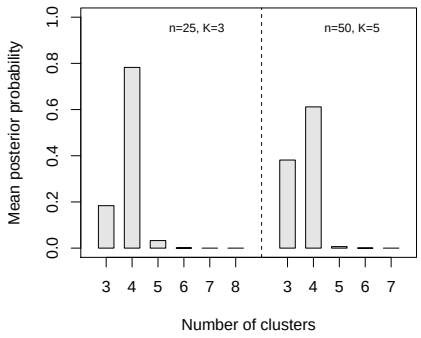
Table 4.4. Scenario IV. Procrustes correlation between estimated and simulated latent spaces and Adjusted Rand Index for the estimated-simulated cluster labels.



(a) $G = 2$ simulated clusters.

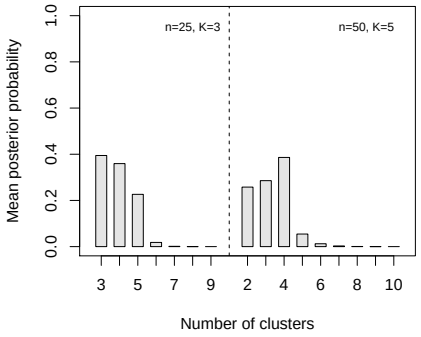


(b) $G = 3$ simulated clusters.

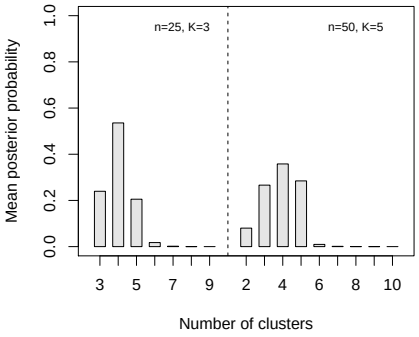


(c) $G = 4$ simulated clusters.

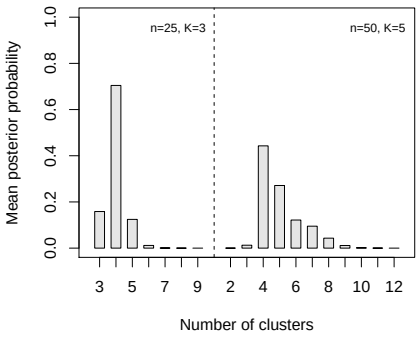
Figure 4.4. Scenario III. (Mean) estimated posterior distribution for the number of clusters G.



(a) $G = 2$ simulated clusters.



(b) $G = 3$ simulated clusters.



(c) $G = 4$ simulated clusters.

Figure 4.5. Scenario IV. (Mean) estimated posterior distribution of the number for clusters G.

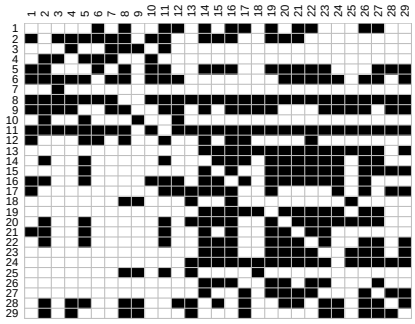
4.6 Vickers multiplex data

To illustrate the proposed model, we propose the re-analysis of the *Vickers-Chan 7th Graders* multidimensional network. The data were collected by Vickers and Chan, 1981 and represent $K = 3$ different social relations among $n = 29$ seventh grade students, twelve boys and seventeen girls, in a school in Victoria, Australia. The analysed relations are:

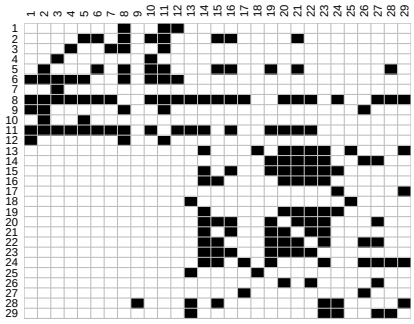
1. *Get on* ($\mathbf{Y}^{(1)}$). This network record whether a student declares to get on with another student;

2. *Best friend* ($\mathbf{Y}^{(2)}$). This network record whether a student declares to be best friend with another student;

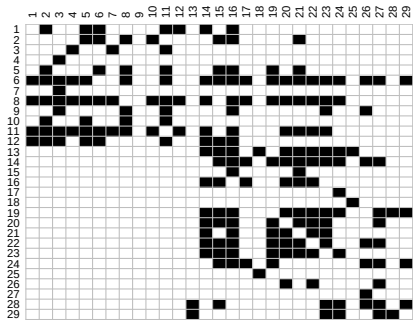
3. *Work with* ($\mathbf{Y}^{(3)}$). This network record whether a student declares to like working with another student;



(a) *Get on* network.



(b) *Best friend* network.



(c) *Work with* network.

Figure 4.6. Vickers data. Adjacency matrices for the three networks.

Figure 4.6 shows the three adjacency matrices for the Vickers multiplex data. Nodes from 1 to 12 are males, while nodes from 13 to 29 are females. Simply by looking at the data, we may hypothesize the presence of sub-groups of students characterized by gender. Indeed, if we partition each adjacency matrix in four sub-matrices,

accordingly to the gender of the students,

$$\left\{ \mathbf{Y}^{(k)}_{mm} = \mathbf{Y}^{(k)}_{[1:12,1:12]}, \quad \mathbf{Y}^{(k)}_{ff} = \mathbf{Y}^{(k)}_{[13:29,13:29]}, \quad \mathbf{Y}^{(k)}_{mf} = \mathbf{Y}^{(k)}_{[1:12,13:29]}, \quad \mathbf{Y}^{(k)}_{fm} = \mathbf{Y}^{(k)}_{[13:29,1:12]} \right\},$$

and compute the observed density in each sub-matrix,

get on				best friend				work with			
$\mathbf{Y}^{(1)}_{mm}$	$\mathbf{Y}^{(1)}_{ff}$	$\mathbf{Y}^{(1)}_{mf}$	$\mathbf{Y}^{(1)}_{fm}$	$\mathbf{Y}^{(2)}_{mm}$	$\mathbf{Y}^{(2)}_{ff}$	$\mathbf{Y}^{(2)}_{mf}$	$\mathbf{Y}^{(2)}_{fm}$	$\mathbf{Y}^{(3)}_{mm}$	$\mathbf{Y}^{(2)}_{ff}$	$\mathbf{Y}^{(3)}_{mf}$	$\mathbf{Y}^{(3)}_{fm}$
0.59	0.60	0.40	0.19	0.44	0.35	0.14	0.00	0.44	0.36	0.21	0.00

we see that the density is much higher in sub-matrices composed by students of the same gender. However, all the three $\mathbf{Y}^{(k)}_{mf}$ sub-matrices have quite high density values too. Therefore, we may expect clustering of the nodes, but it is hard to tell, simply by looking at the data, how many clusters we should expect. Are the students simply separated by gender or not? Are there mixed-gender sub-groups in the data? Also, the four sub-matrices present similar density values in the three networks, with the views "best friend" and "work with" having practically identical densities. Further, the associations computed between couples of adjacency matrices[4] are quite high: $A(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}) = 0.79$, $A(\mathbf{Y}^{(1)}, \mathbf{Y}^{(3)}) = 0.77$ and $A(\mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}) = 0.90$. These last two facts may suggest that a single latent space can be employed to describe the overall similarities between the students.

4.6.1 Vickers data: results

We have run the MCMC procedure described in Section 4.3 for 60000 iterations, with a burn-in of 10000 iterations. The hyperparameters have been set as in the simulation study, see Section 4.5; in particular, the parameters of the Dirichlet prior were set to $\xi_1 = \frac{1}{2}$ and $\xi_2 = 1$. Another uninformative prior specification was tested, $\xi_1 = \xi_2 = 1$, with no substantial change in the results. For visualization purposes, the dimension of the latent space has been set to $p = 2$.
As it can be seen from Figure 4.7, our procedure estimated $G = 4$ mixture components in the latent space representation of the Vickers data. The posterior probability associated with this solution is quite high, almost 0.80. To recover cluster labels for the students, we adopted the post processing procedure by Dahl, 2006, described in Section 4.4.2. Figure 4.8 shows the average co-occurrence matrix, displaying how often two nodes have been assigned to the same component, and the final cluster label matrix, resulting from the post-processing procedure. The assignment of the nodes to different components is quite clear, as it can be seen from the co-occurrence matrix. These assignments can be visualized more clearly in the estimated cluster labels matrix:

1. Green component ($g = 1$). This component has the largest estimated size, 11, with 11 out of the 12 male students in Vickers data.

[4]The association between couple of adjacency matrices is defined as:

$$A(\mathbf{Y}^{(k)}, \mathbf{Y}^{(l)}) = \frac{\sum_{i=1}^{n} \sum_{j \neq i} \mathbf{I}(y^{(k)}_{ij} = (y^{(l)}_{ij})}{n(n-1)}, \quad k, l = 1, \ldots, K.$$

2. Orange component ($g = 2$). This component has the smallest estimate size and it is the only mixed-gender one. Its elements are male student 9 and female students 18 and 25.

3. Blue component ($g = 3$). Most of the female students in Vickers data are assigned to this component, which has size 10.

4. Purple component ($g = 4$). The last component has size 5 and contains female students number $13, 17, 24, 28, 29$.

Combining the information coming from the two matrices in Figure 4.8, we see that, although cluster labels have been assigned quite clearly, there is a bit of uncertainty regarding a small group of nodes. For example, male student 9 was often placed with female students in a different component, but never with other male students. Another example is that of female student number 24, assigned to the purple component, which has non-null posterior probability of co-occurring with many other students outside of this component. In general, segmentation by gender is quite evident. The only estimated mixed-gender component is quite small and contains students with a low number of links, as it can be seen in Figure 4.6. These students might have been placed in the same component, "isolated" by the vast majority of others, due to their "low sociability". The uncertainty in cluster labels assignment for some of the nodes is reflected in the posterior distributions of the component means, displayed in the left plot in Figure 4.9 and obtained using the procedure by Frühwirth-Schnatter, 2011, described in Section 4.4.2. Indeed, while the posterior distribution for the green component mean is well separated from the others, there is some overlap between the orange and the purple component and between the purple and the blue component. The estimated node latent coordinates are displayed in the right plot in Figure 4.9. These are coloured by cluster affiliation, and components are indicated by their estimated means (black dots) and standard deviations (ellipses). The latent space proximity between some of the components is reflected in the average estimated edge posterior probabilities within and between clusters in Table 4.5, for the three Vickers networks. Indeed, between clusters probability values are higher among the three "female" components, purple, blue and orange. Higher probability values are estimated in correspondence of the first network. In the *best friend* and *work with* networks the average posterior probability of observing an edge between a male student of the green component and any other student from other components is quite low, namely below 0.11. Instead, in these last two networks the between cluster probabilities for the orange, blue and purple components remain relatively high, except for the couple orange/blue component. The average estimated edge posterior probabilities within and between clusters have been computed using the estimated latent distances and the estimated logit parameters:

get on				best friend				work			
α		β		α		β		α		β	
mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
2.50	-	1.00	-	0.53	0.15	1.13	0.10	0.42	0.14	0.89	0.08

The α and β parameters have been fixed in the first network for identifiability reasons, as discussed in Section 4.4.1. The values of the estimated intercepts in the last two networks are much lower than that of the first view, as we could have expected by looking at observed networks in Figure 4.6. Indeed, the *get on* network is much denser than the other two. The β parameters are estimated to be close to 1 in all the three networks. This suggests that the effect of the latent space on the edge probabilities is quite constant across the different views.

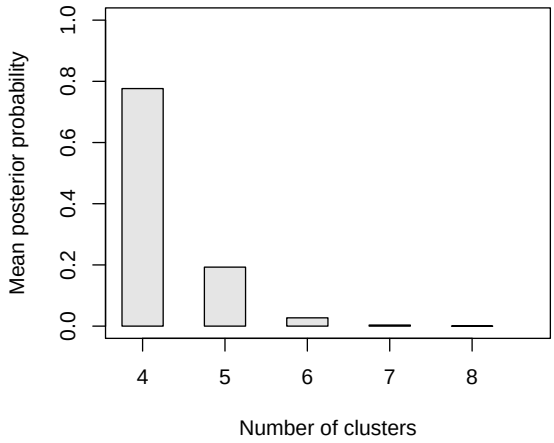
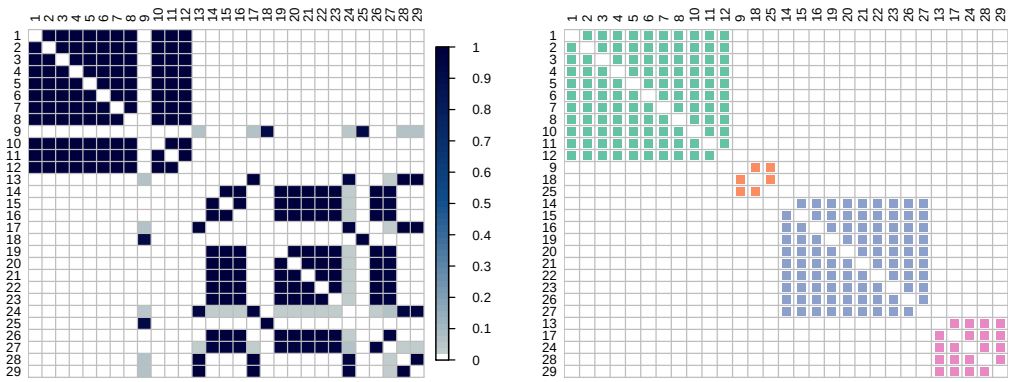


Figure 4.7. Vickers data. Posterior distributions of the number of clusters.

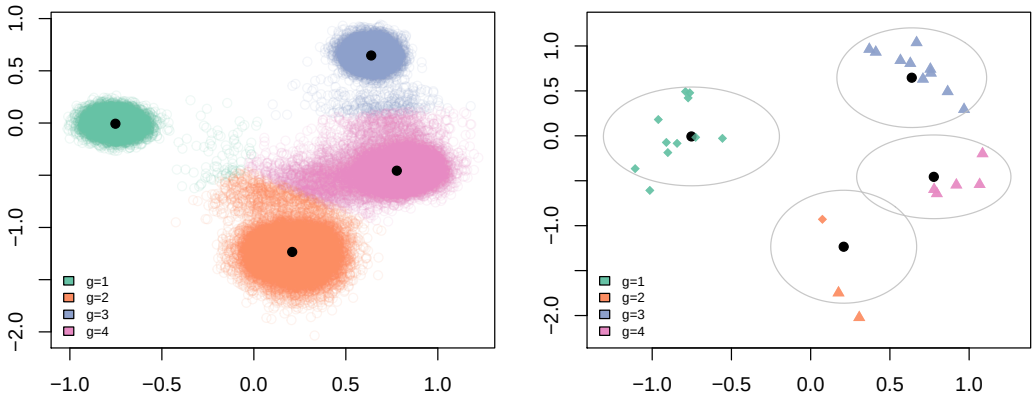


(a) Average co-occurrence **(b)** Estimated cluster labels.

Figure 4.8. Vickers data. Average co-occurrence of the nodes in the same cluster and estimated cluster labels.

4.7 Discussion

In this work, we introduced an *infinite latent position cluster model* for clustering nodes in multidimensional network data. Our model allows to address transitivity and clustering in (multidimensional) networks via a latent space representation of



(a) Component means posterior distributions. **(b)** Latent coordinates, component means (in black) and standard deviations (ellipses).

Figure 4.9. Vickers data. Posterior distributions of component means and estimated latent coordinates. In the right plot, triangles indicate female students, while dots correspond to male students.

	get on				best friend				work with			
	$g=1$	$g=2$	$g=3$	$g=4$	$g=1$	$g=2$	$g=3$	$g=4$	$g=1$	$g=2$	$g=3$	$g=4$
$g=1$	0.90	0.29	0.39	0.27	0.55	0.07	0.07	0.04	0.54	0.10	0.11	0.07
$g=2$	0.29	0.85	0.12	0.63	0.07	0.45	0.02	0.20	0.10	0.46	0.03	0.24
$g=3$	0.39	0.12	0.91	0.67	0.07	0.02	0.58	0.22	0.11	0.03	0.57	0.26
$g=4$	0.27	0.63	0.67	0.92	0.04	0.20	0.22	0.60	0.07	0.24	0.26	0.58

Table 4.5. Vickers data. Average estimated edge probabilities within and between clusters.

the nodes, whose coordinates are assumed to be drawn from an infinite mixture of Gaussian distributions. Differently from existing clustering methods for network data, thanks to the infinite mixture setting, our proposal performs joint inference on the latent coordinates, the component parameters, and the number of mixture components. Model estimation is carried out within a Bayesian hierarchical approach via a MCMC algorithm, whose performance has been tested in a simulation study and showed to give results of good quality. Treating the number of component as a parameter allows to avoid multiple model comparisons. Also, it avoids the problem of choosing which model selection criterion to adopt for the comparison, which is not a trivial task. The proposed framework and estimation procedure have been applied to the well known Vickers multiplex data(Vickers and Chan, 1981). In this application, our procedure has shown to be useful in capturing and visualizing a clustering structure in the nodes/students, which could not be identified by looking at the networks alone. Four sub-groups of students have been detected, corresponding to four components. Cluster allocation appeared to be quite influenced by the gender of the students, with 11 out of the 12 male students placed in the same component. However, this was not the only clustering factor, as female students were split into three different groups: two were large all-female components, while the remaining one is a small mixed-gender group. One of the all-females groups

can be considered the cluster of most "sociable" students, as they are close to the other two main sub-groups. The small component estimated in the *Vickers* example is composed by nodes which tend to make only few connections. Indeed, in social network data there could be some nodes who poorly interact with others and may not present a clear propensity to cluster. Such nodes are "quasi-isolated" from the others, as they do not express a clear "social" behaviour. An interesting extension of the proposed model may be that of accounting for a "quasi-isolated" component in the mixture, which collects all "quasi-isolated" nodes in (multidimensional) network data. An interesting consequence of this extra component would be that of removing the possibility of having components with a single node allocated to it. Indeed, single-node components would be grouped in a wider "quasi-isolated" one. The other components would then contain at least two nodes, reducing the number of possible components from $G \leq n$ to $G \leq \frac{n}{2} + 1$.

Missing edges and edge-specific covariates may be easily included in the model, as done in D'Angelo, Murphy, and Alfò, 2018 for example. The inclusion of covariates would allow to address homophily by attributes, as done in Handcock, Raftery, and Tantrum, 2007. Also, a more general specification of the component covariance matrices may be allowed, e.g. by assuming an Inverse Wishart prior distribution for such terms; only small changes would be needed in the estimation procedure we have described.

The proposed models will be incorporated shortly in the R package *spaceNet* (D'Angelo and Fop, 2018), already available on CRAN.

Chapter 5

Discussion

This thesis contributes to the current research on latent space modelling of network and multidimensional network data, where multiple relations are recorded among the same set of nodes. Chapter 1 provides a general introduction to network analysis. Earlier models in the area are presented, as well as more recent approaches, with particular emphasis on the class of latent space models.
Chapters 2-4 discuss novel models for latent space network analysis. In particular, Chapter 2 extends the latent space network model by Gollini and Murphy, 2016 to analyse voting data from the Eurovision Song Contest. Different editions of the contest are represented as a set of networks, where nodes are the participating countries and edges represent votes exchanged between them. The nodes are assumed to lay in a single Euclidean latent space, common for all the networks. Proximities between countries in such space are taken to be proxies of countries "cultural" similarities. The aim is to investigate possible bias in the exchange of votes. To do so, we estimate the relevance of latent space coordinates in determining the observed exchanged votes in each network/edition of the contest. Also, estimated latent coordinates provide a representation of participating countries in terms of their "cultural locations", at least in terms of musical preferences expressed through votes in the Eurovision Song Contest. Edge covariates might be incorporated in the framework, with geographical and cultural covariates being used for modelling of the Eurovision data.
Although the model presented in Chapter 2 has been developed focusing on the Eurovision application, it can have wide applicability to other types of multiplex data. In particular, it may be applied to multidimensional networks describing different interactions between subjects or to multiplexes regarding trading relations between entities (countries, industries, and so on). In general, this model is apt to capture overall similarities between nodes in a multiplex. For example, latent coordinates may be proxies of cultural similarities (as in the case of the Eurovision) or of character affinities (as for human-interaction networks).
Chapter 3 presents an extension to the latent space network model developed in Chapter 2. In particular, a class of nine latent space models is introduced, to represent various levels of in-/out-degree variability for networks in multiplex data. These nine models are characterized in terms of different sender and receiver node effects. Such effects may be either null, constant or variable across the networks, and they

characterize nodes behaviours in different views of the multiplex. The idea is that persistent characteristics of the nodes may be captured through their latent space representation, while deviations from such characteristics depend on the particular network nodes have to interact in. We use the sender-receiver parametrization to model import-export data between countries. Here, different networks correspond to different sub-markets, and the activity level of each country in a given sub-market may be captured via its sender/receiver effect. In this context, the latent space provides a representation of the overall trading relations between countries. This class of models may be used to analyse all types of multiplex data for which nodes behaviours in single networks are believed to be variations from their usual "common" behaviours.

Chapter 4 introduces the "infinite latent position cluster model" for network and multidimensional network data. Nodes latent coordinates are assumed to arise from an infinite mixture of Gaussian distributions, where each mixture component in the latent space represents a group of nodes in the observed multiplex. The infinite mixture model allows for a flexible characterization of different multiplex data, as the number of components does not have to be specified in advanced. Also, the number of components directly depends on the nodes in the data, meaning that new actors entering the multidimensional network may lead to a different number of estimated components or to a different characterization of the components. This model is illustrated via an application to a three-networks multiplex, representing relations among a class of students. However, it can be applied to all types of network and multiplex data, whenever a clustering structure is hypothesized for the nodes.

The models introduced in this thesis are all developed in the context of binary networks/multidimensional networks. It may be of interest to extend them to allow for weighted edges, see for example Sewell and Chen, 2016 or Hoff, Fosdick, and Stovel, 2013. Also, in the case of the infinite latent position cluster model, it would be interesting to introduce a non-Gaussian component, to capture possible anti-social behaviour of some nodes. Indeed, not all nodes in a network/multiplex may tend to "group" with others; a non-Gaussian component (for example, a Uniform component) could help to locate such nodes and to distinguish between "sociable" and "non-sociable" nodes.

The Bayesian framework is exploited to perform parameter estimation. In particular, a hierarchical Bayesian approach is adopted to derive MCMC algorithms for all the models discussed in this thesis. The performances of the proposed algorithms are investigated through extensive simulation studies. Furthermore, an R package containing some of these algorithms has been developed, *spaceNet*, and it is available on CRAN.

Appendices to the Chapters provide detailed descriptions of the MCMC algorithms, as well as other details on the proposed models. Last, Appendix D discusses a re-implementation of the method by Gollini and Murphy, 2016, with an MCMC algorithm derived as an alternative to the original variational one.

Appendix A

Appendix to Chapter 2

A.1 Posterior distributions for the parameters

This section will present the full conditional and the proposal distributions that have been derived from the posterior distribution of the model. For ease of calculation, the log-posterior distribution is considered:

$$\begin{aligned}\log\Big(P(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z},\mu_\alpha,&\mu_\beta,\sigma^2_\alpha,\sigma^2_\beta|\mathbf{Y})\Big)\propto\\ &\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}\Big[y_{ij}^{(k)}(\alpha^{(k)}-\beta^{(k)}d_{ij})-\log\Big(1+\exp\{\alpha^{(k)}-\beta^{(k)}d_{ij}\}\Big)\Big]\\ &-\frac{1}{2}\Bigg\{\sum_{i=1}^{n}z_i^2+\frac{\sum_{k=1}^{K}(\alpha^{(k)}-\mu_\alpha)^2}{\sigma^2_\alpha}+\frac{\sum_{k=1}^{K}(\beta^{(k)}-\mu_\beta)^2}{\sigma^2_\beta}+K\log(\sigma^2_\alpha)\,\cdot\\ &+K\log(\sigma^2_\beta)+\log(\tau_\alpha\sigma^2_\alpha)+\log(\tau_\beta\sigma^2_\beta)+\frac{\mu^2_\alpha}{\tau_\alpha\sigma^2_\alpha}+\frac{\mu^2_\beta}{\tau_\beta\sigma^2_\beta}+\frac{1}{\sigma^2_\alpha}+\frac{1}{\sigma^2_\beta}\Bigg\}\\ &+\Big(-\frac{\nu_\alpha}{2}-1\Big)\log(\sigma^2_\alpha)+\Big(-\frac{\nu_\beta}{2}-1\Big)\log(\sigma^2_\beta)\end{aligned}$$

Without any loss of information, the latent coordinates are assumed to be univariate. A generalization to the multivariate case is straightforward and will be considered when the proposal distribution for the latent positions is introduced.

A.1.1 Nuisance parameters

All the posterior distributions for the nuisance parameters have closed form. In particular, the variances follow Inverse Gamma distributions

$$\sigma^2_\alpha|\boldsymbol{\alpha},\mu_\alpha,\tau_\alpha,\nu_\alpha,K\sim\text{InvΓ}(r_\alpha,R_\alpha);\qquad\sigma^2_\beta|\boldsymbol{\beta},\mu_\beta,\tau_\beta,\nu_\beta,K\sim\text{InvΓ}(r_\beta,R_\beta),$$

with parameters:

$$r_x=\frac{\nu_x+K+1}{2},\qquad R_x=\frac{\tau_x+\tau_x\sum_{k=1}^{K}(x^{(k)}-\mu_x)^2+\mu_x^2}{2\tau_x}.$$

The mean parameters μ_α, μ_β are distributed as truncated normal distributions:

$$\mu_\alpha|\boldsymbol{\alpha}, \sigma_\alpha^2, \tau_\alpha, m_\alpha, K \sim N_{\left[LB(\alpha),\infty\right]}\left(\frac{\tau_\alpha \sum_{k=1}^K \alpha^{(k)} + m_\alpha}{1+K\tau_\alpha}, \frac{\tau_\alpha \sigma_\alpha^2}{1+K\tau_\alpha}\right),$$

$$\mu_\beta|\boldsymbol{\beta}, \sigma_\beta^2, \tau_\beta, m_\beta, K \sim N_{\left[0,\infty\right]}\left(\frac{\tau_\beta \sum_{k=1}^K \beta^{(k)} + m_\beta}{1+K\tau_\beta}, \frac{\tau_\beta \sigma_\beta^2}{1+K\tau_\beta}\right).$$

A.1.2 Latent positions

In order to derive a proposal distribution for the latent coordinates, only the terms containing the variables $\boldsymbol{z}_i$ in the log-posterior distribution are considered:

$$\log\Big(P(\boldsymbol{z}_i \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\beta})\Big) = -\frac{1}{2}z_i^2 - \sum_{k=1}^K \sum_{j\neq i} h_{ij}^{(k)} \left[y_{ij}^{(k)} \beta^{(k)} d_{ij} + \log\Big(1+\exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\Big)\right].$$

To approximate the logarithmic term in the above equation, let us recall the definition of the LSE (Log Sum Exponential) function,

$$\log\Big(\sum_{g=1}^G \exp(x_g)\Big),$$

that is known to be bounded between:

$$\max\{x_1, \ldots, x_G\} \leq \log\Big(\sum_{g=1}^G \exp(x_g)\Big) \leq \max\{x_1, \ldots, x_G\} + \log(G).$$

The logarithmic term in the log-posterior part of interest for the latent coordinates has indeed LSE form, thus it can be bounded between:

$$\max\{0, \alpha^{(k)} - \beta^{(k)} d_{ij}\} \leq \log\Big(\exp\{0\}\exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\Big) \leq \max\{0, \alpha^{(k)} - \beta^{(k)} d_{ij}\} + \log(2).$$

The lower bound is met when only one of the term in the summation is non-zero, which is the case. Defining the binary variable $w_{ij}^{(k)}$ as,

$$w_{ij}^{(k)} = \begin{cases} 1 & \text{if} \quad \alpha^{(k)} - \beta^{(k)} d_{ij} > 0 \\ 0 & \text{if} \quad \alpha^{(k)} - \beta^{(k)} d_{ij} \leq 0 \end{cases},$$

then the logarithmic term of interest can be approximated using its lower bound. The log-posterior terms in z_i can be approximated as:

$$\begin{aligned}
&-\frac{1}{2}z_i^2 - \sum_{k=1}^K \sum_{j\neq i} h_{ij}^{(k)} \left[y_{ij}^{(k)} \beta^{(k)} d_{ij} + \max\{0, \alpha^{(k)} - \beta^{(k)} d_{ij}\}\right] \\
\propto &-\frac{1}{2}z_i^2 - \sum_{k=1}^K \sum_{j\neq i} h_{ij}^{(k)} \left[y_{ij}^{(k)} \beta^{(k)} (z_i^2 + z_j^2 - 2z_i z_j) + w_{ij}^{(k)} \Big(\alpha^{(k)} \theta_i \gamma_j - \beta^{(k)} (z_i^2 + z_j^2 - 2z_i z_j)\Big)\right] \\
\propto &-\frac{1}{2}z_i^2 - \sum_{k=1}^K \sum_{j\neq i} h_{ij}^{(k)} \left[y_{ij}^{(k)} \beta^{(k)} (z_i^2 - 2z_i z_j) + w_{ij}^{(k)} \Big(-\beta^{(k)} (z_i^2 - 2z_i z_j)\Big)\right] \\
= &-\frac{1}{2}z_i^2 \left[1 + 2\sum_{k=1}^K \beta^{(k)} \sum_{j\neq i} h_{ij}^{(k)} (y_{ij}^{(k)} - w_{ij}^{(k)})\right] + z_i \left[2\sum_{k=1}^K \beta^{(k)} \sum_{j\neq i} h_{ij}^{(k)} (y_{ij}^{(k)} - w_{ij}^{(k)}) z_j\right].
\end{aligned}$$

The last equality returns a quadratic form in z_i and could be used to specify a proposal distribution for the latent coordinates. However, it is not guaranteed that the quantity multiplying z_i^2, which would be the precision of the proposal distribution for latent position i, is strictly positive. Indeed, defining $n_i^{(k)} = \sum_{j \neq i} y_{ij}^{(k)}$,

$$
\begin{aligned}
1 + 2\sum_{k=1}^{K} \beta^{(k)} \sum_{j \neq i} h_{ij}^{(k)} (y_{ij}^{(k)} - w_{ij}^{(k)}) &\geq \min\left\{ 1 + 2\sum_{k=1}^{K} \beta^{(k)} \sum_{j \neq i} h_{ij}^{(k)} (y_{ij}^{(k)} - w_{ij}^{(k)}) \right\} \\
&\propto 1 + 2\Big[\sum_{k=1}^{K} \beta^{(k)} n_i^{(k)} - \max\{ \sum_{k=1}^{K} \beta^{(k)} \sum_{j \neq i} w_{ij}^{(k)} \} \Big]. \\
&= 1 + 2\Big[\sum_{k=1}^{K} \beta^{(k)} \big(n_i^{(k)} - (n-1) \big) \Big]
\end{aligned}
$$

In the second row node indicator variables are fixed $h_{ij}^{(k)} = 1$. Notice that $n_i^{(k)} \leq n - 1 \implies (n_i^{(k)} - (n-1)) \leq 0$. Hencefor the precision to be strictly positive it should be that:

$$
\begin{aligned}
&\min\left\{ 1 + 2\sum_{k=1}^{K} \beta^{(k)} \sum_{j \neq i} (y_{ij}^{(k)} - w_{ij}^{(k)}) \right\} > 0 \iff 2\Big[\sum_{k=1}^{K} \beta^{(k)} \big(n_i^{(k)} - (n-1) \big) \Big] > -1 \\
&\iff \sum_{k=1}^{K} \beta^{(k)} \big(n_i^{(k)} - (n-1) \big) > -\frac{1}{2}.
\end{aligned}
$$

One of the coefficient has to be fixed for identifiability issue. Taking $\beta^{(1)} = 1$, the above inequality may be rewritten as:

$$
\begin{aligned}
\sum_{k=1}^{K} \beta^{(k)} \big(n_i^{(k)} - (n-1) \big) &\geq \sum_{k=1}^{K} \beta^{(k)} (1-n) \\
&= (1-n) + \sum_{k=2}^{K} \beta^{(k)} (1-n) \geq -\frac{1}{2}.
\end{aligned}
$$

The last inequality holds only when $n = 1$. This implies that the quantity

$$
1 + 2\sum_{k=1}^{K} \beta^{(k)} \sum_{j \neq i} h_{ij}^{(k)} (y_{ij}^{(k)} - w_{ij}^{(k)})
$$

may not always be positive. Therefore, an alternative specification for the proposal precision is needed. The binary variables $w_{ij}^{(k)}$ can be thought as potential-link indicator, as they depend on the argument of the numerator in the edge probability. In that sense, the difference term $(y_{ij}^{(k)} - w_{ij}^{(k)})$ can be considered as a measure of the goodness of link classification. However, the sum of these terms, $\sum_{j \neq i} (y_{ij}^{(k)} - w_{ij}^{(k)})$, could be null either in presence of perfect classification or in the presence of perfect misclassification. In fact, $\sum_{j \neq i} y_{ij}^{(k)} = \sum_{j \neq i} w_{ij}^{(k)} \implies \sum_{j \neq i} (y_{ij}^{(k)} - w_{ij}^{(k)}) = 0$. To overcome this problem, the absolute value of the difference is considered. The

proposal distribution specified for latent coordinate i is then: We can then define the proposal distribution for the i^{th} latent coordinate as:

$$\tilde{z}_i \mid \mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{D}, K \sim N\Big(\mu_{\tilde{z}_i}, \sigma^2_{\tilde{z}_i}\Big),$$

where

$$\mu_{\tilde{z}_i} = \sigma^2_{\tilde{z}_i}\Big(2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}h_{ij}^{(k)}(y_{ij}^{(k)}-w_{ij}^{(k)})z_j\Big), \quad \sigma^2_{\tilde{z}_i} = \left(1+2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}h_{ij}^{(k)}|y_{ij}^{(k)}-w_{ij}^{(k)}|\right)^{-1}.$$

In the case of multivariate latent coordinates, with number of dimensions p, $\mu_{\tilde{\boldsymbol{z}}_i}$ would be a $p-$dimensional vector and the covariance matrix would be diagonal, with generic element $\sigma^2_{\tilde{z}_i}$.

A.1.3 Intercept parameters

The part of interest of the log-posterior distribution to derive a proposal distribution for the intercept terms is:

$$\log\Big(P(\alpha^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \beta^{(k)}, \mu_\alpha, \sigma^2_\alpha, n\Big) \propto \sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}\Big[y_{ij}^{(k)}(\alpha^{(k)}-\beta^{(k)}d_{ij}) - \log\Big(1+\exp\{\alpha^{(k)}-\beta^{(k)}d_{ij}\}\Big)\Big] - \frac{1}{2}\frac{(\alpha^{(k)}-\mu_\alpha)^2}{\sigma^2_\alpha}.$$

The logarithmic term is approximate with its second order Taylor expansion around μ_α:

$$\log\Big(1+\exp\{\alpha^{(k)}-\beta^{(k)}d_{ij}\}\Big)\Big|_{\alpha^{(k)}=\mu_\alpha} \approx \log\Big(1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}\Big) + (\alpha^{(k)}-\mu_\alpha)\frac{\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}}{1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}} + \frac{1}{2}(\alpha^{(k)}-\mu_\alpha)^2\frac{\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}}{\big(1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}\big)^2}.$$

Substituting the logarithmic term with its approximation in the log posterior yields to:

$$\log\Big(P(\alpha^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \beta^{(k)}, \mu_\alpha, \sigma^2_\alpha)\Big) \propto -\frac{1}{2}\frac{(\alpha^{(k)}-\mu_\alpha)^2}{\sigma^2_\alpha} + \sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}\left[y_{ij}^{(k)}\alpha^{(k)} - \alpha^{(k)}\frac{\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}}{1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}} - \frac{1}{2}(\alpha^{(k)}-\mu_\alpha)^2\frac{\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}}{\big(1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}\}\big)^2}\right]$$

that is a quadratic form in $\alpha^{(k)}$. Defining $E^{(k)} = \sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}y_{ij}^{(k)}$, the proposal distribution for intercept $\alpha^{(k)}$ is taken to be:

$$\tilde{\alpha}^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \beta^{(k)}, \mu_\alpha, \sigma^2_\alpha, n \sim N\Big(\mu_{\tilde{\alpha}^{(k)}}, \sigma^2_{\tilde{\alpha}^{(k)}}\Big),$$

with

$$\mu_{\tilde{\alpha}^{(k)}} = \sigma^2_{\tilde{\alpha}^{(k)}} \left\{ E^{(k)} - \sum_{i=1}^{n} \sum_{j \neq i} \frac{h_{ij}^{(k)} \exp\{\mu_\alpha - \beta^{(k)} d_{ij}\}}{1 + \exp\{\mu_\alpha - \beta^{(k)} d_{ij}\}} \right\} + \mu_\alpha$$

$$\sigma^2_{\tilde{\alpha}^{(k)}} = \left\{ \sum_{i=1}^{n} \sum_{j \neq i} \frac{h_{ij}^{(k)} \exp\{\mu_\alpha - \beta^{(k)} d_{ij}\}}{\left(1 + \exp\{\mu_\alpha - \beta^{(k)} d_{ij}\}\right)^2} + \frac{1}{\sigma^2_\alpha} \right\}^{-1}.$$

A.1.4 Coefficient parameters (distances)

The part of interest for the coefficient parameters in the posterior distribution is:

$$\log\Big(P(\beta^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \alpha^{(k)}, \mu_\beta, \sigma^2_\beta, n\Big) \propto \sum_{i=1}^{n} \sum_{j \neq i} h_{ij}^{(k)} \Big[y_{ij}^{(k)} (\alpha^{(k)} - \beta^{(k)} d_{ij}) - \log\Big(1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\Big)\Big] - \frac{1}{2} \frac{(\beta^{(k)} - \mu_\beta)^2}{\sigma^2_\beta}.$$

The logarithmic term is approximated with its second order Taylor expansion in μ_β:

$$\log\Big(1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\Big)\Big|_{\beta^{(k)} = \mu_\beta} \approx \log\Big(1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}\Big) - (\beta^{(k)} - \mu_\beta) \frac{d_{ij} \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}} + \frac{1}{2} (\beta^{(k)} - \mu_\beta)^2 \frac{d_{ij}^2 \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{\left(1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}\right)^2}.$$

The logarithmic term is replaced with its approximation, leading to:

$$\log\Big(P(\beta^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \alpha^{(k)}, \mu_\beta, \sigma^2_\beta)\Big) \propto -\frac{1}{2} \frac{(\beta^{(k)} - \mu_\beta)^2}{\sigma^2_{beta}} + \sum_{i=1}^{n} \sum_{j \neq i} h_{ij}^{(k)} \Bigg[-y_{ij}^{(k)} \beta^{(k)} d_{ij} + \beta^{(k)} \frac{d_{ij} \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}} - \frac{1}{2} (\beta^{(k)} - \mu_\beta)^2 \frac{d_{ij}^2 \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{\left(1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}\right)^2}.$$

The above expression is quadratic in $\beta^{(k)}$. Then, the proposal distribution specified for intercept coefficient k is:

$$\tilde{\beta}^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \alpha^{(k)}, \mu_\beta, \sigma^2_\beta \sim N\Big(\mu_{\tilde{\beta}^{(k)}}, \sigma^2_{\tilde{\beta}^{(k)}}, n\Big),$$

where

$$\mu_{\tilde{\beta}^{(k)}} = \sigma^2_{\tilde{\beta}^{(k)}} \left\{ \sum_{i=1}^{n} \sum_{j \neq i} h_{ij}^{(k)} d_{ij} \left(\frac{\exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}} - y_{ij}^{(k)} \right) \right\} + \mu_\beta,$$

$$\sigma^2_{\tilde{\beta}^{(k)}} = \left\{ \sum_{i=1}^{n} \sum_{j \neq i} \frac{h_{ij}^{(k)} d_{ij}^2 \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}}{\left(1 + \exp\{\alpha^{(k)} - \mu_\beta d_{ij}\}\right)^2} + \frac{1}{\sigma^2_\beta} \right\}^{-1}.$$

A.1.5 Coefficient parameters (covariates)

The coefficient parameters of the covariates are assumed to be independent but not identically distributed. We consider the prior distribution for coefficient parameter of set of covariates f being:

$$\lambda_f \sim N_{[0,\infty]}(\mu_{\lambda_f}, \sigma^2_{\lambda_f})$$

The parameters μ_{λ_f} and $\sigma^2_{\lambda_f}$ are nuisance parameters, distributed as:

$$\mu_{\lambda_f}|\sigma^2_{\lambda_f} \sim N_{[0,\infty]}(m_\lambda, \tau_\lambda\sigma^2_{\lambda_f}) \quad \sigma^2_{\lambda_f} \sim \text{Inv}\chi^2_{\nu_\lambda}$$

Then the log-posterior distribution has to be slightly modified with the introduction of covariates and the extra parameters:

$$\begin{aligned}
\log\Big(P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}, \mu_\alpha, \mu_\beta, \sigma^2_\alpha, \sigma^2_\beta, \boldsymbol{\lambda}, \mu_\lambda, \sigma^2_\lambda|\mathbf{Y}, \mathbf{X})\Big) \propto \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i} h_{ij}^{(k)}\Big[y_{ij}^{(k)}(\alpha^{(k)} - \beta^{(k)}d_{ij} \\
- \sum_{f=1}^{F}\lambda_f x_{ijf}) - \log\Big(1 + \exp\{\alpha^{(k)} - \beta^{(k)}d_{ij}\sum_{f=1}^{F}\lambda_f x_{ijf}\}\Big)\Big] \\
- \frac{1}{2}\Bigg\{\sum_{i=1}^{n} z_i^2 + \frac{\sum_{k=1}^{K}(\alpha^{(k)} - \mu_\alpha)^2}{\sigma^2_\alpha} + \frac{\sum_{k=1}^{K}(\beta^{(k)} - \mu_\beta)^2}{\sigma^2_\beta} + K\log(\sigma^2_\alpha) \\
+ K\log(\sigma^2_\beta) + \log(\tau_\alpha\sigma^2_\alpha) + \log(\tau_\beta\sigma^2_\beta) + \frac{\mu^2_\alpha}{\tau_\alpha\sigma^2_\alpha} + \frac{\mu^2_\beta}{\tau_\beta\sigma^2_\beta} + \frac{1}{\sigma^2_\alpha} + \frac{1}{\sigma^2_\beta} \\
- \frac{1}{2}\sum_{f=1}^{F}\left(\frac{(\lambda_f - \mu_{\lambda_f})^2}{\sigma^2_{\lambda_f}} + \log(\sigma^2_{\lambda_f}) + \log(\tau_\lambda\sigma^2_{\lambda_f}) + \frac{(\mu_{\lambda_f} - m_\lambda)^2}{\tau_\lambda\sigma^2_{\lambda_f}} + \frac{1}{\sigma^2_{\lambda_f}}\right)\Bigg\} \\
+ \Big(-\frac{\nu_\alpha}{2} - 1\Big)\log(\sigma^2_\alpha) + \Big(-\frac{\nu_\beta}{2} - 1\Big)\log(\sigma^2_\beta) + \sum_{f=1}^{F}\Big(-\frac{\nu_{\lambda_f}}{2} - 1\Big)\log(\sigma^2_{\lambda_f}).
\end{aligned}$$

Given a set F covariates f, the corresponding nuisance parameters The nuisance parameters μ_{λ_f} and $\sigma^2_{\lambda_f}$ are update via the following full conditional distributions:

$$\sigma^2_{\lambda_f} \mid \lambda_f, \mu_{\lambda_f}, \tau_\lambda, \nu_\lambda \sim \text{Inv}\Gamma\left(\frac{\nu_\lambda + 2}{2}, \frac{\tau_\lambda + \tau_\lambda(\lambda_f - \mu_{\lambda_f})^2 + \mu^2_{\lambda_f}}{2\tau_\lambda}\right),$$

$$\mu_{\lambda_f} \mid \lambda_f, \sigma^2_{\lambda_f}, \tau_\lambda, m_\lambda \sim N_{[0,\infty]}\left(\frac{\tau_\lambda\lambda_f + m_\lambda}{1 + \tau_\lambda}, \frac{\tau_\lambda\sigma^2_{\lambda_f}}{1 + \tau_\lambda}\right).$$

The update of the parameters λ_f is carried out with a Metropolis-Hastings algorithm, where the proposal distribution used for the parameter f is derived approximating the logarithmic term in the log-posterior with its second order Taylor series expansion in $\lambda_f = \mu_{\lambda_f}$. Then, the proposal distribution for λ_f is:

$$\tilde{\lambda}_f \sim N_{(0,\infty)}(\mu_{\tilde{\lambda}_f}, \sigma^2_{\tilde{\lambda}_f}),$$

where

$$\mu_{\tilde{\lambda}_f} = \sigma^2_{\tilde{\lambda}_f}\left[\sum_{k=1}^{K}\left(\sum_{i=1}^{n}\sum_{j\neq i}\frac{x_{ijf}\exp\left(\alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{l\neq f}\lambda_l x_{ijl}\right)}{1+\exp\left(\alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{l\neq f}\lambda_l x_{ijl}\right)}-y_{ij}^{(k)}x_{ijf}\right)\right]+\mu_{\lambda_f},$$

$$\sigma^2_{\tilde{\lambda}_f} = \left\{\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}\frac{x_{ijf}^2\exp\left(\alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{l\neq f}\lambda_l x_{ijl}\right)}{\left(1+\exp\left(\alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{l\neq f}\lambda_l x_{ijl}\right)\right)^2}+\frac{1}{\sigma^2_{\lambda_f}}\right\}^{-1}.$$

A.1.6 Other proposal distributions

The introduction of covariates in the model requires a convenient adjustment of the proposal distribution for the latent coordinates, the intercepts and the coefficients. The proposal distribution for latent position i has the same form as the one shown in section A.1.2, but the definition of the auxiliary variables $w_{ij}^{(k)}$ changes in:

$$w_{ij}^{(k)} = \begin{cases} 1 & \text{if} \quad \alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf} > 0 \\ 0 & \text{if} \quad \alpha^{(k)}-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf} \leq 0 \end{cases},$$

The mean and variance parameters of the proposal distribution for the intercept terms $\alpha^{(k)}$ are modified by the introduction of the covariates, while the form of the distribution is left unchanged:

$$\mu_{\tilde{\alpha}^{(k)}} = \sigma^2_{\tilde{\alpha}^{(k)}}\left\{E^{(k)}-\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}\exp\{\mu_\alpha-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}{1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}\right\}+\mu_\alpha,$$

$$\sigma^2_{\tilde{\alpha}^{(k)}} = \left\{\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}\exp\{\mu_\alpha-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}{\left(1+\exp\{\mu_\alpha-\beta^{(k)}d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}\right)^2}+\frac{1}{\sigma^2_\alpha}\right\}^{-1}$$

The mean and variance parameters of the proposal distribution for the coefficient parameters $\beta^{(k)}$ changes, though the form of the distribution remains the same:

$$\mu_{\tilde{\beta}^{(k)}} = \sigma^2_{\tilde{\beta}^{(k)}}\left\{\sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}d_{ij}\left(\frac{\exp\{\alpha^{(k)}-\mu_\beta d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}{1+\exp\{\alpha^{(k)}-\mu_\beta d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}-y_{ij}^{(k)}\right)\right\}+\mu_\beta,$$

$$\sigma^2_{\tilde{\beta}^{(k)}} = \left\{\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}d_{ij}^2\exp\{\alpha^{(k)}-\mu_\beta d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}}{\left(1+\exp\{\alpha^{(k)}-\mu_\beta d_{ij}-\sum_{f=1}^{F}\lambda_f x_{ijf}\}\right)^2}+\frac{1}{\sigma^2_\beta}\right\}^{-1}$$

A.2 ISO3 codes

Country name	iso3 code	Country name	iso3 code
Albania	ALB	Italy	ITA
Andorra	AND	Latvia	LVA
Armenia	ARM	Lithuania	LTU
Australia	AUS	Malta	MLT
Austria	AUT	Moldova	MDA
Azerbaijan	AZE	Monaco	MCO
Belarus	BLR	Montenegro	MNE
Belgium	BEL	Norway	NOR
Bosnia and Herzegovina	BIH	Poland	POL
Bulgaria	BGR	Portugal	PRT
Croatia	HRV	Romania	ROU
Cyprus	CYP	Russia	RUS
Czech Republic	CZE	San Marino	SMR
Denmark	DNK	Serbia	SRB
Estonia	EST	Serbia and Montenegro	SCG
Federal Republic of Macedonia	MKD	Slovakia	SVK
Finland	FIN	Slovenia	SVN
France	FRA	Spain	ESP
Georgia	GEO	Sweden	SWE
Germany	DEU	Switzerland	CHE
Greece	GRC	Netherlands	NLD
Hungary	HUN	Turkey	TUR
Iceland	ISL	Ukraine	UKR
Ireland	IRL	United Kingdom	GBR
Israel	ISR		

Table A.1. Eurovision data: country ISO3 codes.

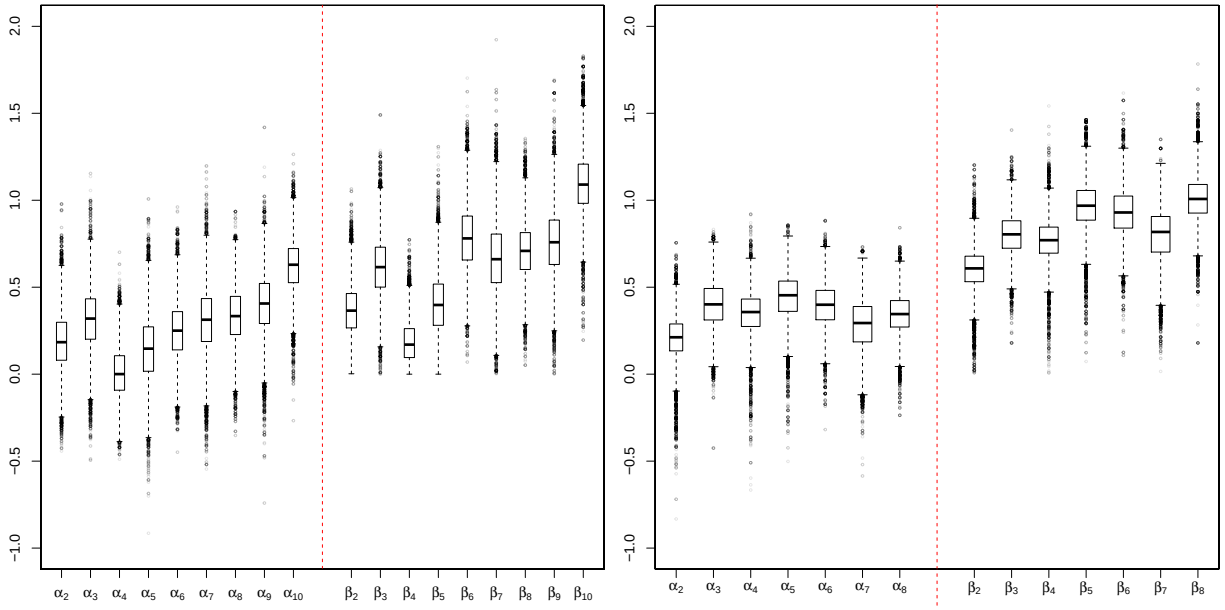
A.3 Results for the Eurovision sub-periods

Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$	Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$
1998	0	-	1	-	2003	0.25	0.17	0.79	0.19
1999	0.19	0.17	0.37	0.15	2004	0.31	0.19	0.67	0.21
2000	0.32	0.18	0.62	0.17	2005	0.33	0.16	0.71	0.17
2001	0.01	0.15	0.19	0.12	2006	0.40	0.18	0.76	0.19
2002	0.15	0.19	0.40	0.17	2007	0.63	0.15	1.10	0.18

Table A.2. Eurovision data: estimated averages and standard deviations for the network parameters in the multiplex 1998-2007.

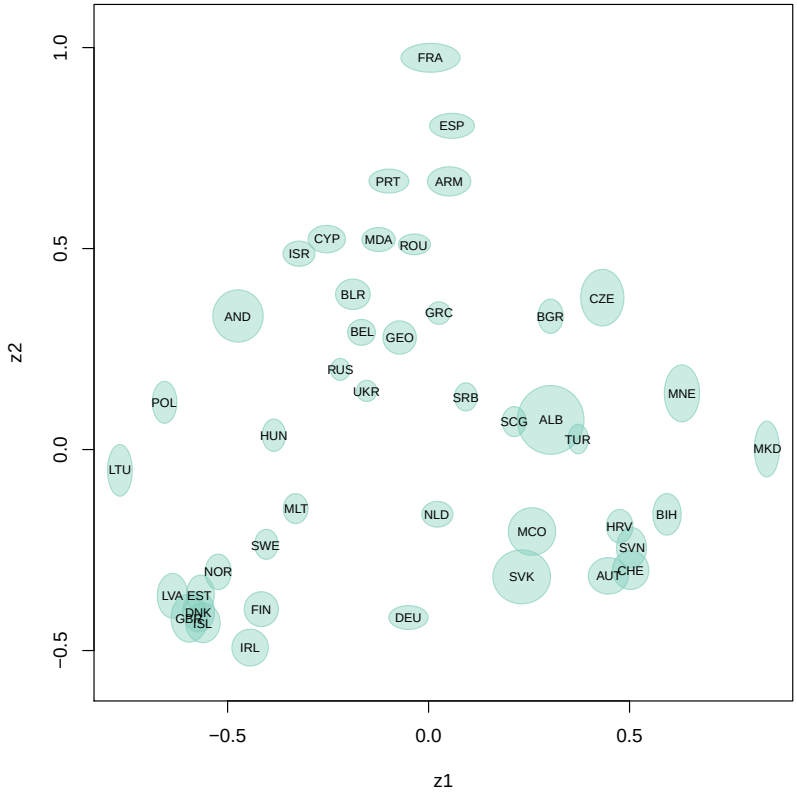
Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$	Year	$\hat{\alpha}$	$sd(\alpha)$	$\hat{\beta}$	$sd(\beta)$
2008	0	-	1	-	2012	0.44	0.14	0.96	0.15
2009	0.21	0.14	0.60	0.13	2013	0.40	0.13	0.93	0.14
2010	0.40	0.14	0.80	0.13	2014	0.28	0.16	0.80	0.16
2011	0.35	0.13	0.77	0.13	2015	0.35	0.12	1.01	0.14

Table A.3. Eurovision data: estimated averages and standard deviations for the network parameters in the multiplex 2008-2015.

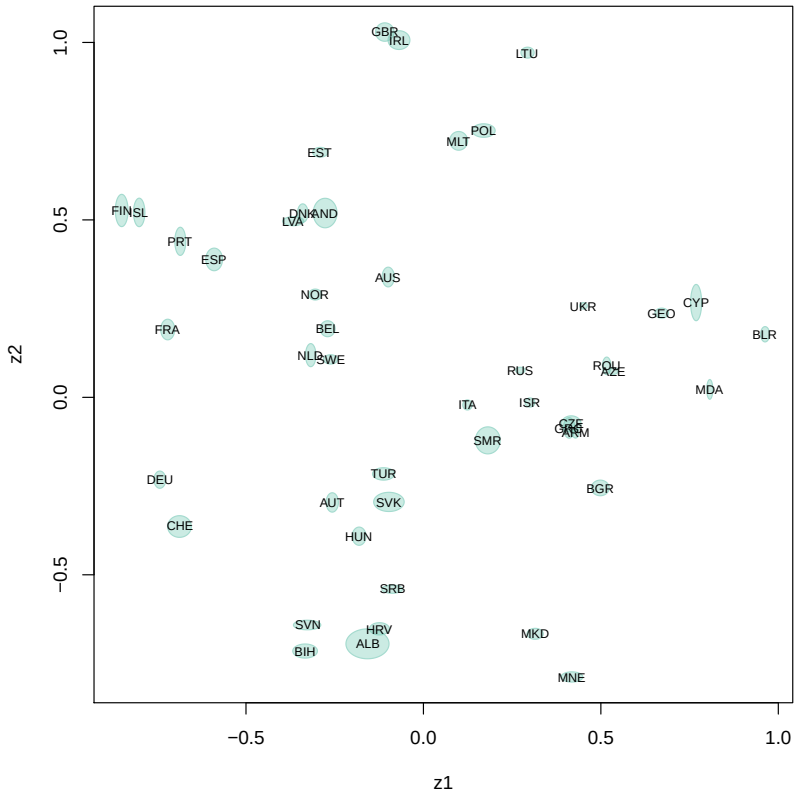


(a) 1998-2007. **(b)** 2008-2015

Figure A.1. Eurovision data: boxplots for the estimates of the logistic parameters, periods 1998-2007 and 2008-2015.



(a) 1998-2007.



(b) 2008-2015.

Figure A.2. Eurovision data: estimated latent positions, periods 1998-2007 and 2008-2015.

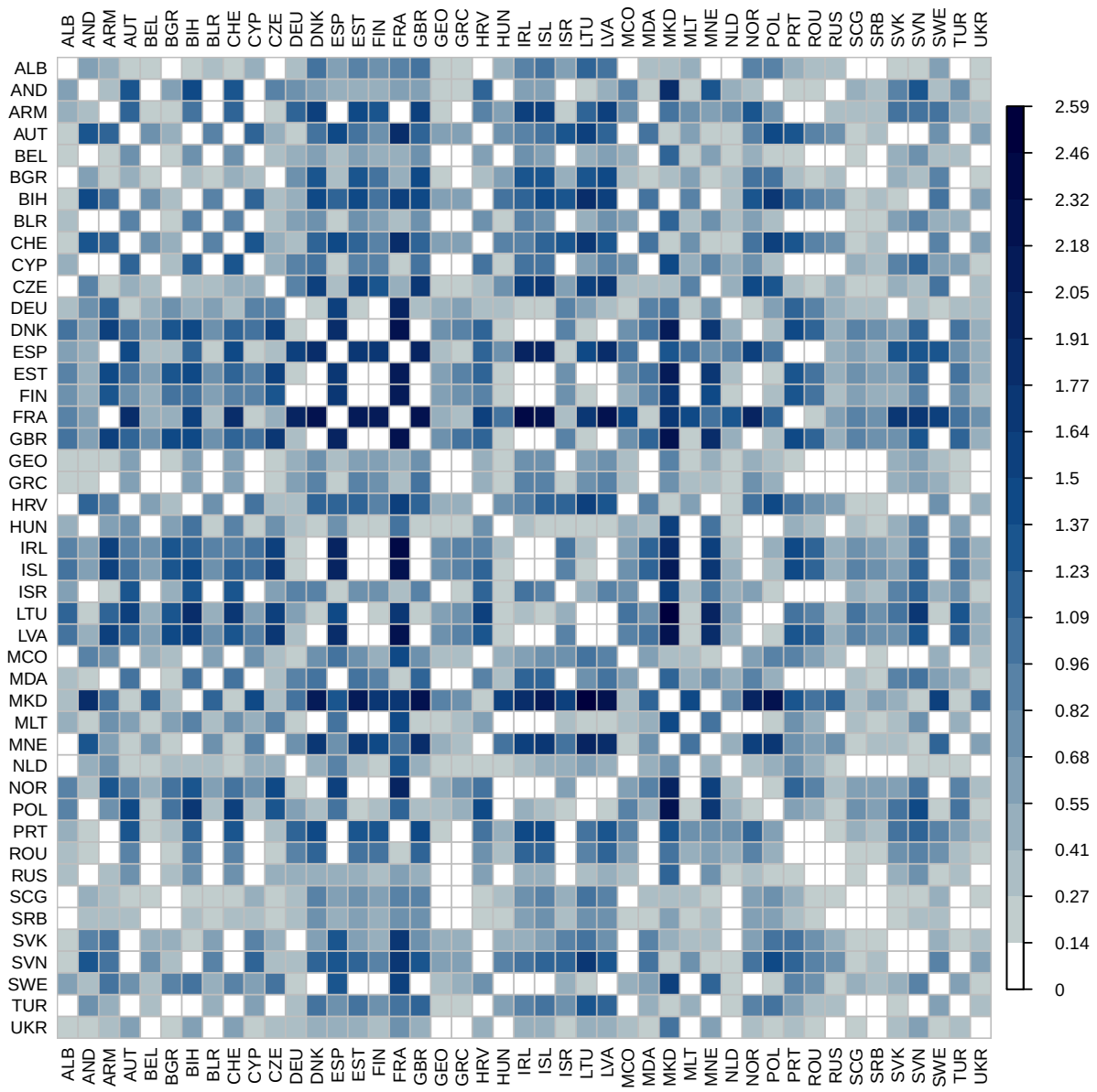


Figure A.3. Eurovision data: estimated distances between couple of countries for the period 1998-2007.

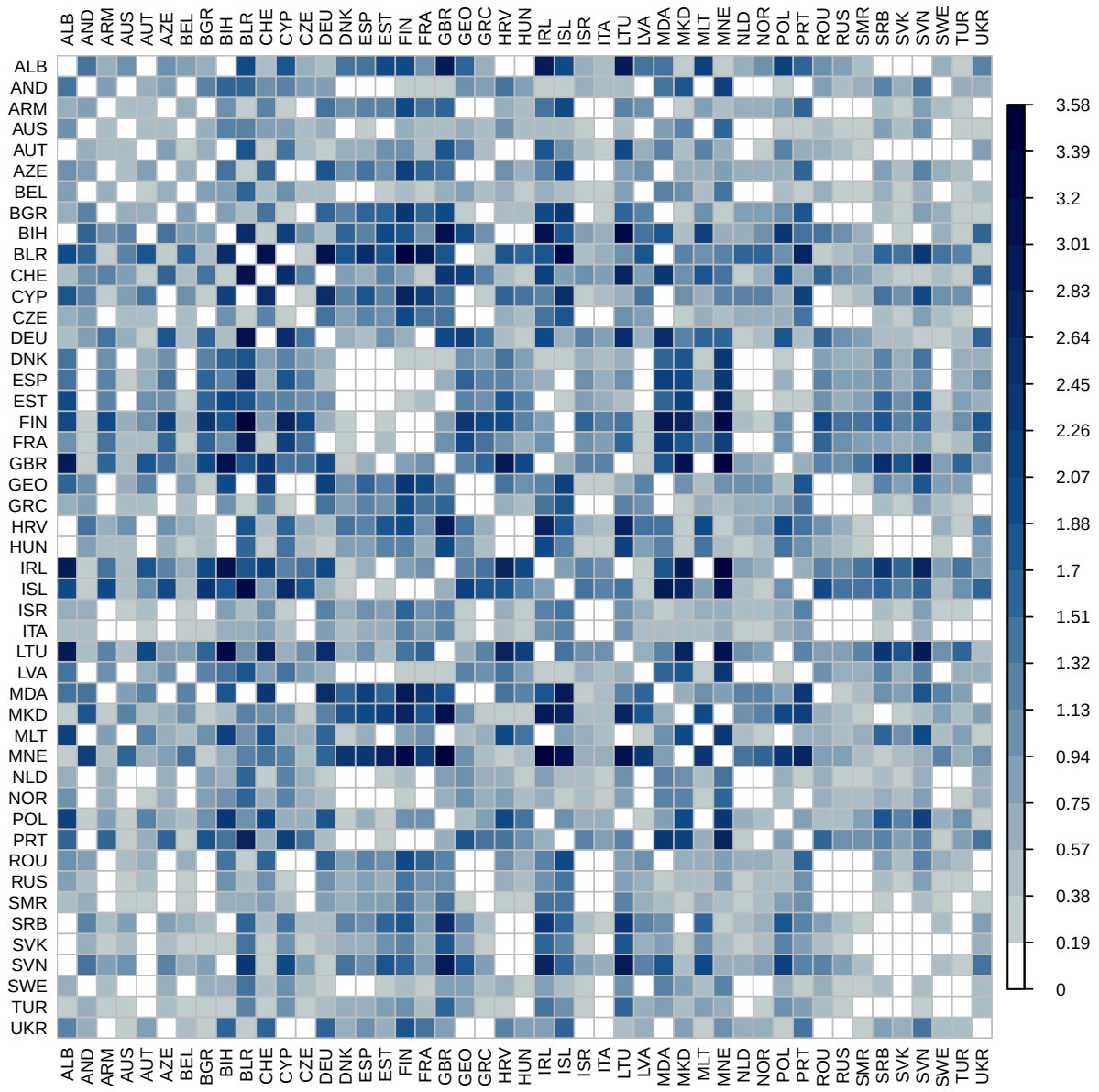
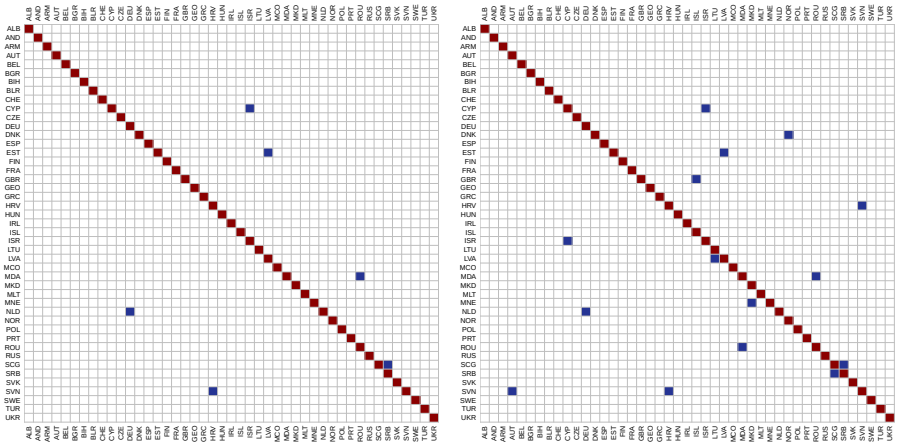
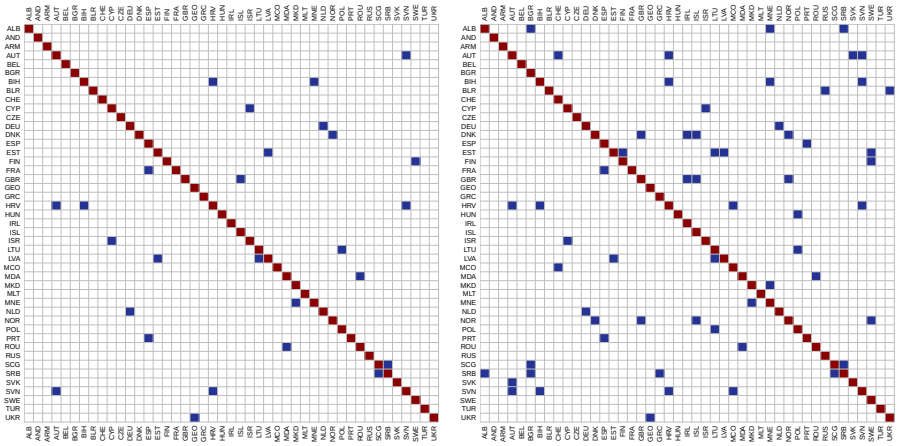


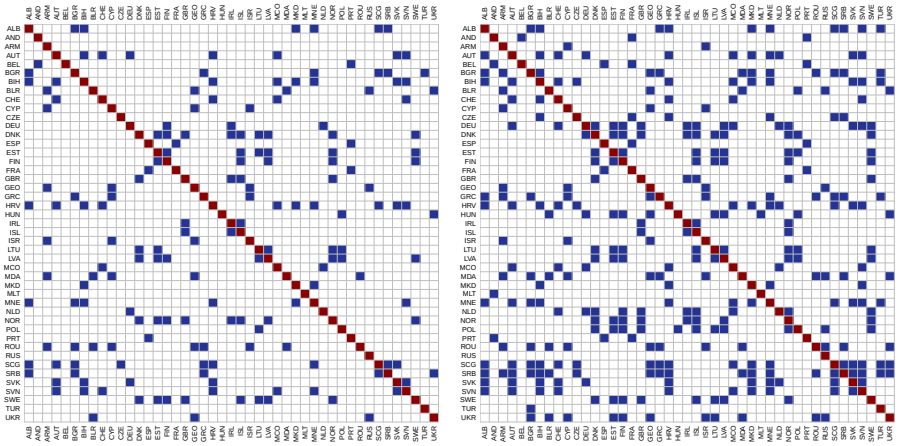
Figure A.4. Eurovision data: estimated distances between couple of countries for the period 2008-2015.



(a) $neigh = 1$ **(b)** $neigh = 2$



(c) $neigh = 3$ **(d)** $neigh = 5$



(e) $neigh = 10$ **(f)** $neigh = 15$

Figure A.5. Eurovision data: intersections of the set of neighbours $LN_{i,r}$ and $GN_{i,r}$, for the period 1998-2007.

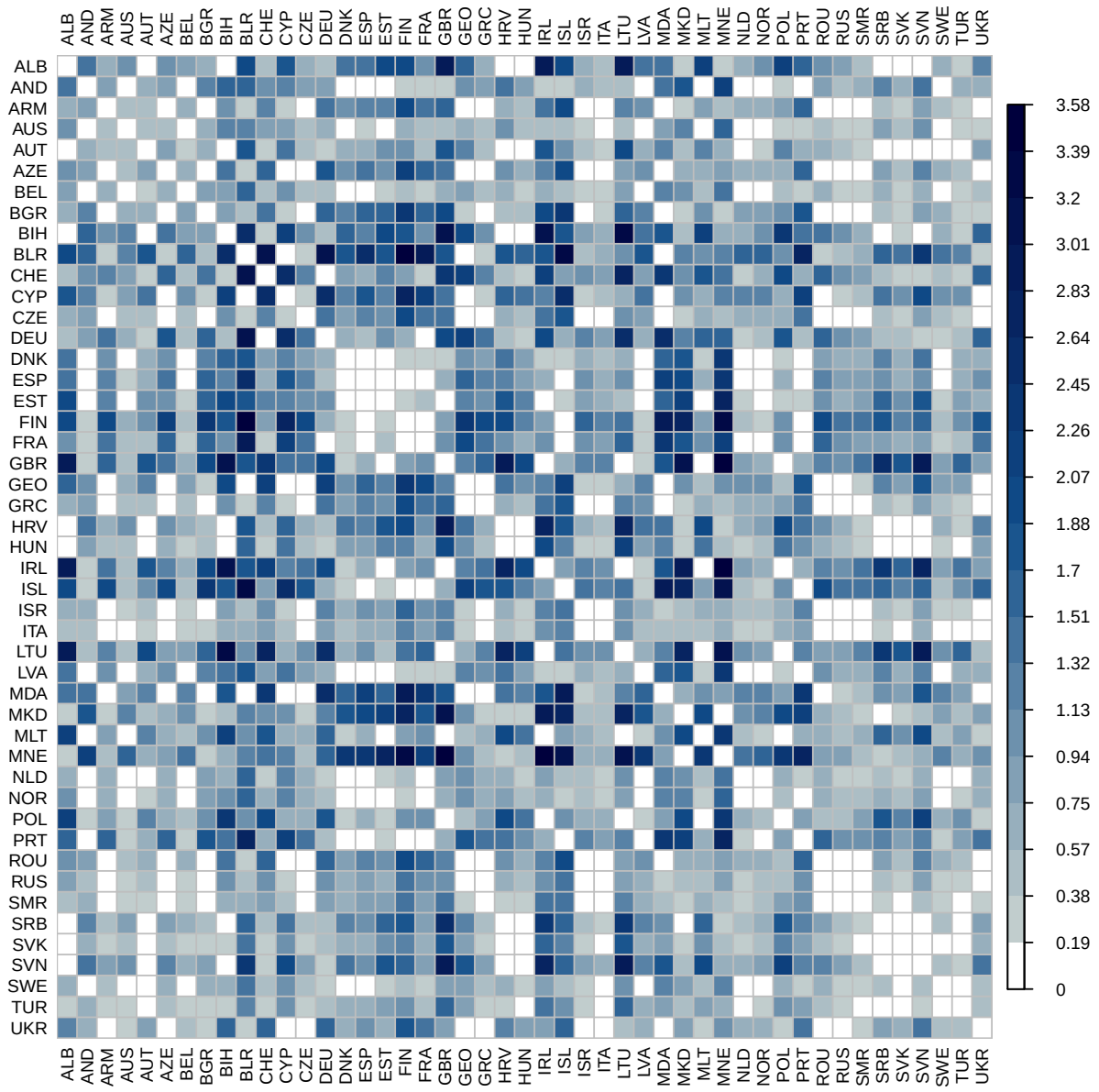
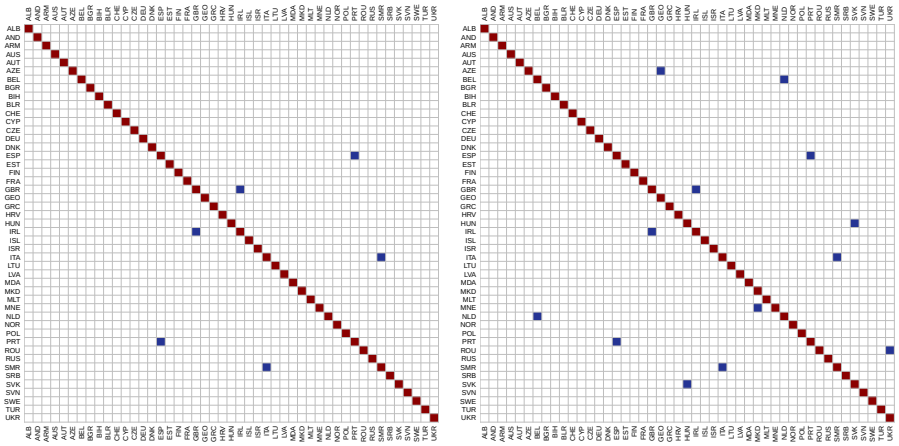
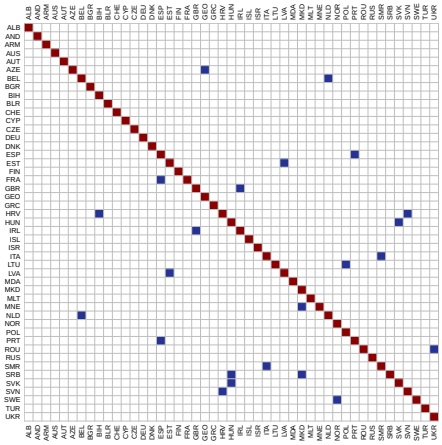


Figure A.6. Eurovision data: estimated distances between couple of countries for the period 2008-2015.

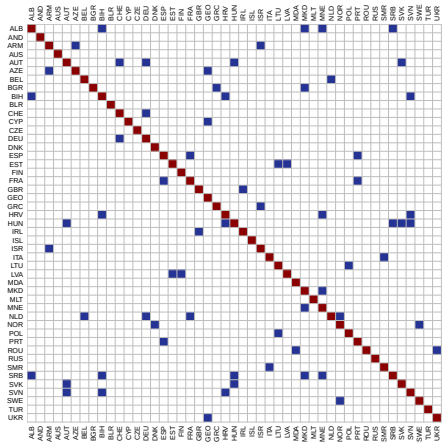


(a) $neigh = 1$

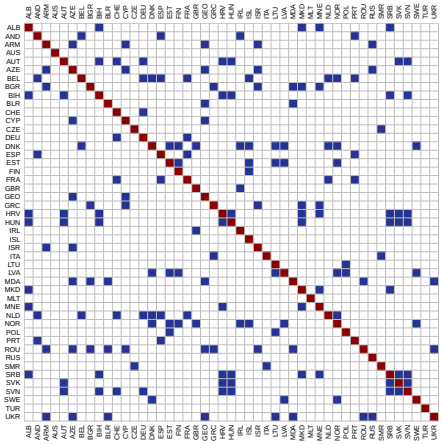
(b) $neigh = 2$



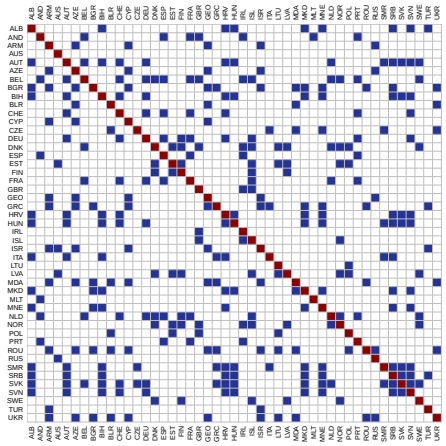
(c) $neigh = 3$



(d) $neigh = 5$



(e) $neigh = 10$



(f) $neigh = 15$

Figure A.7. Eurovision data: intersections of the set of neighbours $LN_{i,r}$ and $GN_{i,r}$, for the period 2008-2015.

A.4 Simulations results

A.4.1 Results for block I

Table A.4. Simulated values for the intercept and the coefficient terms in the multidimensional networks considered in scenarios I-III.

	$K=2$		$K=3$		$K=4$		$K=5$	
Multiplex	α	β	α	β	α	β	α	β
$n=25, K=3$	-0.22	0.91	0.69	0.22	-	-	-	-
$n=50, K=3$	0.51	0.68	-0.83	0.12	-	-	-	-
$n=100, K=3$	0.21	0.70	-0.74	1.09	-	-	-	-
$n=50, K=5$	1.10	1.36	0.23	0.45	0.47	0.07	-0.52	0.95

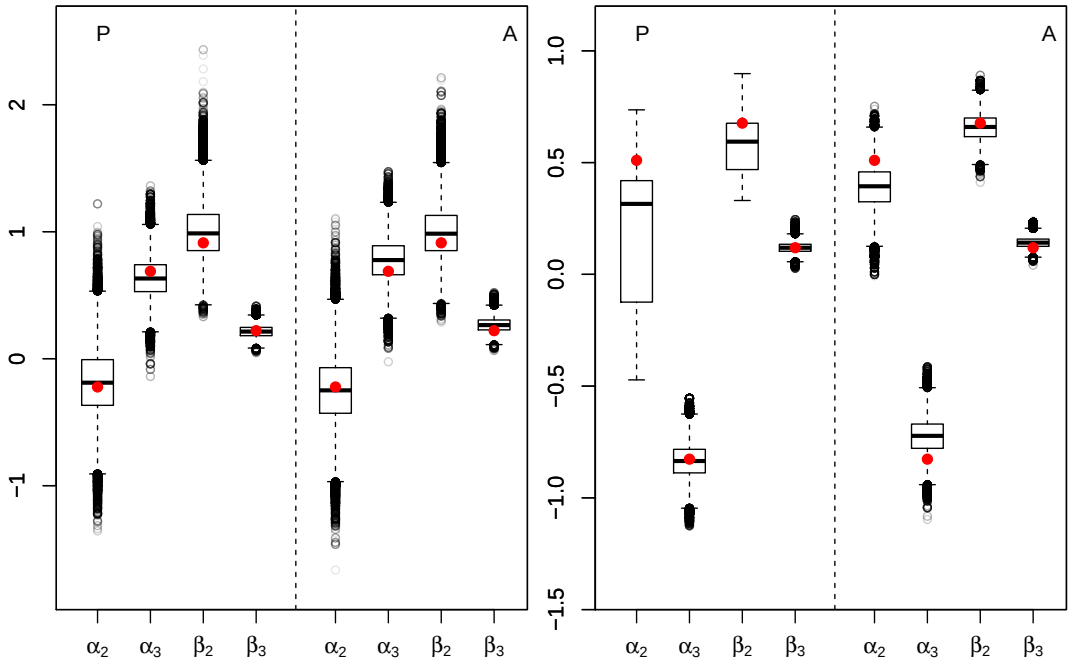
First scenario

Table A.5. Multivariate Gaussian latent coordinates. Averages for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$\alpha^{(1)}$	$\hat{\alpha}^{(2)}$	$\hat{\alpha}^{(3)}$	$\hat{\alpha}^{(4)}$	$\hat{\alpha}^{(5)}$	$\beta^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\beta}^{(4)}$	$\hat{\beta}^{(5)}$	PC	
$n=25,K=3$	0	-0.19	0.63	-	-	1	1.00	0.22	-	-	0.92	P
	0	-0.25	0.78	-	-	1	1.00	0.27	-	-	0.91	A
$n=50,K=3$	0	0.18	-0.84	-	-	1	0.58	0.12	-	-	0.91	P
	0	0.39	-0.72	-	-	1	0.65	0.14	-	-	0.96	A
$n=100,K=3$	0	0.09	-0.97	-	-	1	0.90	1.32	-	-	0.95	P
	0	0.16	-0.95	-	-	1	0.60	0.85	-	-	0.96	A
$n=50,K=5$	0	0.83	0.16	0.60	-0.71	1	1.35	0.48	0.09	1.05	0.97	P
	0	0.85	0.15	0.64	-0.73	1	1.39	0.49	0.09	1.01	0.96	A

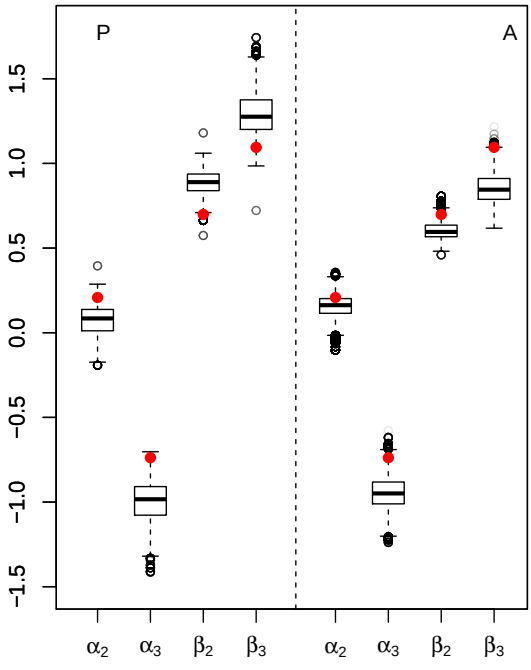
Table A.6. Multivariate Gaussian latent coordinates. Standard deviations for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$sd(\hat{\alpha}^{(2)})$	$sd(\hat{\alpha}^{(3)})$	$sd(\hat{\alpha}^{(4)})$	$sd(\hat{\alpha}^{(5)})$	$sd(\hat{\beta}^{(2)})$	$sd(\hat{\beta}^{(3)})$	$sd(\hat{\beta}^{(4)})$	$sd(\hat{\beta}^{(5)})$	sd(PC)	
$n=25,\ K=3$	0.06	0.04	-	-	0.05	0.03	-	-	0.02	P
	0.02	0.01	-		0.03	0.02	-	-	0.05	A
$n=50,\ K=3$	0.28	0.04	-	-	0.11	0.01	-	-	0.06	P
	0.05	0.02	-	-	0.04	0.01	-	-	0.01	A
$n=100,\ K=3$	0.07	0.09	-	-	0.06	0.10	-	-	0.02	P
	0.05	0.06	-		0.05	0.08	-	-	0.02	A
$n=50,\ K=5$	0.08	0.03	0.03	0.04	0.16	0.08	0.01	0.11	0.01	P
	0.11	0.04	0.02	0.09	0.16	0.04	0.01	0.13	0.01	A

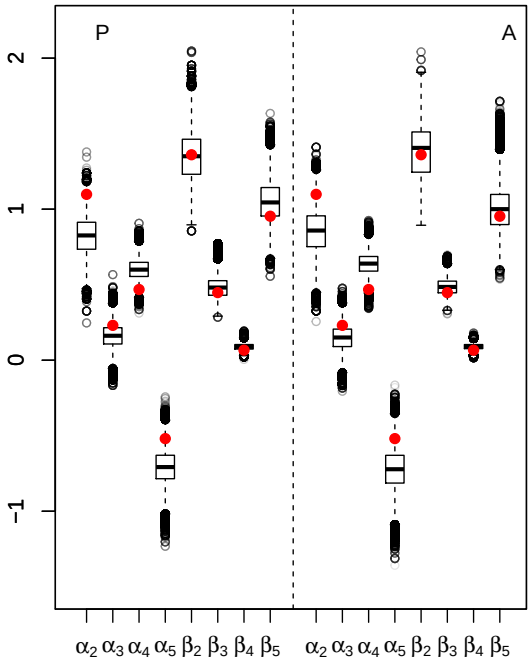


(a) $n = 25, K = 3$

(b) $n = 50, K = 3$



(c) $n = 100, K = 3$



(d) $n = 50, K = 5$

Figure A.8. Simulations: first scenario.

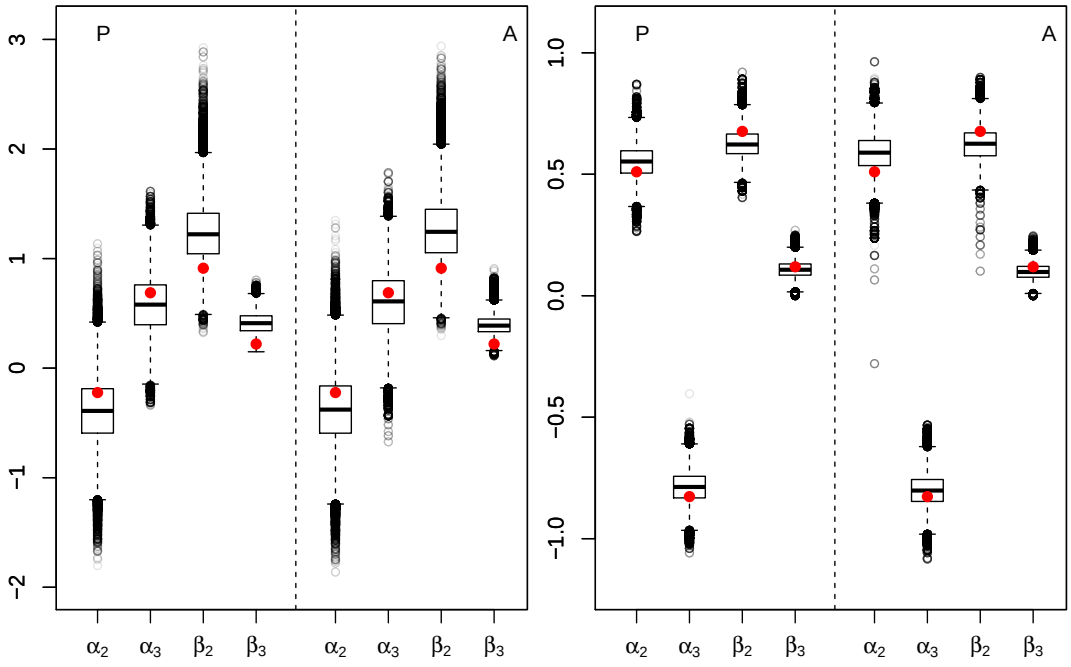
Second scenario

Table A.7. Mixture of multivariate Gaussian distributions latent coordinates. Averages for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$\alpha^{(1)}$	$\hat{\alpha}^{(2)}$	$\hat{\alpha}^{(3)}$	$\hat{\alpha}^{(4)}$	$\hat{\alpha}^{(5)}$	$\beta^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\beta}^{(4)}$	$\hat{\beta}^{(5)}$	PC	
$n = 25,\ K = 3$	0	−0.39	0.58	-	-	1	1.24	0.41	-	-	0.85	P
	0	−0.38	0.60	-	-	1	1.26	0.40	-	-	0.86	A
$n = 50,\ K = 3$	0	0.55	−0.79	-	-	1	0.63	0.11	-	-	0.93	P
	0	0.59	−0.80	-	-	1	0.63	0.10	-	-	0.93	A
$n = 100,\ K = 3$	0	0.10	−0.79	-	-	1	0.54	0.91	-	-	0.95	P
	0	0.12	−0.81	-	-	1	0.53	0.91	-	-	0.95	A
$n = 50,\ K = 5$	0	1.03	0.29	0.56	−0.64	1	1.10	0.39	0.04	0.70	0.95	P
	0	1.03	0.24	0.57	−0.7	1	1.13	0.37	0.04	0.68	0.95	A

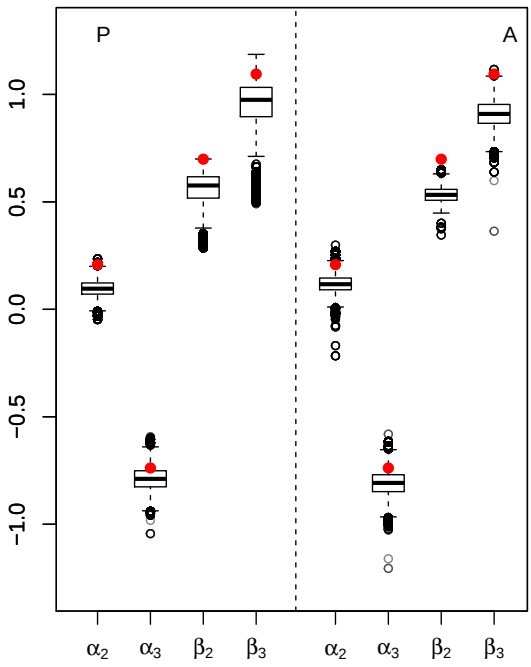
Table A.8. Mixture of multivariate Gaussian distributions latent coordinates. Standard deviations for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$sd(\hat{\alpha}^{(2)})$	$sd(\hat{\alpha}^{(3)})$	$sd(\hat{\alpha}^{(4)})$	$sd(\hat{\alpha}^{(5)})$	$sd(\hat{\beta}^{(2)})$	$sd(\hat{\beta}^{(3)})$	$sd(\hat{\beta}^{(4)})$	$sd(\hat{\beta}^{(5)})$	sd(PC)	
$n = 25,\ K = 3$	0.07	0.15	-	-	0.06	0.07	-	-	0.17	P
	0.08	0.17	-	-	0.08	0.6	-	-	0.18	A
$n = 50,\ K = 3$	0.02	0.01	-	-	0.04	0.01	-	-	0.02	P
	0.06	0.01	-	-	0.05	0.01	-	-	0.02	A
$n = 100,\ K = 3$	0.02	0.03	-	-	0.12	0.18	-	-	0.01	P
	0.02	0.03	-	-	0.03	0.05	-	-	0.01	A
$n = 50,\ K = 5$	0.04	0.02	0.01	0.02	0.09	0.03	0.00	0.06	0.01	P
	0.04	0.02	0.01	0.02	0.08	0.03	0.01	0.04	0.02	A

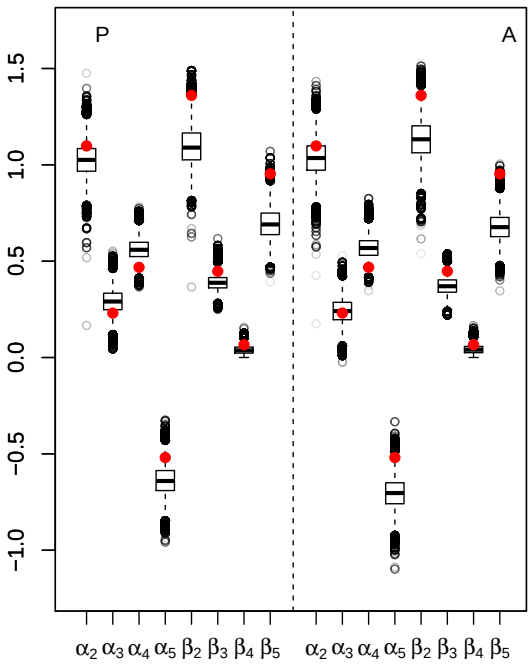


(a) $n = 25, K = 3$

(b) $n = 50, K = 3$



(c) $n = 100, K = 3$



(d) $n = 50, K = 5$

Figure A.9. Simulations: second scenario.

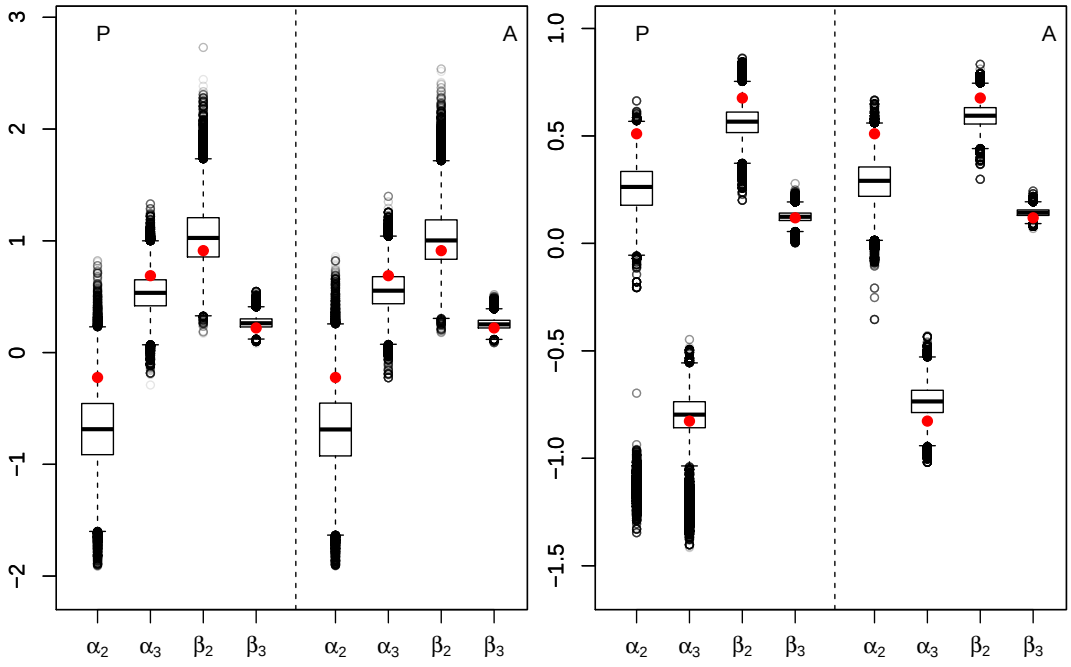
A.4.2 Third scenario

Table A.9. Hotelling T squared latent coordinates. Averages for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$\alpha^{(1)}$	$\hat{\alpha}^{(2)}$	$\hat{\alpha}^{(3)}$	$\hat{\alpha}^{(4)}$	$\hat{\alpha}^{(5)}$	$\beta^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\beta}^{(4)}$	$\hat{\beta}^{(5)}$	PC	
$n = 25,\ K = 3$	0	−0.69	0.54	-	-	1	1.04	0.27	-	-	0.92	P
	0	−0.69	0.56	-	-	1	1.02	0.26	-	-	0.90	A
$n = 50,\ K = 3$	0	0.18	−0.82	-	-	1	0.51	0.12	-	-	0.95	P
	0	0.29	−0.74	-	-	1	0.59	0.14	-	-	0.95	A
$n = 100,\ K = 3$	0	0.02	−0.93	-	-	1	0.93	1.39	-	-	0.80	P
	0	0.04	−0.87	-	-	1	0.90	1.40	-	-	0.80	A
$n = 50,\ K = 5$	0	0.47	0.12	0.60	−0.65	1	1.26	0.60	0.11	1.16	0.97	P
	0	0.53	0.13	0.58	−0.72	1	1.23	0.56	0.10	1.06	0.96	A

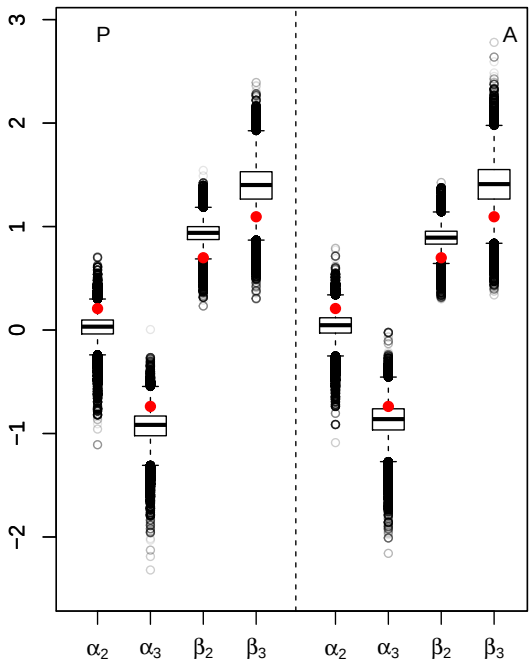
Table A.10. Hotelling T squared latent latent coordinates. Standard deviations for the estimated logistic parameters and the procrustes correlation between true and estimated latent spaces.

	$sd(\hat{\alpha}^{(2)})$	$sd(\hat{\alpha}^{(3)})$	$sd(\hat{\alpha}^{(4)})$	$sd(\hat{\alpha}^{(5)})$	$sd(\hat{\beta}^{(2)})$	$sd(\hat{\beta}^{(3)})$	$sd(\hat{\beta}^{(4)})$	$sd(\hat{\beta}^{(5)})$	sd(PC)	
$n = 25,\ K = 3$	0.04	0.03	-	-	0.02	0.02	-	-	0.02	P
	0.04	0.02	-	-	0.03	0.01	-	-	0.02	A
$n = 50,\ K = 3$	0.31	0.10	-	-	0.06	0.02	-	-	0.02	P
	0.04	0.01	-	-	0.04	0.01	-	-	0.01	A
$n = 100,\ K = 3$	0.01	0.01	-	-	0.02	0.02		-	- 0.02	P
	0.01	0.01	-	-	0.01	0.02	-	-	0.02	A
$n = 50,\ K = 5$	0.10	0.03	0.01	0.07	0.08	0.04	0.01	0.08	0.01	P
	0.08	0.03	0.02	0.05	0.12	0.06	0.01	0.08	0.01	A

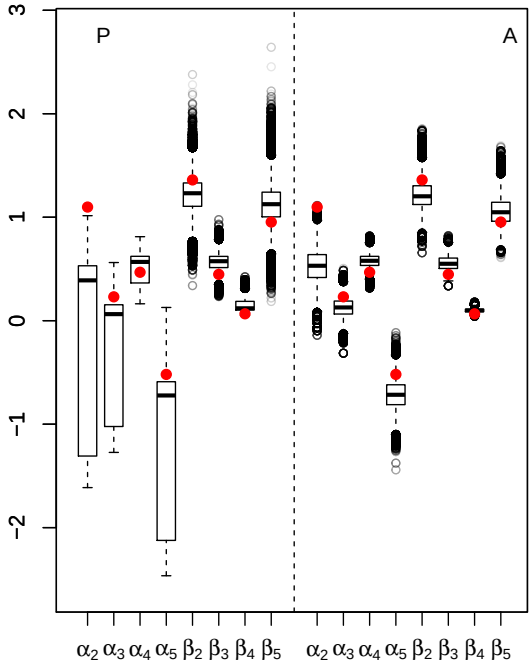


(a) $n = 25, K = 3$

(b) $n = 50, K = 3$



(c) $n = 100, K = 3$



(d) $n = 50, K = 5$

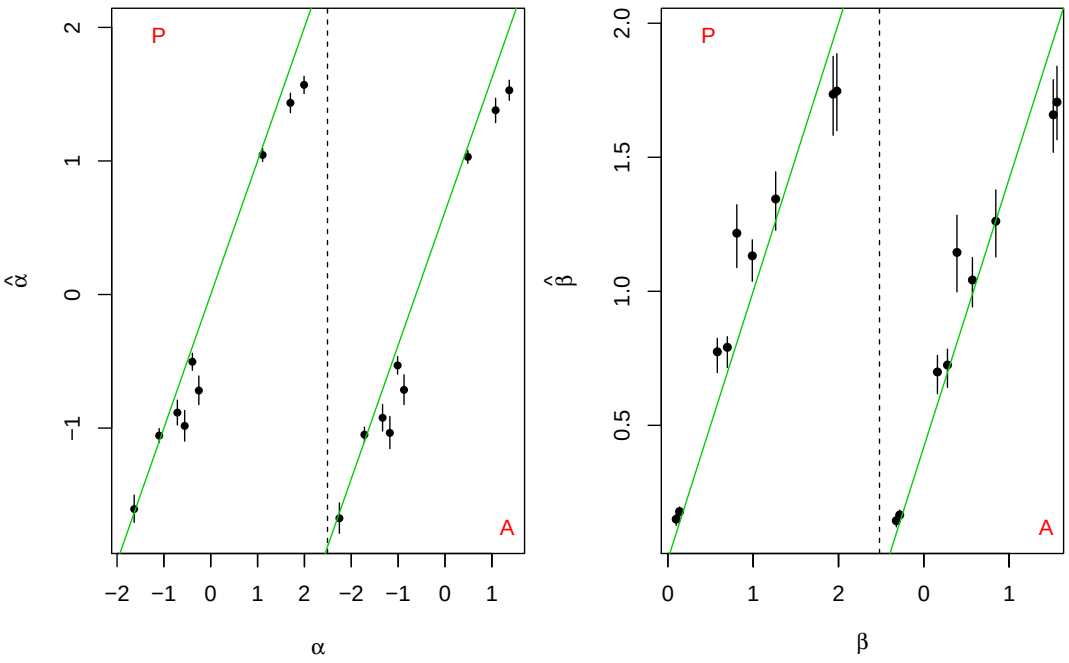
Figure A.10. Simulations: third scenario.

A.4.3 Results for block II

Fourth scenario

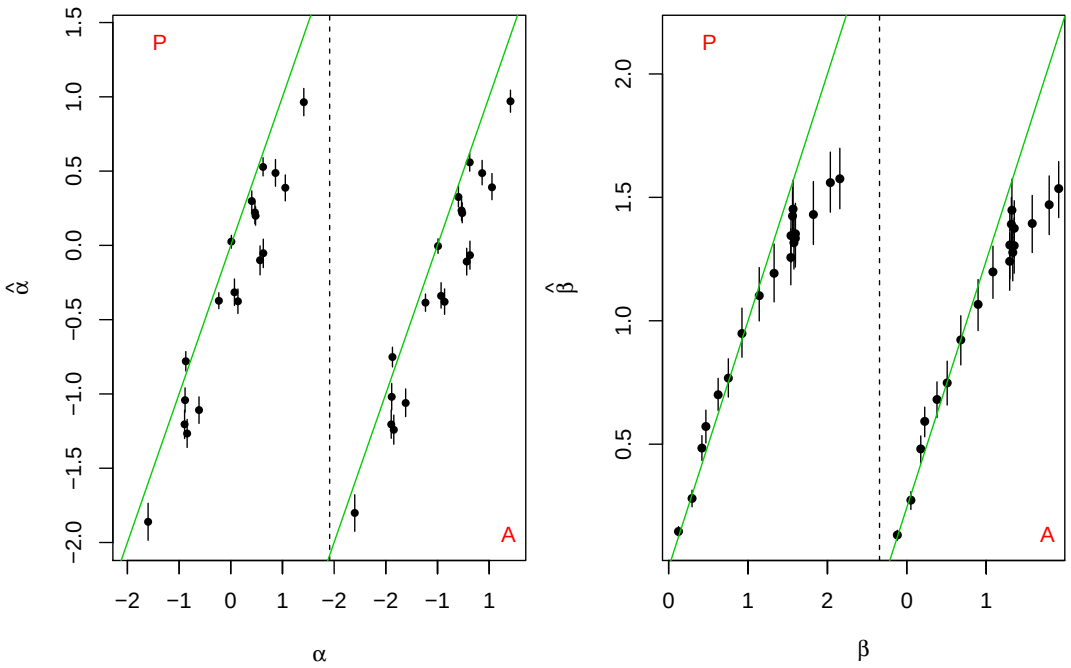
Table A.11. Simulations: fourth scenario. Procrustes correlation.

	$K = 10$		$K = 20$		$K = 30$	
	P	A	P	A	P	A
PC	0.97	0.96	0.97	0.97	0.97	0.97
sd(PC)	0.01	0.01	0.01	0.01	0.01	0.01



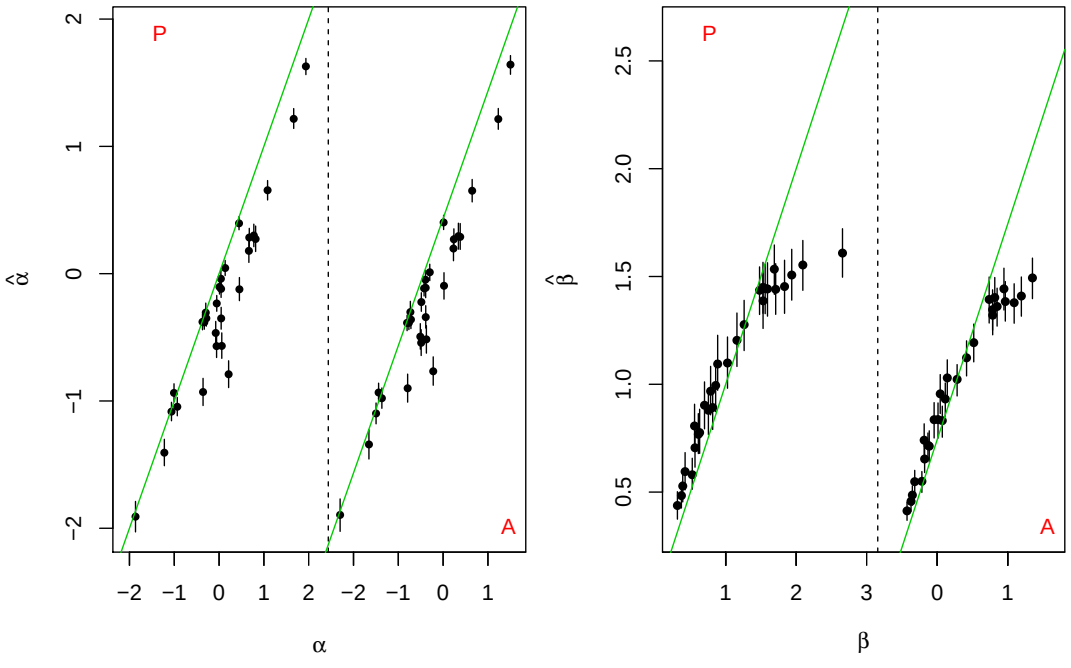
(a) α parameters. **(b)** β parameters.

Figure A.11. Simulations: fourth scenario, multiplex with $n = 50$ and $K = 10$. Estimated averages and standard deviations for the network parameters.



(a) α parameters. **(b)** β parameters.

Figure A.12. Simulations: fourth scenario, multiplex with $n = 50$ and $K = 20$. Estimated averages and standard deviations for the network parameters.



(a) α parameters. **(b)** β parameters.

Figure A.13. Simulations: fourth scenario, multiplex with $n = 50$ and $K = 30$. Estimated averages and standard deviations for the network parameters.

A.4.4 Comparison with the *lsjm* model. Results

Table A.12. Results of the comparison with the *lsjm* model. Averages and standard deviations for the estimated intercepts and the procrustes correlation between true and estimated latent spaces. The acronym *lsmmn* is used to indicate the model presented in this work.

	$\hat{\alpha}^{(1)}$	$\hat{\alpha}^{(2)}$	$\hat{\alpha}^{(3)}$	$sd(\hat{\alpha}^{(1)})$	$sd(\hat{\alpha}^{(2)})$	$sd(\hat{\alpha}^{(3)})$	PC	sd(PC)	
$n = 50$, $K = 3$	−201.89	−211.10	−193.75	1501.55	1500.95	1496.83	0.46	0.27	`lsjm`
	0	−0.73	−0.64	-	0.04	0.07	0.96	0.01	`lsmmn`
$n = 70$, $K = 2$	−0.84	−1.30	-	0.06	0.06	-	0.79	0.05	`lsjm`
	0	−1.36	-		-	0.01	0.85	0.05	`lsmmn`
$n = 25$, $K = 3$	−44.38	15.40	−16.80	360.49	353.41	356.76	0.57	0.31	`lsjm`
	0	0.69	0.26	-	0.08	0.06	0.95	0.05	`lsmmn`

Concerning the *lsjm* estimates, the instability in the estimation of the intercept parameters in the first and third scenario could be caused by similar identifiability issues as the one described in the paper. The second scenario returns stable and truthful estimates of the intercept parameters and high correlation on average between the true and the estimated latent space. The *lsmmn* model always recovers a set of latent coordinates which is highly correlated with the simulated one. The average value for this correlation in the estimates for the second multiplex is lower than the values estimated for multidimensional network one and three. This could be due to the presence of only 2 views in the second multiplex, meaning that there are only two replications of the latent space in the data. The *lsmmn* model estimated also the coefficient parameters $\beta^{(k)}$, which in the simulations have all been set to 1. The estimates produced for all the three multiplex returned faithful values for the coefficients, as in all the cases $.93 \leq \hat{\beta}^{(k)} \leq 1.05$, $\forall k$. The intercept coefficient in the first network as been fixed $\alpha^{(1)} = 0$ to estimate all the three multiplex.

A.5 Pseudo-code of the mcmc algorithm

Algorithm 1 MCmc for LSM-MN

procedure

Given: $\mathbf{Y}, K, n, LB(\alpha), \mathbf{H}$

Fix: η

Initialize: $\mathbf{z}^{\{0\}}, \mathbf{D}^{\{0\}}, \alpha^{\{0\}}, \beta^{\{0\}}, \mu_\alpha^{\{0\}}, \mu_\beta^{\{0\}}, \sigma_\alpha^{2\{0\}}, \sigma_\beta^{2\{0\}}$

Fix: $\alpha^{(1)\{0\}} = 0, \quad \beta^{(1)\{0\}} = 1$

for iter in 1:niter **do**

Simulate $\sigma_\alpha^{2\{iter\}}$ from $g\Big(\sigma_\alpha^2 \mid \alpha^{\{iter-1\}}, \mu_\alpha^{\{iter-1\}}, \tau_\alpha, \nu_\alpha, K\Big)$;

Simulate $\sigma_\beta^{2\{iter\}}$ from $g\Big(\sigma_\beta^2 \mid \beta^{\{iter-1\}}, \mu_\beta^{\{iter-1\}}, \tau_\beta, \nu_\beta, K\Big)$;

Simulate $\mu_\alpha^{\{iter\}}$ from $g\Big(\mu_\alpha \mid \alpha^{\{iter-1\}}, \sigma_\alpha^{2\{iter\}}, \tau_\alpha, K\Big)$;

Simulate $\mu_\beta^{\{iter\}}$ from $g\Big(\mu_\beta \mid \beta^{\{iter-1\}}, \sigma_\beta^{2\{iter\}}, \tau_\beta, K\Big)$;

for k in 2:K **do**

Propose $\tilde{\alpha}^{(k)}$ from $g\Big(\alpha^{(k)} \mid \mathbf{Y}, \mathbf{D}^{\{iter-1\}}, \mathbf{H}, \beta^{(k)\{iter-1\}}, \mu_\alpha^{\{iter\}}, \sigma_\alpha^{2\{iter\}}, n\Big)$;

Propose $\tilde{\beta}^{(k)}$ from $g\Big(\beta^{(k)} \mid \mathbf{Y}, \mathbf{D}^{\{iter-1\}}, \mathbf{H}, \alpha^{(k)\{iter-1\}}, \mu_\beta^{\{iter\}}, \sigma_\beta^{2\{iter\}}, n\Big)$;

Accept $\Big(\tilde{\alpha}^{(k)}, \tilde{\beta}^{(k)}\Big)$ with probability $A1$

if Accept **then**

$\alpha^{(k)\{iter\}}, \beta^{(k)\{iter\}} \leftarrow \tilde{\alpha}^{(k)}, \tilde{\beta}^{(k)}$

else

$\alpha^{(k)\{iter\}}, \beta^{(k)\{iter\}} \leftarrow \alpha^{(k)\{iter-1\}}, \beta^{(k)\{iter-1\}}$

Assign $\mathbf{z}^{\{old\}} \leftarrow \mathbf{z}^{\{iter-1\}}$

for i in 1:n **do**

Propose $\tilde{z}_i$ from $g\Big(z_i \mid \mathbf{Y}, \alpha^{\{iter\}}, \beta^{\{iter\}}, \mathbf{D}^{\{iter-1\}}, K\Big)$;

Accept $\tilde{z}_i$ with probability $A2$

if Accept **then**

$z_i^{\{iter\}} \leftarrow \tilde{z}_i$

else

$z_i^{\{iter\}} \leftarrow z_i^{\{iter-1\}}$

if Procrustes Check$\big(\mathbf{z}^{\{old\}}, \mathbf{z}^{\{iter\}}\big) == 0$ **then**

$\mathbf{z}^{\{iter\}} \leftarrow \mathbf{z}^{\{iter\}}$

else

$\mathbf{z}^{\{iter\}} \leftarrow \mathbf{z}^{\{iter-1\}}$

If edge-specific covariates are considered, the parameters λ_f are updated after the latent positions, with a Metropolis-Hastings step, using their proposal distributions. The nuisance parameters μ_{λ_f} and $\sigma^2_{\lambda_f}$ are updated via their full conditional distribution, right after the update of the nuisance parameters for the prior distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In addition, the proposal distributions to use for the coordinates, the coefficients and the intercepts are the ones described in appendix A.1.

Appendix B

Appendix to Chapter 3

B.1 FAO data

Below, we report the country ISO3 codes for the FAO dataset

Country name	iso3 code	Country name	iso3 code
Australia	AUS	Malaysia	MYS
Austria	AUT	Maldives	MDV
Bahrain	BHR	Montenegro	MNE
Belarus	BLR	Morocco	MAR
Belgium	BEL	Netherlands	NLD
Bosnia and Herzegovina	BIH	New Zealand	NZL
Bulgaria	BGR	Norway	NOR
Canada	CAN	Oman	OMN
Hong Kong	HKG	Pakistan	PAK
China	CHN	Poland	POL
Croatia	HRV	Portugal	PRT
Czech Republic	CZE	Qatar	QAT
Denmark	DNK	Republic of Korea	KOR
Egypt	EGY	Republic of Moldova	MDA
Estonia	EST	Romania	ROU
Finland	FIN	Russian Federation	RUS
France	FRA	Saudi Arabia	SAU
Germany	DEU	Serbia	SRB
Greece	GRC	Singapore	SGP
Hungary	HUN	Slovakia	SVK
India	IND	Slovenia	SVN
Iran	IRN	South Africa	ZAF
Ireland	IRL	Spain	ESP
Italy	ITA	Sweden	SWE
Jordan	JOR	Switzerland	CHE
Kazakhstan	KAZ	Thailand	THA
Kenya	KEN	Republic of Macedonia	MKD
Kuwait	KWT	Turkey	TUR
Latvia	LVA	Ukraine	UKR
Lebanon	LBN	United Arab Emirates	ARE
Lithuania	LTU	United Kingdom	GBR
Luxembourg	LUX	United States of America	USA

Table B.1. Fao data: country ISO3 codes.

B.2 Estimation: proposal and full conditional distributions

The log-posterior distribution for the eight models with sender and/or receiver effect, presented in section 3.2, is proportional to:

$$
\begin{aligned}
&\log\Big(P(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{z},\mu_\alpha,\mu_\beta,\sigma^2_\alpha,\sigma^2_\beta|\mathbf{Y})\Big) \propto \\
&\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i} h_{ij}^{(k)}\Big[y_{ij}^{(k)}(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}) - \log\Big(1+\exp\{\alpha^{(k)}\phi_{ij}^{(k)} \\
&- \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}\Big)\Big] - \frac{1}{2}\Bigg\{\sum_{i=1}^{n} z_i^2 + \frac{\sum_{k=1}^{K}(\alpha^{(k)}-\mu_\alpha)^2}{\sigma^2_\alpha} + \frac{\sum_{k=1}^{K}(\beta^{(k)}-\mu_\beta)^2}{\sigma^2_\beta} \\
&+ K\log(\sigma^2_\alpha) + K\log(\sigma^2_\beta) + \log(\tau_\alpha\sigma^2_\alpha) + \log(\tau_\beta\sigma^2_\beta) + \frac{\mu^2_\alpha}{\tau_\alpha\sigma^2_\alpha} + \frac{\mu^2_\beta}{\tau_\beta\sigma^2_\beta} + \frac{1}{\sigma^2_\alpha} + \frac{1}{\sigma^2_\beta} \\
&+ \sum_{f=1}^{F}\left(\frac{(\lambda_f-\mu_{\lambda_f})^2}{\sigma^2_{\lambda_f}} + \log(\sigma^2_{\lambda_f}) + \log(\tau_\lambda\sigma^2_{\lambda_f}) + \frac{(\mu_{\lambda_f}-\mu_\lambda)^2}{\tau_\lambda\sigma^2_{\lambda_f}} + \frac{1}{\sigma^2_{\lambda_f}}\right)\Bigg\} \\
&+ \Big(-\frac{\nu_\alpha}{2}-1\Big)\log(\sigma^2_\alpha) + \Big(-\frac{\nu_\beta}{2}-1\Big)\log(\sigma^2_\beta) + \sum_{f=1}^{F}\Big(-\frac{\nu_{\lambda_f}}{2}-1\Big)\log(\sigma^2_{\lambda_f}),
\end{aligned}
\tag{B.1}
$$

where, without any loss of information, the latent coordinates are assumed to be univariate.

B.2.1 Nuisance parameters

The variances of the intercept and coefficient parameters have Inverse Gamma full conditional distributions

$$
\sigma^2_\alpha|\boldsymbol{\alpha},\mu_\alpha,\tau_\alpha,\nu_\alpha,K \sim \text{Inv}\Gamma(r_\alpha,R_\alpha); \qquad \sigma^2_\beta|\boldsymbol{\beta},\mu_\beta,\tau_\beta,\nu_\beta,K \sim \text{Inv}\Gamma(r_\beta,R_\beta),
$$

with parameters:

$$
r_x = \frac{\nu_x + K + 1}{2}, \qquad R_x = \frac{\tau_x + \tau_x\sum_{k=1}^{K}(x^{(k)}-\mu_x)^2 + \mu_x^2}{2\tau_x}.
$$

The nuisance parameters μ_α, μ_β are distributed as truncated normal distributions:

$$
\mu_\alpha|\boldsymbol{\alpha},\sigma^2_\alpha,\tau_\alpha,m_\alpha,K \sim N_{[0,\infty]}\left(\frac{\tau_\alpha\sum_{k=1}^{K}\alpha^{(k)} + m_\alpha}{1+K\tau_\alpha}, \frac{\tau_\alpha\sigma^2_\alpha}{1+K\tau_\alpha}\right),
$$

$$
\mu_\beta|\boldsymbol{\beta},\sigma^2_\beta,\tau_\beta,m_\beta,K \sim N_{[0,\infty]}\left(\frac{\tau_\beta\sum_{k=1}^{K}\beta^{(k)} + m_\beta}{1+K\tau_\beta}, \frac{\tau_\beta\sigma^2_\beta}{1+K\tau_\beta}\right).
$$

B.2.2 Latent positions

The proposal distribution for the i^{th} latent coordinate is derived from the log-posterior distribution of the model in equation B.1. The logarithmic term in the log-likelihood, which is indeed a LSE function, is approximated with its lower bound.

$$\tilde{\boldsymbol{z}}_i \mid \mathbf{Y}, \boldsymbol{\alpha}, \theta_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{x}_i, \mathbf{D}, K \sim N\Big(\mu_{\tilde{z}_i}, \sigma^2_{\tilde{z}_i}\mathbf{I}\Big),$$

where

$$\mu_{\tilde{z}_i} = \sigma^2_{\tilde{z}_i}\Big(2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}h_{ij}^{(k)}(y_{ij}^{(k)}-w_{ij}^{(k)})z_j\Big), \quad \sigma^2_{\tilde{z}_i} = \left(1+2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}h_{ij}^{(k)}|y_{ij}^{(k)}-w_{ij}^{(k)}|\right)^{-1}.$$

where $w_{ij}^{(k)}$ is a binary indicator variable, defined as:

$$w_{ij}^{(k)} = \begin{cases} 1 & \text{if} \quad \alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf} > 0 \\ 0 & \text{if} \quad \alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf} \leq 0 \end{cases},$$

B.2.3 Intercept parameters

The proposal distribution for the k^{th} intercept parameter $\alpha^{(k)}$ is derived from the log-posterior distribution, where the logarithmic term is approximated via its second order Taylor expansion in $\alpha^{(k)} = \mu_\alpha$. Defining $E^{(k)} = \sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}y_{ij}^{(k)}\phi_{ij}^{(k)}$, the proposal distribution for intercept $\alpha^{(k)}$ is taken to be:

$$\tilde{\alpha}^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \beta^{(k)}, \theta^{(k)}, \gamma^{(k)}, \lambda, \mathbf{X}, \mu_\alpha, \sigma^2_\alpha \sim N\Big(\mu_{\tilde{\alpha}^{(k)}}, \sigma^2_{\tilde{\alpha}^{(k)}}\Big),$$

with

$$\mu_{\tilde{\alpha}^{(k)}} = \sigma^2_{\tilde{\alpha}^{(k)}}\left\{E^{(k)} - \sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}\phi_{ij}^{(k)}\exp\{\mu_\alpha\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}}{1+\exp\{\mu_\alpha\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}}\right\} + \mu_\alpha,$$

$$\sigma^2_{\tilde{\alpha}^{(k)}} = \left\{\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}(\phi_{ij}^{(k)})^2\exp\{\mu_\alpha\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}}{\left(1+\exp\{\mu_\alpha\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}\right)^2} + \frac{1}{\sigma^2_\alpha}\right\}^{-1}$$

B.2.4 Coefficient parameters (distances)

The proposal distribution for the k^{th} coefficient parameter $\beta^{(k)}$ is derived from the log-posterior distribution, where the logarithmic term is approximated via its second order Taylor expansion in $\beta^{(k)} = \mu_\beta$.
Then, the proposal distribution specified for intercept coefficient k is:

$$\tilde{\beta}^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \alpha^{(k)}, \theta^{(k)}, \gamma^{(k)}, \lambda, \mathbf{X}, \mu_\beta, \sigma^2_\beta \sim N\Big(\mu_{\tilde{\beta}^{(k)}}, \sigma^2_{\tilde{\beta}^{(k)}}, n\Big),$$

where

$$\mu_{\tilde{\beta}^{(k)}} = \sigma^2_{\tilde{\beta}^{(k)}}\left\{\sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}d_{ij}\left(\frac{\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \mu_\beta d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}}{1+\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \mu_\beta d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}} - y_{ij}^{(k)}\right)\right\} + \mu_\beta,$$

$$\sigma^2_{\tilde{\beta}^{(k)}} = \left\{\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}d_{ij}^2\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \mu_\beta d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}}{\left(1+\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \mu_\beta d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}\right)^2} + \frac{1}{\sigma^2_\beta}\right\}^{-1}$$

B.2.5 Coefficient parameters (covariates)

The proposal distribution for λ_f is:

$$\tilde{\lambda}_f \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \mu_{\lambda_f}, \sigma^2_{\lambda_f}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(0,\infty)}(\mu_{\tilde{\lambda}_f}, \sigma^2_{\tilde{\lambda}_f}),$$

where

$$\mu_{\tilde{\lambda}_f} = \sigma^2_{\tilde{\lambda}_f}\left[\sum_{k=1}^{K}\left(\sum_{i=1}^{n}\sum_{j\neq i}\frac{x_{ijf}h_{ij}^{(k)}\exp(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{l\neq f}\lambda_l x_{ijl})}{1+\exp(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{l\neq f}\lambda_l x_{ijl})} - y_{ij}^{(k)}x_{ijf}\right)\right] + \mu_{\lambda_f},$$

$$\sigma^2_{\tilde{\lambda}_f} = \left\{\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}\frac{h_{ij}^{(k)}x_{ijf}^2\exp(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{l\neq f}\lambda_l x_{ijl})}{\left(1+\exp(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{l\neq f}\lambda_l x_{ijl})\right)^2} + \frac{1}{\sigma^2_{\lambda_f}}\right\}^{-1}.$$

B.2.6 Sender and receiver parameters

To define proposal distributions for the sender and receiver parameter, we consider only the relevant part of the posterior distribution. In particular, we will derive them from the log-likelihood of the model, that is, the first line of the log-posterior distribution presented in equation B.1,

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i}h_{ij}^{(k)}\Big[y_{ij}^{(k)}(\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}) - \log\Big(1+\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij} - \sum_{f=1}^{F}\lambda_f x_{ijf}\}\Big)\Big].$$

Sender parameters

When the sender effect is variable across the networks, we need to propose a new value for the $i^{(th)}$ sender parameter in the k^{th} network at the t^{th} iteration of the MCMC algorithm. We start by approximating the logarithmic term in the log-likelihood via its second order Taylor's expansion in $\theta_i^{(k)t} \approx \theta_i^{(k)t-1}$:

$$\begin{aligned}
\log\Big(1+\exp\{\alpha^{(k)}\phi_{ij}^{(k)t} - \beta^{(k)}d_{ij}\}\Big) &= \log\Big(1+\exp\{\frac{\alpha^{(k)}\theta_i^{(k)t}}{g} + \frac{\alpha^{(k)}\gamma_j^{(k)}}{g} - \beta^{(k)}d_{ij}\}\Big) \\
&\approx \left(\theta_i^{(k)t} - \theta_i^{(k)t-1}\right)\left\{\frac{\frac{\alpha^{(k)}}{g}\exp\{\frac{\alpha^{(k)}\theta_i^{(k)t-1}}{g} + \frac{\alpha^{(k)}\gamma_j^{(k)}}{g} - \beta^{(k)}d_{ij}\}}{1+\exp\{\frac{\alpha^{(k)}\theta_i^{(k)t-1}}{g} + \frac{\alpha^{(k)}\gamma_j^{(k)}}{g} - \beta^{(k)}d_{ij}\}}\right\} \\
&+ \frac{1}{2}\left(\theta_i^{(k)t} - \theta_i^{(k)t-1}\right)^2\left\{\frac{(\frac{\alpha^{(k)}}{g})^2\exp\{\frac{\alpha^{(k)}\theta_i^{(k)t-1}}{g} + \frac{\alpha^{(k)}\gamma_j^{(k)}}{g} - \beta^{(k)}d_{ij}\}}{(1+\exp\{\frac{\alpha^{(k)}\theta_i^{(k)t-1}}{g} + \frac{\alpha^{(k)}\gamma_j^{(k)}}{g} - \beta^{(k)}d_{ij}\})^2}\right\} \\
&+ const
\end{aligned}$$

Then, the logarithmic term is substituted with its approximation. The approximated log-likelihood is now a quadratic function in $\theta_i^{(k)t}$, which permits to define the following proposal distribution

$$\tilde{\theta}_i^{(k)t} \mid \alpha^{(k)}, \beta^{(k)}, \boldsymbol{\lambda}, \gamma^{(k)}, \theta_i^{(k)t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\theta}_i^{(k)t}}, \sigma^2_{\tilde{\theta}_i^{(k)t}}\right)$$

where

$$\mu_{\tilde{\theta}_i^{(k)t}} = \left[\frac{\alpha^{(k)}}{g}\sum_{j\neq i} h_{ij}^{(k)}\left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\theta}_i^{(k)t}} + \theta_i^{(k)t-1}$$

$$\sigma^2_{\tilde{\theta}_i^{(k)t}} = \left[\left(\frac{\alpha^{(k)}}{g}\right)^2\sum_{j\neq i} h_{ij}^{(k)}\left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}$$

Instead, when the sender effect is constant, the logarithmic term in the log-likelihood is approximated with its second order Taylor's expansion in $\theta_i^t \approx \theta_i^{t-1}$a and the resulting proposal distribution is:

$$\tilde{\theta}_i^t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \theta_i^{t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\theta}_i^t}, \sigma^2_{\tilde{\theta}_i^t}\right)$$

where

$$\mu_{\tilde{\theta}_i^t} = \left[\sum_{k=1}^{K}\frac{\alpha^{(k)}}{g}\sum_{j\neq i} h_{ij}^{(k)}\left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\theta}_i^t} + \theta_i^{t-1}$$

$$\sigma^2_{\tilde{\theta}_i^t} = \left[\sum_{k=1}^{K}\left(\frac{\alpha^{(k)}}{g}\right)^2\sum_{j\neq i} h_{ij}^{(k)}\left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}$$

The proposal distribution defined depend on the previous value of the parameter of interest, as in random walk proposals, but also on the current configuration of all the other parameters. This allows to explore efficiently the parameter space for the $\theta_i^{(k)}$s: Indeed, classical random walks are usually slow in the exploration of the parameter space and, when employed in the estimation procedure of this class of models, they returned unstable chains. "Correcting" the random walk by incorporating the past value of the sender parameter of interest in a more complex parametrization of the proposal distribution has proven to be a valid approach to estimate these parameters.

Receiver parameters

The proposal distributions for the receiver parameters, both in the variable and in the constant case, are recovered with the same procedure used to define those of the sender parameters. The only, obvious, difference is that the logarithmic term of the log-likelihood is approximated in the receiver parameter. Then, the proposal distributions for the j^{th} receiver parameter at the t^{th} iteration are:

- If the effect is variable:

$$\tilde{\gamma}_j^{(k)t} \mid \alpha^{(k)}, \beta^{(k)}, \boldsymbol{\lambda}, \theta^{(k)}, \gamma_j^{(k)t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\gamma}_j^{(k)t}}, \sigma^2_{\tilde{\gamma}_j^{(k)t}}\right)$$

  where

$$\mu_{\tilde{\gamma}_j^{(k)t}} = \left[\frac{\alpha^{(k)}}{g}\sum_{i\neq j} h_{ij}^{(k)}\left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\gamma}_j^{(k)t}} + \gamma_j^{(k)t-1}$$

$$\sigma^2_{\tilde{\gamma}_j^{(k)t}} = \left[\left(\frac{\alpha^{(k)}}{g}\right)^2 \sum_{i\neq j} h_{ij}^{(k)} \left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}.$$

- If the effect is constant:

$$\tilde{\gamma}_j^t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \gamma_j^{t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\gamma}_j^t}, \sigma^2_{\tilde{\gamma}_j^t}\right)$$

where

$$\mu_{\tilde{\gamma}_j^t} = \left[\sum_{k=1}^{K} \frac{\alpha^{(k)}}{g} \sum_{i\neq j} h_{ij}^{(k)} \left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\gamma}_j^t} + \gamma_j^{t-1}$$

$$\sigma^2_{\tilde{\gamma}_j^t} = \left[\sum_{k=1}^{K}\left(\frac{\alpha^{(k)}}{g}\right)^2 \sum_{i\neq j} h_{ij}^{(k)} \left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}.$$

Undirected networks

The proposal distributions for the $\delta_i^{(k)}$ parameters, both in the variable and in the constant case, are recovered with the same procedure used to define those of the sender and the receiver parameters. This time, the logarithmic term in the log-likelihood, which this time is defined via the edge probabilities in equation 3.7, is approximated in the i^{th} parameter. The proposal distributions for the i^{th} parameter at the t^{th} iteration are:

- If the effect is variable:

$$\tilde{\delta}_i^{(k)t} \mid \alpha^{(k)}, \beta^{(k)}, \boldsymbol{\lambda}, \delta^{(k)t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\delta}_i^{(k)t}}, \sigma^2_{\tilde{\delta}_i^{(k)t}}\right)$$

where

$$\mu_{\tilde{\delta}_i^{(k)t}} = \left[\frac{\alpha^{(k)}}{g} \sum_{j\neq i} h_{ij}^{(k)} \left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\delta}_i^{(k)t}} + \delta_i^{(k)t-1}$$

$$\sigma^2_{\tilde{\delta}_i^{(k)t}} = \left[\left(\frac{\alpha^{(k)}}{g}\right)^2 \sum_{j\neq i} h_{ij}^{(k)} \left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{(k)t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}.$$

- If the effect is constant:

$$\tilde{\delta}_i^t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}^{t-1}, \mathbf{D}, \mathbf{H}, \mathbf{X}, \mathbf{Y} \sim N_{(-1,1)}\left(\mu_{\tilde{\delta}_i^t}, \sigma^2_{\tilde{\delta}_i^t}\right)$$

where

$$\mu_{\tilde{\delta}_i^t} = \left[\sum_{k=1}^{K} \frac{\alpha^{(k)}}{g} \sum_{j\neq i} h_{ij}^{(k)} \left(y_{ij}^{(k)} - \frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{1+\exp\left(\alpha^{(k)}\phi_{ij}^{(t-1} - \beta^{(k)}d_{ij}\right)}\right)\right]\sigma^2_{\tilde{\delta}_i^t} + \delta_i^{t-1}$$

$$\sigma^2_{\tilde{\delta}_i^t} = \left[\sum_{k=1}^{K}\left(\frac{\alpha^{(k)}}{g}\right)^2 \sum_{j\neq i} h_{ij}^{(k)} \left(\frac{\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)}{\left(1+\exp\left(\alpha^{(k)}\phi_{ij}^{t-1} - \beta^{(k)}d_{ij}\right)\right)^2}\right)\right]^{-1}.$$

B.3 Scenario I: simulation results

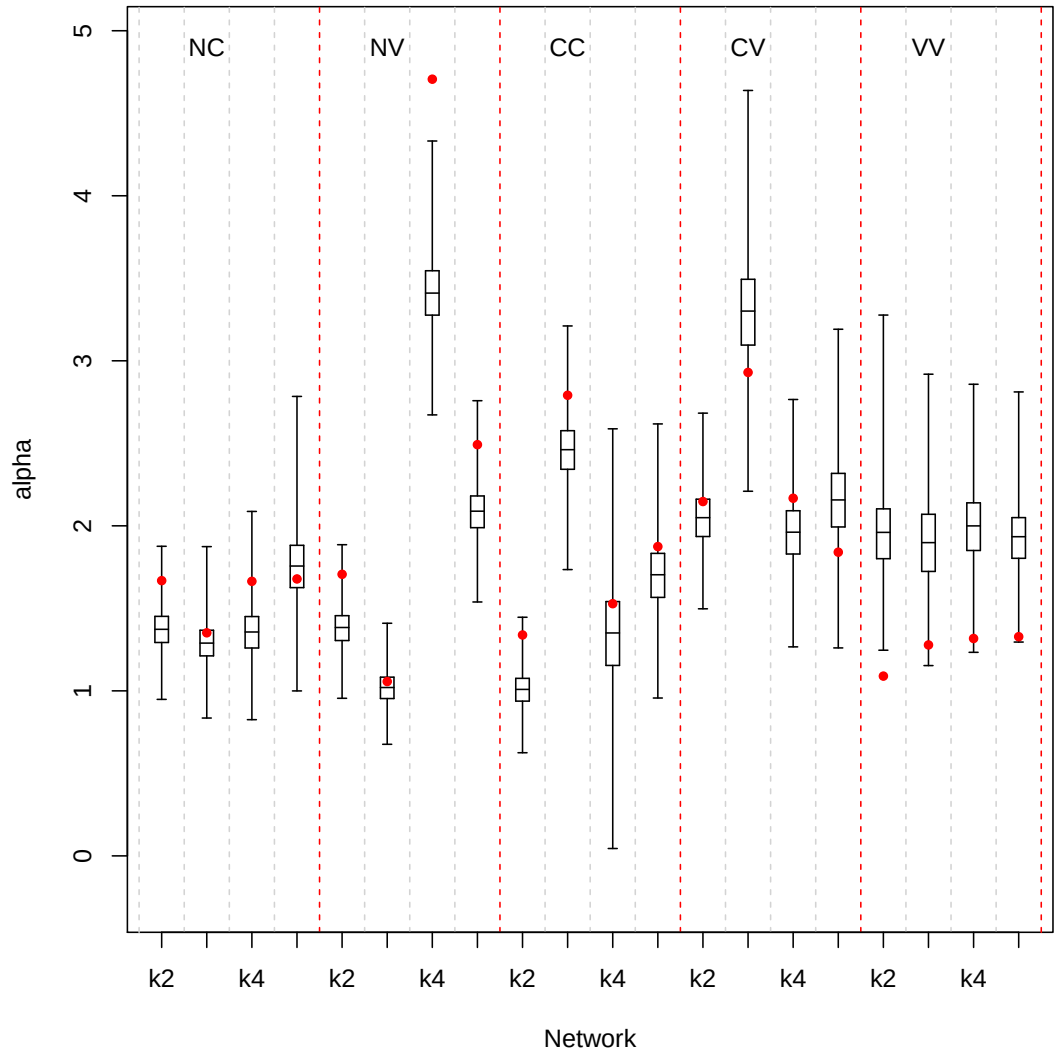


Figure B.1. Boxplots of the estimated posterior distributions for the intercept parameters $\alpha^{(k)}$. Red dots indicate the true, simulated, values of the intercepts.

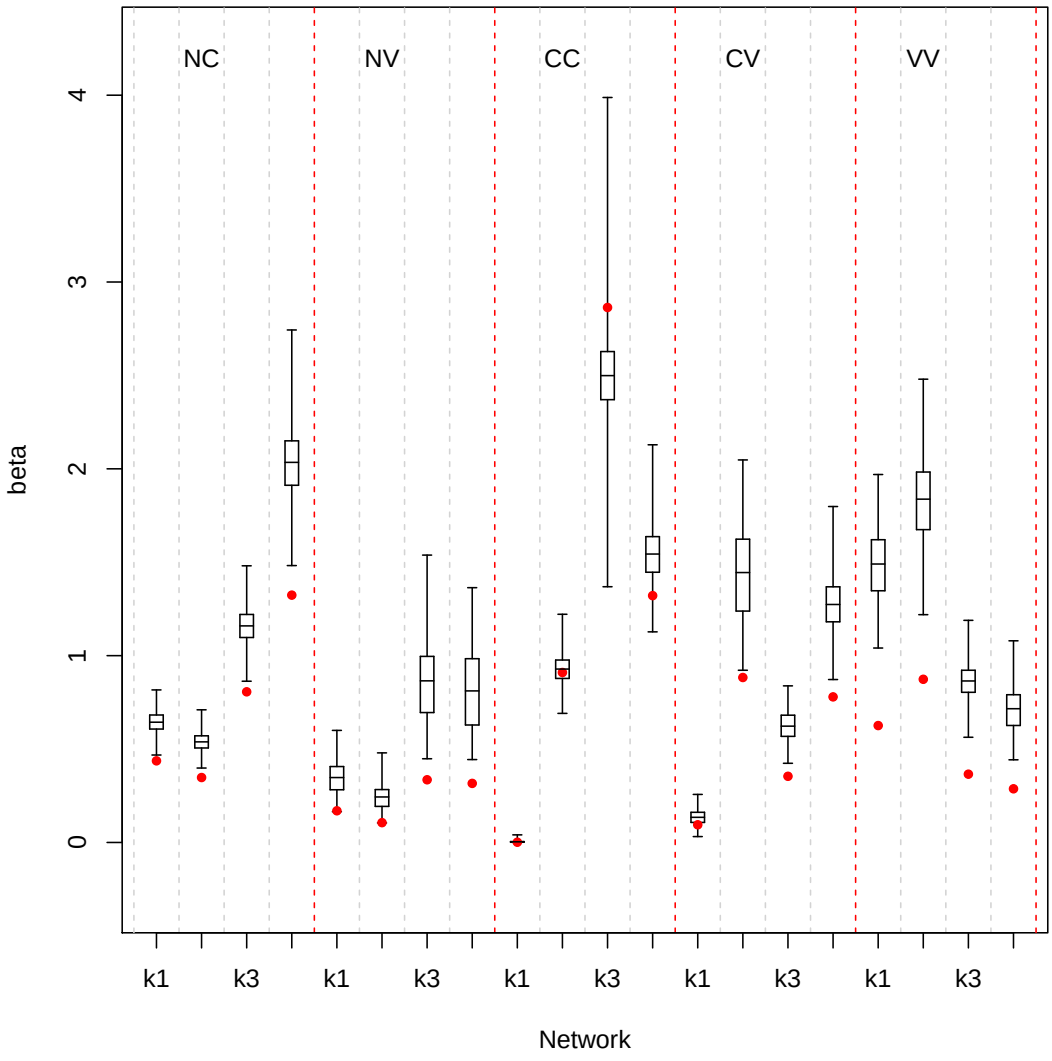


Figure B.2. Boxplots of the estimated posterior distributions for the coefficient parameters $\beta^{(k)}$. Red dots indicate the true, simulated, values of the coefficients.

B.4 Heuristic model search

		Simulated								
		NN	CN	NC	CC	VN	NV	VC	CV	VV
Selected (%)	NN	**84**	0	0	15	0	0	0	0	1
	CN	1	**94**	0	4	0	0	0	1	0
	NC	1	0	**95**	3	0	0	1	0	0
	CC	17	0	0	**80**	0	0	1	1	1
	VN	0	0	0	0	**84**	0	16	0	0
	NV	0	0	0	0	0	**84**	0	16	0
	VC	1	0	0	1	5	0	**92**	0	1
	CV	1	0	0	1	0	5	0	**92**	1
	VV	2	0	0	5	0	0	0	0	**93**

Table B.2. Block II, $n = 50$, $K = 10$. Percentage of time each model is selected by the Heuristic search.

		Simulated								
		NN	CN	NC	CC	VN	NV	VC	CV	VV
Selected (%)	NN	**81**	0	0	19	0	0	0	0	0
	CN	0	**94**	0	6	0	0	0	0	0
	NC	0	0	**93**	6	0	0	1	0	0
	CC	11	0	0	**89**	0	0	0	0	0
	VN	0	0	0	0	**89**	0	11	0	0
	NV	0	0	0	0	0	**89**	0	11	0
	VC	0	0	0	1	10	0	**89**	0	0
	CV	0	0	0	1	0	9	0	**90**	0
	VV	0	0	0	10	0	0	0	0	**90**

Table B.3. Block II, $n = 65$, $K = 10$. Percentage of time each model is selected by the Heuristic search.

Appendix C

Appendix to Chapter 4

C.1 Proposal and full conditional distributions

This appendix presents the full conditional distributions and the proposal distributions that have been derived for the latent coordinates and the latent space parameters, introduced in Sections 4.2.2 and 4.3.

C.1.1 Latent coordinates

To derive a proposal distribution for the latent coordinate $\boldsymbol{z}_i$, $i = 1, \dots, n$, we fix the number of components to a constant G, as is the case at each iteration of the estimation procedure[1] illustrated in Section 4.3. Then, from equations 4.3 and 4.1, it can be easily derived the log posterior distribution for the i^{th} latent coordinate. Without loss of information, we consider such log posterior for the r^{th} latent space dimension, $r = 1, \dots, p$, being proportional to:

$$\sum_{j \neq i}^{n} \sum_{k=1}^{K} \Big[y_{ij}^{(k)} \beta^{(k)} d_{ij} - \log\big(1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}\big) \Big] + \log\Big[\sum_{g=1}^{G} \pi_g \frac{1}{\sqrt{2\pi\sigma_{rg}^2}} \exp\{ -\frac{1}{2\sigma_{rg}^2} (z_{ri} - \mu_{rg})^2 \} \Big]$$

The double summation term in the above equation is the log-likelihood of the model, while the other term depends on the i^{th} latent position prior distribution. If we define

$$a_{irg} = \pi_g \frac{1}{\sqrt{2\pi\sigma_{rg}^2}} \exp\{ -\frac{1}{2\sigma_{rg}^2} (z_{ri} - \mu_{rg})^2 \} = \exp(x_{irg}) \quad \Leftrightarrow \quad x_{irg} = \log(a_{irg}),$$

we may rewrite the last term in the above log-posterior as an LSE (LogSumExp) function:

$$\log\Big[\sum_{g=1}^{G} \pi_g \frac{1}{\sqrt{2\pi\sigma_{rg}^2}} \exp\{ -\frac{1}{2\sigma_{rg}^2} (z_{ri} - \mu_{rg})^2 \} \Big] = \log\big[\sum_{g=1}^{G} \exp(x_{rg}) \big].$$

This LSE function is bounded:

$$\max_g \{x_{ir1}, \dots, x_{irG}\} \leq \log\big[\sum_{g=1}^{G} \exp(x_{irg}) \big] \leq \max_g \{x_{ir1}, \dots, x_{irG}\} + \log(G),$$

[1] For ease of notation, we avoid iteration indexes $t = 1, \dots, T$.

that is

$$\max_g\{\log(a_{ir1}),\dots,\log(a_{irG})\} \leq \log[\sum_{g=1}^{G} a_{irg}] \leq \max_g\{\log(a_{ir1}),\dots,\log(a_{irG})\}+\log(G).$$

Now, note that

$$\log(a_{irg}) \propto -\frac{1}{2\sigma^2_{rg}}(z_{ri}-\mu_{rg})^2 \tag{C.1}$$

is the log density of a Gaussian distribution with mean μ_{rg} and variance σ^2_{rg}. Note also that, at each iteration of the estimation procedure, $\max_g\{\log(a_{ir1}),\dots,\log(a_{irG})\}$ should be found in correspondence of the g^{th} component, where $\boldsymbol{z}_i$ is currently allocated. Hence, we may approximate the LSE function with its lower bound, according to equation C.1,

$$\log\Big[\sum_{g=1}^{G}\pi_g\frac{1}{\sqrt{2\pi\sigma^2_{rg}}}\exp\{-\frac{1}{2\sigma^2_{rg}}(z_{ri}-\mu_{rg})^2\}\Big] \approx -\frac{1}{2\sigma^2_{rg}}(z_{ri}-\mu_{rg})^2,$$

with $c_{ig}=1$ and $r=1,\dots,p$. Combining this approximation with a similar one of the log-likelihood term in the first equation (see the appendix results in D'Angelo, Murphy, and Alfò, 2018, for all $r=1,\dots,p$, we may derive the following proposal distribution for the latent position of a unit in the g^{th} component:

$$\tilde{\boldsymbol{z}}_i \mid \dots \sim N(\boldsymbol{\mu}_{\tilde{\boldsymbol{z}}_i}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{z}}_i}),$$

with

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{z}}_i} = \left(\boldsymbol{\mu}_{1\tilde{\boldsymbol{z}}_i},\dots,\boldsymbol{\mu}_{p\tilde{\boldsymbol{z}}_i}\right) \quad ; \quad \Sigma_{\tilde{\boldsymbol{z}}_i} = \begin{bmatrix}\sigma^2_{1\tilde{\boldsymbol{z}}_i} & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \sigma^2_{p\tilde{\boldsymbol{z}}_i}\end{bmatrix}\mathbf{I},$$

and

$$\mu_{r\tilde{\boldsymbol{z}}_i} = \sigma^2_{r\tilde{\boldsymbol{z}}_i}\Big[2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}(y^{(k)}_{ij}-w^{(k)}_{ij})z_{rj}+\mu_{rg}\Big] \quad ; \quad \sigma^2_{r\tilde{\boldsymbol{z}}_i} = \Big(\frac{1}{\sigma^2_{rg}}+2\sum_{k=1}^{K}\beta^{(k)}\sum_{j\neq i}|y^{(k)}_{ij}-w^{(k)}_{ij}|\Big)^{-1},$$

with $w^{(k)}_{ij}=1$ if $\alpha^{(k)}-\beta^{(k)}d_{ij}>0$ and $w^{(k)}_{ij}=0$ otherwise.

C.1.2 Component parameters

The component parameters follow a Dirichlet Process, whose base distribution is a Normal Inverse Gamma (Section 4.3). The conjugacy allows a straightforward derivation of the full conditional distributions for $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$, $g=1,\dots,G$[2]:

$$\mu_{rg} \mid \dots \sim N\Big(\frac{\tau_z\sum_{i\in g}z_{ri}+m_{rz}}{1+n_g\tau_z}, \frac{\tau_z\sigma^2_{rg}}{1+n_g\tau_z}\Big) \quad ; \quad \sigma^2_{rg} \mid \dots \sim Inv\Gamma(x_g, X_{rg}), \quad r=1,\dots p,$$

[2] As done for the latent coordinates, we fix the number of components to a constant G.

with n_g the number of latent coordinates currently allocated in the g^{th} component, that is $n_g = \pi_g n$. Also,

$$x_g = \frac{n_g + 1 + 2\nu_1}{2} \quad ; \quad X_{rg} = \frac{\tau_z \sum_{i \in g}(z_{ri} - \mu_{rg})^2 + (\mu_{rg} - m_{rg})^2 + 2\tau_z\nu_2}{2\tau_z}.$$

The components mean and covariance are then:

$$\boldsymbol{\mu}_g = \Big(\mu_{1g}, \ldots, \mu_{pg}\Big) \quad ; \quad \boldsymbol{\Sigma}_g = \begin{bmatrix} \sigma_{1g}^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{pg}^2 \end{bmatrix} \mathbf{I}, \quad g = 1, \ldots, G.$$

If the hyperparameter $\boldsymbol{m} = (m_1, \ldots, m_r, \ldots, m_p)$ is given standard multivariate Gaussian distribution, it may be updated from the following conditional distribution:

$$m_r \mid \cdots \sim N\left(\sum_{g=1}^{G} \frac{\mu_{rg}}{\sigma_{rg}^2} / (\tau_z + x_r), \frac{\tau_z}{\tau_z + x_r}\right), \quad r = 1, \ldots, p,$$

with $x_r = \sum_{g=1}^{G} \frac{1}{\sigma_{rg}^2}$.

C.1.3 Concentration parameter

Following Escobar and West, 1995, the update of the Dirichlet process concentration parameter ψ is performed as follows. Given a fixed value of G,

1. Simulate the latent variable $x \sim BETA(\xi_1 + 1, n)$;
2. Compute $\pi = \frac{p}{1+p}$, with $p = \frac{\xi_1 + G - 1}{n(\xi_2 - \log(x))}$;
3. Generate a new value of the concentration parameter ψ from:

$$\begin{cases} \Gamma\Big(\xi_1 + G, \xi_2 - \log(x)\Big) & \text{with probability} \quad \pi \\ \Gamma\Big(\xi_1 + G - 1, \xi_2 - \log(x)\Big) & \text{with probability} \quad 1 - \pi \end{cases}.$$

C.1.4 Cluster labels

Given the i^{th} cluster label $\boldsymbol{c}_i$, following Bush and MacEachern, 1996, we may compute the posterior probability for a the i^{th} latent coordinate to be assigned to the g^{th} component as:

$$P_{ig} = Pr(c_{ig} = 1 \mid \ldots) = \begin{cases} n_g^{-(i)} MVN_p(\boldsymbol{z}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) & g = 1, \ldots, G \\ \psi \int MVN_p(\boldsymbol{z}_i \mid \mu, \sigma^2) d(\mu \mid \sigma^2) d(\sigma^2) & g = G + 1 \end{cases},$$

where G is the current number of components and n_g is the number of latent coordinates in the g^{th} component, discarding the i^{th} one, if present. The above integral is that of a multivariate Student t distribution:

$$\int MVN_p(\boldsymbol{z}_i \mid \mu, \sigma^2) d(\mu \mid \sigma^2) d(\sigma^2) = MVNt_{2\nu_1}\big(m, \frac{\nu_2}{\nu_1}(1 + \tau_z)\mathbf{I}\big).$$

New cluster label vectors c_i, $i = 1, \ldots, n$, are sampled sequentially from multinomial distributions with probabilities $\big(P_{i1}, \ldots, P_{ig}, \ldots, P_{iG}\big)$. If, for the i^{th} latent coordinate, we have that $\boldsymbol{c}_i = \big(c_{i1}, \ldots, c_{ig}, \ldots, c_{iG}, c_{i(G+1)}\big) = \big(0, \ldots, 0, \ldots, 0, 1\big)$, a new component is formed, which includes, for now, only $\boldsymbol{z}_i$. In this case, new component parameters need to be sampled from the corresponding full conditional distributions, previously defined. We propose to initialize first $\boldsymbol{\mu}_{G+1}$ with $\boldsymbol{z}_i$, than sample a value for $\boldsymbol{\Sigma}_{(G+1)}$ and update the value of $\boldsymbol{\mu}_{(G+1)}$. Also, to all the other $\boldsymbol{c}_j$ vectors, $j \neq i$ is added a $(G+1)^{\text{th}}$ 0 element. If instead, $c_{ig} = 1$ for some $g \leq G+1$, the $(G+1)^{\text{th}}$ element is removed from vector $\boldsymbol{c}_i$, to uniform its length with those of other label vectors. Last, if the update of a vector $\boldsymbol{c}_i$ empties a given component g, then the g^{th} element is removed from all $\boldsymbol{c}_i$ vectors, $i = 1 \ldots, n$.

Appendix D

Estimation of the Latent Space Joint Model with an MCMC Approach

D.1 Introduction

In the first chapter we have introduced a latent space model for multidimensional network data. Even though the model was developed to analyze the Eurovision Song Contest multiplex, it is quite general and can be used to study other multidimensional network data. Our model is an extension of the *lsjm* model by Gollini and Murphy, 2016, where the authors assume a joint latent space underlying the observed multiplex. In that context, the authors proposed to estimate the node-specific latent coordinates using a variational EM algorithm, where the joint latent space is approximated locally in each network. The result is a collection of network-specific latent spaces, which are then used to represent the association structure observed in the different views of the multiplex. Hence, a priori, the multiplex is generated by a single latent space but, a posteriori, this *joint* space splits in K spaces, one for each network. Instead, in our model we estimate a single latent space for the whole set of networks in the multiplex and introduce a network-specific scale coefficient to weight the influence of the latent space in each network. This coefficient allow us to verify whether a latent space is really needed to depict the association structure in a multiplex, or if the latter is just random.
According to the research question at hand, one could choose to use either the model we propose or the *lsjm* model. If, for example, one is interested in obtaining a joint representation of the data and locating the single network-specific variations of the nodes coordinates, the *lsjm* model can be useful. On the other hand, when the interest is at depicting the overall association between the nodes in the multiplex, or at verifying whether the association is potentially random, the model we propose seems more appropriate.
In Chapter 2 we have compared the performance of the two models, using synthetic multiplex data generated from a single latent space with unitary scale coefficients for all the networks. We found out that the *lsjm* model, implemented in the *lvm4net* package, did very poorly in recovering the latent coordinates, whose estimates were

also quite unstable. Such a problem could be the result of the variational approach rather than ascribed to the model itself. Usually, a variational algorithm is much faster when compared to a standard MCMC algorithm, but it suffers from high risk of getting stuck in local optima.
In this appendix, we present a different approach to estimate the *lsjm* model, where inference is carried out via an MCMC algorithm. Our formulation of the *lsjm* slightly differs from that of Gollini and Murphy, 2016, and we will point out and motivate such differences in Section D.2. Then, in Section D.3, we explain how to estimate this model and show the improvements we obtain for the estimates using the simulation study described in Section D.4.

D.2 Latent Space Joint Model

Let us consider a binary observed multiplex $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(K)}\}$, with $k = 1, \dots, K$ indexing the networks and $i, j = 1, \dots, n$ the nodes. The *lsjm* model assumes that the probability of observing an edge between nodes i and j in the k^{th} network of the multiplex is given by:

$$P\Big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \boldsymbol{z}_i, \boldsymbol{z}_j\Big) = \frac{\exp\{\alpha^{(k)} - d(\boldsymbol{z}_i, \boldsymbol{z}_j)\}}{1 + \exp\{\alpha^{(k)} - d(\boldsymbol{z}_i, \boldsymbol{z}_j)\}} = p_{ij}^{(k)}, \tag{D.1}$$

where the function $d(\cdot, \cdot)$ represents the squared Euclidean distance. To estimate the parameters and the latent coordinates in this model, Gollini and Murphy, 2016 use a variational inference approach, in which they introduce the variational latent coordinates $\boldsymbol{z}_i^{(k)}$, $k = 1, \dots, K$. In this way, they recover estimates for both the overall positions, $\boldsymbol{z}_i$, and the position given in a particular network view, $\boldsymbol{z}_i^{(k)}$.
To mimic the *lsjm*, we start by assuming that nodes-specific coordinates vary with the network, that is, we consider $\boldsymbol{z}_i^{(k)}$, with $k = 1, \dots, K$. Then, we assume that the probability of observing an edge between nodes i and j in the k^{th} network of the multiplex is given by:

$$P\Big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \boldsymbol{z}_i^{(k)}, \boldsymbol{z}_j^{(k)}\Big) = \frac{\exp\{\alpha^{(k)} - d(\boldsymbol{z}_i^{(k)}, \boldsymbol{z}_j^{(k)})\}}{1 + \exp\{\alpha^{(k)} - d(\boldsymbol{z}_i^{(k)}, \boldsymbol{z}_j^{(k)})\}} = p_{ij}^{(k)}, \tag{D.2}$$

with $d(\boldsymbol{z}_i^{(k)}, \boldsymbol{z}_j^{(k)}) = \sum_{l=1}^{p} (z_{il}^{(k)} - z_{jl}^{(k)})^2 = d_{ij}^{(k)}$, and $\mathbf{D}^{(k)}$ the distance matrix in the k^{th} network. The dimension of the latent space is denoted by p. In this way, each node is allowed to adapt to the different networks via (potentially) different values of its network-specific coordinates $z_i^{(k)}$. To reduce our formulation to that of the *lsjm* model, we assume that all the K sets of latent coordinates are generated by a single, joint, latent space, where each node has "central" coordinate $\boldsymbol{c}_i$. Also, as in Gollini and Murphy, 2016, we assume that the latent coordinates are normally distributed:

$$\boldsymbol{z}_i^{(k)} \sim MVN_p(\boldsymbol{c}_i, \sigma_z^2 \mathbf{I}),$$

with

$$\boldsymbol{c}_i \sim MVN_p(0, \sigma_c^2 \mathbf{I}).$$

Hence, the overall behaviour of the nodes is still captured, via the distances between the central coordinates in the joint space, collected in a matrix that we denote by $\boldsymbol{C}$.

D.2.1 Further issues

Note that, in the particular case of null distances, we obtain

$$p_{ij}^{(k)} = \frac{\exp\{\alpha^{(k)}\}}{1+\exp\{\alpha^{(k)}\}} = p_{RG},$$

and the model describes a random graph with edge probability p_{RG}; in this sense, edge probabilities are bounded from above by p_{RG}. Therefore, as detailed in Chapter 2, to avoid numerical identifiability problem, we bound the intercept parameter $\alpha^{(k)}$ so that the graph corresponding to p_{RG} is not disconnected:

$$\alpha^{(k)} > \log\left(\frac{\log(n)}{n-\log(n)}\right) = LB(\alpha^{(k)}) = LB(\alpha).$$

Note that bounding the intercepts, which are not bounded in the approach proposed by Gollini and Murphy, 2016, does not prevent us from comparing the estimates of the latent coordinates obtained with the proposed approach and that by Gollini and Murphy, 2016. Indeed, the intercept parameters do not influence the relative distances among the nodes in the network-specific spaces and in the joint space. The estimates of the latent coordinates are still comparable via Procrustes correlation, which is invariant to translations and rotations of the space.
Note also that by defining the edge probabilities as in equation (D.2), an identifiability issue arises:

$$\alpha^{(k)} - d_{ij}^{(k)} = (\alpha^{(k)} + r) - (d_{ij}^{(k)} + r) = \alpha^{(k)\star} - d_{ij}^{(k)\star}.$$

To overcome such an issue, a "reference" distance $d_{ij}^{(k)}$ must be fixed in each network k. We propose to fix $d_{ij}^{(k)} = d_{12}^{(k)}$ by imposing $\boldsymbol{z}_1^{(k)} = (0)_p$ and $\boldsymbol{z}_2^{(k)} = (1)_p$. This constraint does not prevent us from recovering the distances between the nodes in the different latent spaces, as we are not interested in the absolute value of such distances. Indeed, the interest is in the relative distances, that is, in which nodes are close and which are far apart.

D.3 Parameter estimation

The log-likelihood according to equation (D.2) is:

$$\begin{aligned}\ell(\boldsymbol{\alpha}, \boldsymbol{D} \mid \mathbf{Y}) &= \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i} y_{ij}^{(k)} \log\left(\frac{\exp\{\alpha^{(k)} - d_{ij}^{(k)}\}}{1+\exp\{\alpha^{(k)} - d_{ij}^{(k)}\}}\right) + (1-y_{ij}^{(k)})\log\left(\frac{1}{1+\exp\{\alpha^{(k)} - d_{ij}^{(k)}\}}\right) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j\neq i} y_{ij}^{(k)}(\alpha^{(k)} - d_{ij}^{(k)}) - \log\Big(1+\exp\{\alpha^{(k)} - d_{ij}^{(k)}\}\Big)\end{aligned} \tag{D.3}$$

The prior distribution for the network intercepts is described by a truncated Gaussian distribution:

$$\alpha^{(k)} \sim N_{\left[LB(\alpha),\infty\right]}\left(\mu_\alpha, \sigma_\alpha^2\right)$$

The unknown $\mu_\alpha, \sigma^2_\alpha$ are nuisance parameters. Indeed, their value is of limited interest but their specification is relevant, as they determine the estimates of $\alpha^{(k)}$, $k = 1, \dots, K$. As done in the model presented in Chapter 2, we decide to estimate these parameters, via an extra layer of dependence in the model:

$$\mu_\alpha | \sigma^2_\alpha \sim N_{\left[LB(\alpha), \infty\right]}(m_\alpha, \tau_\alpha \sigma^2_\alpha),$$

$$\sigma^2_\alpha \sim \text{Inv}\chi^2_{\nu_\alpha}.$$

The posterior distribution for the model parameters is

$$P(\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{c}, \mu_\alpha, \sigma^2_\alpha | \mathbf{Y}) = L(\boldsymbol{\alpha}, \boldsymbol{z} | \mathbf{Y}) \pi(\boldsymbol{z} \mid \boldsymbol{c}) \pi(\boldsymbol{c}) \pi(\boldsymbol{\alpha} | \mu_\alpha, \sigma^2_\alpha) \pi(\mu_\alpha | \sigma^2_\alpha, \tau_\alpha) \pi(\sigma^2_\alpha | \nu_\alpha), \tag{D.4}$$

and Figure D.1 synthesizes the structure of the model. The nuisance parameters have full conditional distributions in closed form, which are equivalent to those we have defined for the model in Chapter 2; they can be found in A.1. Instead, the posterior distributions for the parameters $\alpha^{(k)}$ are not available in closed form. However, a proposal distribution for the intercept terms can be derived in a similar way to what we have done in Chapter 2:

$$\tilde{\alpha}^{(k)} \mid \mathbf{Y}, \mathbf{D}, \mathbf{H}, \mu_\alpha, \sigma^2_\alpha \sim N\left(\mu_{\tilde{\alpha}^{(k)}}, \sigma^2_{\tilde{\alpha}^{(k)}}\right),$$

with

$$\mu_{\tilde{\alpha}^{(k)}} = \sigma^2_{\tilde{\alpha}^{(k)}} \left\{ E^{(k)} - \sum_{i=1}^{n} \sum_{j \neq i} \frac{\exp\{\mu_\alpha - d^{(k)}_{ij}\}}{1 + \exp\{\mu_\alpha - d^{(k)}_{ij}\}} \right\} + \mu_\alpha,$$

$$\sigma^2_{\tilde{\alpha}^{(k)}} = \left\{ \sum_{i=1}^{n} \sum_{j \neq i} \frac{\exp\{\mu_\alpha - d^{(k)}_{ij}\}}{\left(1 + \exp\{\mu_\alpha - d^{(k)}_{ij}\}\right)^2} + \frac{1}{\sigma^2_\alpha} \right\}^{-1}.$$

with $E^{(k)} = \sum_{i=1}^{n} \sum_{j \neq i} y^{(k)}_{ij}$.
Similarly, the posterior distribution for the i^{th} node coordinate can not be expressed in closed form; however, we can use the following proposal distribution:

$$\tilde{\boldsymbol{z}}_i \mid \mathbf{Y}, \boldsymbol{\alpha}, \mathbf{D}, K \sim MVN_p\left(\mu_{\tilde{z}_i}, \sigma^2_{\tilde{z}_i} \mathbf{I}\right),$$

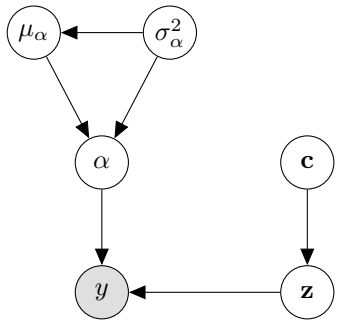


Figure D.1. Hierarchy structure of our formulation of the *lsjm* model.

where

$$\mu_{\tilde{z}_i} = \sigma^2_{\tilde{z}_i}\Big(2\sum_{k=1}^{K}\sum_{j\neq i}(y_{ij}^{(k)} - w_{ij}^{(k)})\boldsymbol{z}_j\Big), \quad \sigma^2_{\tilde{z}_i} = \left(1 + 2\sum_{k=1}^{K}\sum_{j\neq i}|y_{ij}^{(k)} - w_{ij}^{(k)}|\right)^{-1},$$

and

$$w_{ij}^{(k)} = \begin{cases} 1 & \text{if} \quad \alpha^{(k)} - d_{ij}^{(k)} > 0 \\ 0 & \text{if} \quad \alpha^{(k)} - d_{ij}^{(k)} \leq 0 \end{cases}.$$

This proposal has been derived according to Appendix A.1. For more details on the proposal distributions of network-specific intercepts and latent coordinates, we refer to Appendix A.1.

Last, the posterior distribution for the i^{th} node central coordinate does have closed form,

$$\boldsymbol{c}_i \sim MVN_p(\mu_{c_i}, \sigma^2_{c_i}\mathbf{I})$$

where

$$\sigma^2_{c_i} = \Big(\frac{1+\sigma_c^2}{\sigma_c^2}\Big)^{-1}, \quad \mu_{c_i} = \sigma^2_{c_i}\Big(\frac{\sum_{k=1}^{K}\boldsymbol{z}_i^{(k)}}{\sigma_z^2}\Big)$$

D.3.1 Algorithm for model estimation

The model parameters are estimated using a Metropolis within Gibbs MCMC algorithm, subdivided into sequential steps. First, new values for the nuisance parameters and the joint latent coordinates are sampled from their full conditional distributions (see Section D.3). Later, the intercepts and the network-specific latent coordinates are updated via a Metropolis step, using the proposal distributions defined in Section D.3. At every iteration of the algorithm, we check via Procrustes correlation whether the new sets of latent coordinates are identical to the old ones (see Chapter 2).

The hyperparameters $\{m_\alpha, \tau_\alpha, \nu_\alpha, \sigma_c^2, \sigma_z^2\}$ have to be specified a priori. We propose to fix the first three, which are the intercept-related nuisance hyperparameters, as done in Chapter 2. For the variances of the prior distributions for the joint and the network-specific latent coordinates we assume $\sigma_c^2 = \sigma_z^2 = 1$. The last quantity that needs to be specified is the number of latent dimensions p. As no specific criterion has jet been proposed in the literature regarding this issue, the specification of p is left to the researcher, depending on the type of data analysed and the aim of the analysis. In the present context, we fix $p = 2$, as we propose to compare in a simulation setting the performance of our implementation of the *lsjm* model and that obtained via the *lvm4net* package, the R package by Gollini and Murphy, 2016 implementing the variational EM algorithm for *lsjm* model estimation. Indeed, the results presented in the paper by Gollini and Murphy, 2016 refer to latent spaces with $p = 2$ dimensions. The simulation study is presented in the following section.

D.4 Simulation study

To compare the performance of the two different implementations of the *lsjm* model, we have designed the following simulation study. Four different sets of multiplex data have been simulated:

1. Scenario 1: 20 multidimensional networks with $n = 25$ and $K = 3$;
2. Scenario 2: 20 multidimensional networks with $n = 25$ and $K = 5$;
3. Scenario 3: 20 multidimensional networks with $n = 50$ and $K = 5$;
4. Scenario 4: 20 multidimensional networks with $n = 100$ and $K = 5$.

The different multiplexes have been simulated according to the framework described in Sections D.2 and D.3, to represent both small multiplex, $(n = 25, K = 3)$ and $(n = 25, K = 5)$, and moderately big multiplex, $(n = 50, K = 5)$ $(n = 50, K = 5)$.
On the different sets we estimated the following models:

- The *lsjm* model as implemented in the R package *lvm4net*. We will denote it by *lvm4net*;
- Our version of the *lsjm* model, as proposed in this appendix. We will denote it by *lsjmMCMC*;
- The latent space model proposed in Chapter 2, as implemented in the R package *spaceNet*. We will denote it by *spaceNet*.

The first two approaches will be compared both for the estimates of the network-specific latent coordinates $\boldsymbol{z}_i^{(k)}$ and the central-coordinates $\boldsymbol{c}_i$. In particular, it will be checked how well the two approaches can recover the simulated latent coordinates, by means of Procrustes correlation between estimated and simulated latent spaces. The model implemented in *spaceNet* will be tested in a stressed scenario, that is, when the observed networks have not been generated by a common latent space but by network-specific latent spaces, defined by random variations of the common latent space.
Both types of MCMC algorithms are run for 50000 iteration, with a burn-in of 15000 iterations. The variational EM algorithm in *lvm4net* is run using its default values with the first two sets of multiplexes. The last two, bigger, multiplexes required an adjustment of the tolerance level in the EM algorithm, which was raised from 0.01 to 0.1, as well the variances for the prior distributions of the intercept parameters, raised from 2 to 40. These adjustments were needed as lower values of the tolerance level and the variances lead to problems in the estimation of the latent coordinates covariance matrices, and, consequently, to continuous crushes of the EM algorithm. Tables D.1, D.2, D.3 and D.4 report the results of this comparison. Our implementation of the *lsjm* model always returns faithful estimations of the latent coordinates, both at a network-specific and at a joint level. The Procrustes correlation values are always above 0.80 on average and get better as the size of the multiplex increases. On the contrary, the *lsjm* model implemented in *lvm4net* does poorly in recovering the true latent coordinates, with an average Procrustes correlation always below 0.30. The performances of the *lvm4net* EM algorithm get worse as the size of the multiplexes increases, but this could be due to the higher values of the tolerance level and variance fixed in the estimation.
The *spaceNet* MCMC algorithm recovered a single latent space for each multidimensional network data, which was then compared with the corresponding simulated joint latent space $\mathbf{c}$. In all the different scenarios, the *spaceNet* algorithm always

recovered a latent space highly correlated with the *true* joint space (average Procrustes correlations above 0.79), proving that it is actually able to find the common latent space, even when this is subject to variations at single-network level.

		$\mathbf{z}^{(1)}$	$\mathbf{z}^{(2)}$	$\mathbf{z}^{(3)}$	$\mathbf{c}$
lsjmMCMC	mean PC	0.86	0.87	0.85	0.83
	sd PC	0.08	0.05	0.07	0.07
lvm4net	mean PC	0.19	0.22	0.23	0.23
	sd PC	0.07	0.06	0.15	0.07
spaceNet	mean PC	-	-	-	0.79
	sd PC	-	-	-	0.06

Table D.1. Simulated multiplex with $n = 25$ and $K = 3$. Mean and standard deviations of the Procrustes correlations between simulated and estimated latent spaces, according to the three different approaches.

		$\mathbf{z}^{(1)}$	$\mathbf{z}^{(2)}$	$\mathbf{z}^{(3)}$	$\mathbf{z}^{(4)}$	$\mathbf{z}^{(5)}$	$\mathbf{c}$
lsjmMCMC	mean PC	0.82	0.82	0.82	0.81	0.82	0.83
	sd PC	0.15	0.14	0.13	0.14	0.15	0.13
lvm4net	mean PC	0.23	0.25	0.27	0.23	0.21	0.21
	sd PC	0.12	0.09	0.12	0.09	0.08	0.07
spaceNet	mean PC	-	-	-	-	-	0.85
	sd PC	-	-	-	-	-	0.08

Table D.2. Simulated multiplex with $n = 25$ and $K = 5$. Mean and standard deviations of the Procrustes correlations between simulated and estimated latent spaces, according to the three different approaches.

		$\mathbf{z}^{(1)}$	$\mathbf{z}^{(2)}$	$\mathbf{z}^{(3)}$	$\mathbf{z}^{(4)}$	$\mathbf{z}^{(5)}$	$\mathbf{c}$
lsjmMCMC	mean PC	0.92	0.92	0.92	0.91	0.89	0.92
	sd PC	0.05	0.02	0.03	0.06	0.10	0.03
lvm4net	mean PC	0.13	0.10	0.11	0.11	0.10	0.10
	sd PC	0.06	0.04	0.06	0.05	0.05	0.05
spaceNet	mean PC	-	-	-	-	-	0.91
	sd PC	-	-	-	-	-	0.04

Table D.3. Simulated multiplex with $n = 50$ and $K = 5$. Mean and standard deviations of the Procrustes correlations between simulated and estimated latent spaces, according to the three different approaches.

D.5 Discussion

In the present appendix we have proposed a different implementation of the *lsjm* model by Gollini and Murphy, 2016, where parameter estimation is based on an MCMC algorithm. Our method has been tested in a simulation study and

		$\mathbf{z}^{(2)}$	$\mathbf{z}^{(2)}$	$\mathbf{z}^{(3)}$	$\mathbf{z}^{(4)}$	$\mathbf{z}^{(5)}$	$\mathbf{c}$
lsjmMCMC	mean PC	0.86	0.82	0.83	0.88	0.84	0.84
	sd PC	0.09	0.11	0.12	0.06	0.13	0.10
lvm4net	mean PC	0.10	0.11	0.10	0.10	0.07	0.11
	sd PC	0.04	0.06	0.03	0.03	0.05	0.04
spaceNet	mean PC	-	-	-	-	-	0.90
	sd PC	-	-	-	-	-	0.03

Table D.4. Simulated multiplex with $n = 100$ and $K = 5$. Mean and standard deviations of the Procrustes correlations between simulated and estimated latent spaces, according to the three different approaches.

returned faithful estimation of the latent coordinates; the original *lsjm* model, as implemented in the *lvm4net* package, proves to be less able to recover the simulated latent spaces, at a network-specific and a joint level. We have also proved that the model proposed in Chapter 2 can be seen as a parsimonious version of the *lsjm* model, and it proved able to recover the joint space when the networks in the multiplex have been generated from different latent spaces, defined as a variation of a common latent space.

Bibliography

Airoldi, E.M. et al. (2008). "Mixed-membership Stochastic Blockmodels". In: *Journal of Machine Learning Research* 9, 1981–2014 (cit. on pp. 8, 16, 62).

Akaike, H. (1974). "A New Look at the Statistical Model Identification". In: *IEEE Transactions on Automatic Control* 19.6, 716–723 (cit. on p. 53).

Aldous, D. (1985). "Exchangeability and related topics". In: *École d' été de probabilités de Saint-Flour XIII–1983*. Vol. 1117 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1–198 (cit. on p. 69).

Antoniak, C.E. (1974). "Mixtures of Dirichlet processes with applications to bayesian nonparametric problems". In: *The Annals of Statistics* 2.6, 1152–1174 (cit. on pp. 63–65, 67).

Aquastat (2011). *Maldives*. URL: \url{http://www.fao.org/nr/water/aquastat/countries_regions/MDV/MDV-CP_eng.pdf} (cit. on p. 53).

Bartolucci, F., M.F. Marino, and S. Pandolfi (2018). "Dealing with reciprocity in dynamic stochastic block models". In: *Computational Statistics and Data Analysis* 123, pp. 86–100 (cit. on p. 62).

Bernardo, J. and F. Giron (1998). "A Bayesian analysis of simple mixture problems". In: *Bayesian Statistics* 3, pp. 67–78 (cit. on p. 74).

Blangiardo, M. and G. Baio (2014). "Evidence of bias in the Eurovision song contest: modelling the votes using Bayesian hierarchical models". In: *Journal of Applied Statistics* 41.10, pp. 2312–2322 (cit. on pp. 19, 29).

Bouveyron, C., P. Latouche, and R. Zreik (2018). "The stochastic topic block model for the clustering of vertices in networks with textual edges". In: *Statistics and Computing* 28.1, pp. 11–31 (cit. on p. 62).

Bush, C.A. and S.N. MacEachern (1996). "A semiparametric Bayesian model for randomised block designs". In: *Biometrika* 83, 275–285 (cit. on pp. 69, 127).

Butts, C.T. and K.M. Carley (2005). "Some simple algorithms for structural comparison". In: *Computational and Mathematical Organization Theory* 11.4, 291–305 (cit. on p. 17).

Carmona, C., L. Nieto-Barajas, and A. Canale (2018). "Model-based approach for household clustering with mixed scale variables". In: *Advances in Data Analysis and Classification*, pp. 1–25 (cit. on p. 71).

CBI (2015). *Fresh fruit and vegetables in Europe*. URL: \url{https://www.cbi.eu/sites/default/files/market_information/researches/trade-statistics-europe-fresh-fruit-vegetables-2015.pdf} (cit. on p. 53).

Clerides, S. and T. Stengos (2006). "Love thy neighbour, love thy kin: Strategy and bias in the Eurovision Song Contest". In: *Discussion paper series of Centre for Economic Policy Research* 5732, pp. 1–28 (cit. on p. 19).

Dahl, D.B. (2006). "Model-based clustering for expression data via a Dirichlet process mixture model". In: *Bayesian inference for gene expression and proteomics*. Cambridge University Press, Cambridge (cit. on pp. 71, 81).

D'Angelo, S., M. Alfò, and T.B. Murphy (2018). "Node-specific effects in latent space modelling of multidimensional networks". In: *arXiv:1807.03874 [stat.ME]* (cit. on p. 13).

D'Angelo, S. and M. Fop (2018). *spaceNet: Latent Space Models for Multidimensional Networks*. R package version 1.0.1. URL: \url{https://CRAN.R-project.org/package=spaceNet} (cit. on p. 85).

D'Angelo, S., T.B. Murphy, and M. Alfò (2018). "Latent Space Modeling of Multidimensional Networks with Application to the Exchange of Votes in Eurovision Song Contest". In: *Annals of applied statistics (accepted). arXiv:1803.07166 [stat.AP]* (cit. on pp. 12, 42, 45, 46, 66, 67, 71, 85, 126).

Dryden, I.L. and K.V. Mardia (1999). *Statistical shape analysis*. Wiley Series in Probability and Statistics (cit. on p. 25).

Duijn, M.A.J. van, T.A.B. Snijders, and B.H. Zijlstra (2004). "p_2: a random effects model with covariates for directed graphs". In: *Statistica Neerlandica* 58, pp. 234–254 (cit. on pp. 7, 42, 61).

Durante, D. and D. Dunson (2016). "Locally Adaptive Dynamic Networks". In: *Annals of Applied Statistics* 10.4, pp. 2203 –2232 (cit. on p. 42).

Durante, D., D.B. Dunson, and J.T. Vogelstein (2017). "Nonparametric Bayes Modeling of Populations of Networks". In: *Journal of the American Statistical Association* 112.520, pp. 1516–1530 (cit. on pp. 12, 17).

Erdős, P. and A. Rényi (1959). "On random graphs.I". In: *Publicationes Mathematicae* 6, pp. 290–297 (cit. on pp. 6, 15, 23, 61, 66).

— (1960). "On the evolution of random graphs". In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, pp. 17–61 (cit. on pp. 15, 23).

Escobar, M.D. and M. West (1995). "Bayesian density estimation and inference using mixtures". In: *Journal of the American Statistical Association* 90, 577–588 (cit. on pp. 69, 127).

FAOStat (2013). *Food and agriculture data*. URL: \url{http://www.fao.org/faostat/en} (cit. on p. 51).

Fenn, D. et al. (2006). "How does Europe Make Its Mind Up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest". In: *Physica A: Statistical Mechanics and its Applications* 360.2, pp. 576–598 (cit. on p. 18).

Ferguson, T.S. (1973). "A bayesian analysis of some nonparametric problems". In: *The Annals of Statistics* 1.2, 209–230 (cit. on pp. 64, 67).

Fienberg, S.E., M.M. Meyer, and S.S. Wasserman (1985). "Statistical Analysis of Multiple Sociometric Relations". In: *Journal of the American Statisical Association* 80.389, pp. 51–67 (cit. on p. 16).

Food, ITE and Drink (2017). *Fruit and vegetables in Kazakhstan: a billion dollar sector*. URL: \url{https://www.food-exhibitions.com/Market-Insights/Turkey-and-Eurasia/Fruit-vegetables-in-Kazakhstan} (cit. on p. 53).

Fosdick, B.K. et al. (2018). "Multiresolution network models". In: *Journal of Computational and Graphical Statistics (accepted)* (cit. on p. 62).

Fraley, C. and A. E. Raftery (2002). "Model-based clustering, discriminant analysis and density estimation". In: *Journal of the American Statistical Association* 97, 611–631 (cit. on p. 62).

Frank, O. and D. Strauss (1986). "Markov Graphs". In: *Journal of the American Statistical Association* 81.395, pp. 832–842 (cit. on pp. 7, 16).

FreshPlaza (2015a). *Germany: worldwide 2nd market fresh fruit and vegetables*. URL: \url{http://www.freshplaza.com/article/138874/Germany-worldwide-2nd-market-fresh-fruit-and-vegetables} (cit. on p. 53).

— (2015b). *Netherlands world's biggest (re-)exporter for 11 fruit and veg products*. URL: \url{http://www.freshplaza.com/article/135083/Netherlands-worlds-biggest-(re-)exporter-for-11-fruit-and-veg-products} (cit. on p. 53).

— (2018). *Spain became the world's largest watermelon exporter in 2017*. URL: \url{http://www.freshplaza.com/article/196648/Spain-became-the-worlds-largest-watermelon-exporter-in-2017} (cit. on p. 53).

Frühwirth-Schnatter, S. (2011). "Dealing with Label Switching under Model Uncertainty". In: *Mixtures*. Wiley-Blackwell. Chap. 10, pp. 213–239 (cit. on pp. 72, 82).

Frühwirth-Schnatter, S. and G. Malsiner-Walli (2018). "From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering". In: *Advances in Data Analysis and Classification* (cit. on p. 63).

Gelman, A. et al. (2014). *Bayesian Data Analysis*. CRC press, pp. 173–174 (cit. on p. 53).

Ginsburgh, V. and A. Noury (2008). "The eurovision song contest. Is voting political or cultural?" In: *European Journal of Political Economy* 24, pp. 41–52 (cit. on p. 19).

Goldenberg, A. et al. (2010). "A Survey of Statistical Network Models". In: *Foundations and Trends in Machine Learning* 2, 129–233 (cit. on p. 16).

Gollini, I. and T.B. Murphy (2016). "Joint Modeling of Multiple Network Views". In: *Journal of Computational and Graphical Statistics* 25.1, pp. 246–265 (cit. on pp. 8, 12, 13, 17, 20–22, 24, 39–41, 66, 87, 88, 129–131, 133, 135).

Gormley, I.C. and T.B. Murphy (2010). "A mixture of experts latent position cluster model for social network data". In: *Statistical Methodology* 7, pp. 385–405 (cit. on p. 62).

Grazian, C. and C.P. Robert (2015). "Jeffreys prior for mixture estimation". In: *Bayesian Statistics from Methods to Models and Applications*. Vol. 126. Springer Proceedings in Mathematics and Statistics, Springer, pp. 37–48 (cit. on p. 74).

Green, P.J. (1995). "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination". In: *Biometrika* 82.4, 711–732 (cit. on p. 63).

Greene, D. and P. Cunningham (2013). "Producing a unified graph representation from multiple social network views". In: *Proceedings of the 5th Annual ACM Web Science Conference (Web-Sci'13)*, 118–121 (cit. on p. 16).

Handcock, M., A. Raftery, and J. Tantrum (2007). "Model-based clustering for social networks (with discussion)". In: *Journal of the Royal Statistical Society: Series A* 170.2, pp. 1–22 (cit. on pp. 16, 24, 62–65, 68, 85).

Hjort, N.L. et al. (2010). *Bayesian Nonparametrics*. Cambridge University Press (cit. on pp. 67, 69).

Hoff, P. (2005). "Bilinear Mixed-Effects Models for Dyadic Data". In: *Journal of the American Statistical Association* 100.469, pp. 286–295 (cit. on pp. 9, 16, 41, 42).
— (2011). "Hierarchical multilinear models for multiway data". In: *Computational Statistics and Data Analysis* 55, 530–543 (cit. on pp. 17, 42).
— (2015). "Dyadic data analysis with amen". In: URL: \url{arXivpreprint.arXiv:1506.08237} (cit. on pp. 17, 20, 22).
Hoff, P., B. Fosdick, and K. Stovel (2013). "Likelihoods for fixed rank nomination networks". In: *Network Science* 1.3, pp. 253–277 (cit. on p. 88).
Hoff, P.D. (2003). "Random effects models for network data". In: *Dynamic social network modeling and analysis*. Vol. 126. Breiger, R., Carley, K., and Pattinson, P. (Eds.), pp. 302–322 (cit. on p. 42).
Hoff, P.D., A.E. Raftery, and M.S. Handcock (2002). "Latent Space Approaches to Social Network Analysis". In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098 (cit. on pp. 8, 9, 12, 16, 17, 20, 22, 25, 40, 41, 43, 61, 62, 64, 65).
Holland, P.W., K. Laskey, and S. Leinhardt (1983). "Stochastic Blockmodels: First Steps". In: *Social Networks* 5, 109–137 (cit. on pp. 7, 16, 61, 62).
Holland, P.W. and S. Leinhardt (1981). "An exponential family of probability distributions for directed graphs". In: *Journal of the American Statistical Association* 76.373, pp. 33–50 (cit. on pp. 6, 15, 42, 61).
Hubert, L. and P. Arabie (1985). "Comparing partitions". In: *Journal of Classification* 2.1, 193–218 (cit. on p. 75).
Johnson, Renée (2016). *The U.S. trade situation for fruit and vegetable products*. URL: \url{https://fas.org/sgp/crs/misc/RL34468.pdf} (cit. on p. 54).
Kim, B. et al. (2018). "A review of dynamic network models with latent variables". In: *Statistics Surveys* 12, pp. 105–135 (cit. on p. 11).
Krivitsky, P.N. et al. (2009). "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models". In: *Social Networks* 31.3, 204–213 (cit. on pp. 16, 42, 45).
Lynch, K. (2015). "Eurovision recognised by Guinness World Records as the longest-running annual TV music competition (international)". In: URL: \url{http://www.guinnessworldrecords.com/news/2015/5/eurovision-recognised-by-guinness-world-records-as-the-longest-running-annual-tv-379520} (cit. on p. 18).
Malsiner-Walli, G., S. Frühwirth-Schnatter, and B. Grün (2016). "Model-based clustering based on sparse finite Gaussian mixtures". In: *Statistics and Computing* 26.1, pp. 303–324 (cit. on p. 63).
Mantzaris, A.V., S.R. Rein, and A.D. Hopkins (2018). "Examining collusion and voting biases between countries during the Eurovision song contest since 1957". In: *Journal of Artificial Societies and Social Simulation* 21.1 (cit. on p. 19).
Matias, C., T. Rebafka, and F. Villers (2018). "A semiparametric extension of the stochastic block model for longitudinal networks". In: *Biometrika* 105.3, 665–680 (cit. on pp. 62, 63).
McLachlan, G. J. and S. Rathnayake (2014). "On the number of components in a Gaussian mixture model". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 4, 341–355 (cit. on p. 63).

Müller, P. et al. (2015). *Bayesian Nonparametric Data Analysis*. Springer (cit. on pp. 67, 69).

Murphy, T.B. (2015). "Model-Based Clustering for Network Data". In: *Handbook of Cluster Analysis*. CRC press, pp. 337–358 (cit. on p. 16).

Nowicki, K. and T.A.B. Snijders (2001). "Estimation and prediction of stochastic blockstructures". In: *Journal of the American Statistical Association* 96.455, pp. 1077–1087 (cit. on pp. 8, 61).

Petris, G. and L. Tardella (2003). "A geometric approach to transdimensional Markov chain Monte Carlo". In: *The Canadian Journal of Statistics* 31.4, pp. 469–482 (cit. on p. 63).

Rasmussen, C.E. (2000). "The Infinite Gaussian Mixture Model". In: *Advances in Neural Information Processing Systems*. Vol. 12. MIT Press, 554–560 (cit. on pp. 63, 65).

Robins, G. et al. (2006). "Recent developments in exponential random graph (p*) models for social networks". In: *Social Networks* 29.2, pp. 192–215 (cit. on pp. 7, 61).

Robins, G. et al. (2007). "An introduction to exponential random graph (p*) models for social network". In: *Social networks* 29.2, pp. 173–191 (cit. on p. 16).

Rosen, B. (1972). "Asymptotic theory for successive sampling with varying probabilities without replacement, I". In: *The Annals of Mathematical Statistics* 43.2, pp. 373–397 (cit. on p. 20).

Ryan, C., J. Wyse, and N. Friel (2017). "Bayesian model selection for the latent position cluster model for social networks". In: *Network Science* 5.1, pp. 70–91 (cit. on p. 62).

Saavedraa, S., J. Efstathioua, and F. Reed-Tsochasb (2007). "Identifying the underlying structure and dynamic interactions in a voting network". In: *Physica A: Statistical Mechanics and its Applications* 377.2, pp. 672–688 (cit. on p. 19).

Salter-Townshend, M. and T.H. McCormick (2017). "Latent Space Models for Multiview Network Data". In: *Annals of Applied Statistics* 11.3, pp. 1217–1244 (cit. on pp. 12, 17, 42).

Salter-Townshend, M. and T.B. Murphy (2013). "Variational Bayesian inference for the Latent Position Cluster Model for network data". In: *Computational Statistics and Data Analysis* 57, pp. 661–671 (cit. on p. 62).

Salter-Townshend, M. et al. (2012). "Review of statistical network analysis: models, algorithms, and software". In: *Statistical Analysis and Data Mining* 5.4, pp. 243–264 (cit. on pp. 4, 16).

Sarkar, P. and A. Moore (2005). "Dynamic social network analysis using latent space models". In: *ACM SIGKDD Explorations Newsletter* 7, pp. 31–40 (cit. on p. 11).

Schwarz, G. (1978). "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2, pp. 461–464 (cit. on p. 53).

Schweinberger, M. and T.A.B. Snijders (2003). "Settings in Social Networks: A Measurement Model". In: *Sociological Methodology* 33, pp. 307–341 (cit. on pp. 10, 41).

Sewell, D.K. and Y. Chen (2015). "Latent Space Models for Dynamic Networks". In: *Journal of the American Statistical Association* 110.512, pp. 1646–1657 (cit. on pp. 11, 42).

Sewell, D.K. and Y. Chen (2016). "Latent space models for dynamic networks with weighted edges". In: *Social Networks* 44, pp. 105–116 (cit. on pp. 42, 88).

— (2017). "Latent space approaches to community detection in dynamic networks". In: *Bayesian Analysis* 12.2, pp. 351–377 (cit. on p. 63).

Signorelli, M. and E.C. Wit (2018). "A Bayesian analysis of simple mixture problems". In: *Journal of the Royal Statistical Society. Series C* 67.2, 355–369 (cit. on p. 62).

Snijders, T.A.B. and K. Nowicki (1997). "Estimation and Prediciton for Stochastic Blockmodels for Graphs with latent Block Structure". In: *Journal of Classification* 14, pp. 75–100 (cit. on pp. 8, 16, 61, 62).

Spiegelhalter, D.J. et al. (2002). "Bayesian measures of model complexity and fit (with discussion)". In: *Journal of the Royal Statistical Society- B* 64.4, 583–639 (cit. on p. 31).

Spierdijk, L. and M. Vellekoop (2006). "Geography, culture, and religion: Explaining the bias in Eurovision song contest voting". In: *Memorandum, Department of Applied Mathematics, University of Twente, Enschede* 1794, pp. 1–30 (cit. on pp. 19, 29).

Sweet, T.M., A.C. Thomas, and B.W. Junker (2013). "Hierarchical network models for education research: Hierarchical latent space models". In: *Journal of Educational and Behavioural Statistics* 38.3, pp. 295–318 (cit. on pp. 12, 16).

Vickers, M. and S. Chan (1981). "Representing Classroom Social Structure". In: *Melbourne: Victoria Institue of Secondary Education* (cit. on pp. 13, 80, 84).

Wasserman, S. and K. Faust (1994). *Statistical network analysis. Methods and applications*. Cambridge University Press (cit. on pp. 3, 4).

Wasserman, S. and P. Pattison (1996). "Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p*". In: *Psychometrika* 61, pp. 401–425 (cit. on pp. 7, 61).

Worldatlas (2018). *Top Pineapple Producing Countries*. URL: `\url{https://www.worldatlas.com/articles/top-pineapple-producing-countries.html}` (cit. on p. 54).

Yair, G. (1995). "'Unite Unite Europe' The political and cultural structures of Europe as reflected in the Eurovision Song Contest". In: *Social Networks* 17.2, pp. 147–161 (cit. on pp. 15, 18).

Yang, T. et al. (2011). "Detecting communities and their evolutions in dynamic social networks—a Bayesian approach". In: *Machine learning* 82.2, 157–189 (cit. on p. 63).

Zhang, Z. et al. (2004). "Learning a multivariate Gaussian mixture model with the reversible jump MCMC algorithm". In: *Statistics and Computing* 14.4, pp. 343–355 (cit. on p. 63).