

MODELLING NODE HETEROGENEITY IN LATENT SPACE MODELS FOR MULTIDIMENSIONAL NETWORKS

Silvia D’Angelo¹, Marco Alfò², and Thomas Brendan Murphy¹

¹School Of Mathematics and Statistics, Science Centre - South, Belfield, Dublin 4, Ireland.

²Department of Statistical Sciences, Sapienza University, Piazzale Aldo Moro 5, 00185, Rome, Italy.

Abstract

Multidimensional network data can have different levels of complexity, as nodes may be characterized by heterogeneous individual-specific features, which may vary across the networks. This paper introduces a class of models for multidimensional network data, where different levels of heterogeneity within and between networks can be considered. The proposed framework is developed in the family of latent space models, and it aims to distinguish symmetric relations between the nodes and node-specific features. Model parameters are estimated via a Markov Chain Monte Carlo algorithm. Simulated data and an application to a real example, on fruits import/export data, are used to illustrate and comment on the performance of the proposed models.

Keywords: Latent Space Models, Multiplex, Markov Chain Monte Carlo

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. 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 features, which may directly influence their structure. A frequently studied feature is *transitivity*, which refers to transitive relations. Roughly speaking, transitivity in social networks can be 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 et al. (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 depends on the positions of the nodes in such a space; closer objects in the latent space have an higher propensity to be connected in the network. Different latent space approaches have been proposed in the literature, either based on metrics (see Hoff et al. (2002), Hoff (2005)) or on ultrametrics (see Schweinberger & Snijders (2003)). Among metric latent spaces, models based on the Euclidean distance are 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 models are the projection (Hoff et al. 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 the “direction” of the relation between the nodes, see Hoff et al. (2002). Extensions of the latent space models to multidimensional networks are described in Gollini & Murphy (2016), Salter-Townshend & McCormick (2017), Hoff (2011) and D’Angelo et al. (2018). Sewell & Chen (2015) introduce a latent space model for dynamic binary networks, later extended to include weighted dynamic networks, see Sewell & Chen (2016). Durante & Dunson (2016) developed a framework based on a dynamic latent space model to describe dynamic networks of face-to-face individual contacts. Other works have discussed approaches to cluster networks into sub-groups sharing similar attributes or properties. These modelling frameworks may be of particular interest when the aim is at dimensional reduction, if the number of networks is quite large, or when different networks describe opposite types of relations, e.g. preys and predators networks. For further details we refer to Durante, Dunson & Vogelstein (2017), Durante, Mukherjee & Steorts (2017), Signorelli & Wit (2020). It is worth to point out that multiplex data could be also analysed from the point of view of tensor data analysis. Within this framework, we refer to Hoff (2015), Fosdick & Hoff (2014) and Zhou et al. (2015).

Another interesting feature of network data is that of *degree heterogeneity*, which refers to the differential (higher/lower) propensity of nodes to send/receive edges when compared to others. Holland & Leinhardt (1981) proposed the so called “ p_1 ” model, where node-specific sender and receiver effects are treated as random effects. van Duijn et al. (2004) developed the “ p_2 ” model, an extension of the “ p_1 ” model, where node-specific attributes are introduced as covariates, together with sender/receiver random effects. Other extensions of the basic model are discussed 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 & 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.

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 presented in D’Angelo et al. (2018) to account for degree heterogeneity in multidimensional networks. The previous model was developed to analyse shared similarities among nodes in the specific case of the Eurovision Song Contest multiplex, where heterogeneity across the networks was limited by construction by rules established in the competition. In general, multiplex 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 context 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. A single latent space is able to capture “characterizing” traits of the nodes, i.e. their individuality. Indeed, multiple networks represent multiple expressions of the same nodes, and the use of a single latent space helps distinguishing individual, core features from view-specific ones. Heterogeneity across different views, i.e. node-specific characteristics, 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 2 we formally introduce the concept of multidimensional networks and define the proposed class of models for directed multiplexes. Section 3 details the estimation procedure and discusses issues of model identifiability. Section 4 describes the proposed class of models for the particular case of undirected multidimensional networks. A simulation study is conducted in Section 5, to investigate the properties of the proposed class of models, and to study the performance of a novel heuristic procedure for model selection. An application to FAO trade data is presented in Section 6. We conclude with a discussion in Section 7.

2 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 a set of $n \times n$ adjacency matrices $\{y^{(1)}, \dots, y^{(K)}\} = \mathbf{Y}$, where the index $k = 1, \dots, K$ denotes 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 $i, j = 1, \dots, n$. A multiplex can be either undirected, if $y_{ij}^{(k)} = y_{ji}^{(k)}$ or directed, if the different adjacency matrices are not necessarily 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, $\sum_{j \neq i} y_{ij}^{(k)}$, and the in-degree, $\sum_{j \neq i} y_{ji}^{(k)}$, 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 most active ones. Further, different views may 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. We start by introducing the general latent space framework for directed multidimensional networks. Section 4 discusses the restriction to the specific case of undirected multidimensional networks.

In latent space models based on Euclidean distances we assume that each node is located into an unobserved p -dimensional Euclidean space; according to model specification, the probability of observing an edge between the elements in the dyad (i, j) , conditionally on the latent coordinates $\mathbf{z}_i, \mathbf{z}_j$, $i, j = 1, \dots, n$, does not depend on the other nodes positions, see Hoff et al. (2002). We hold these assumptions in our model, together with a further one, as we assume that the probability of a connection between nodes in a dyad also depends on node-specific propensity to send/receive links. In multidimensional networks, the propensity may vary with the views, as an actor could be quite popular in a network, while receiving few edges in another one. Different levels of heterogeneity in edge probabilities may depend on different levels of node heterogeneity in the networks.

For this purpose, let $\theta_i^{(k)}$ and $\gamma_i^{(k)}$, $i = 1, \dots, n$, $k = 1, \dots, K$ represent the sender and the receiver effects 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. 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:

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

where $\alpha = (\alpha^{(1)}, \dots, \alpha^{(K)})$ and $\beta = (\beta^{(1)}, \dots, \beta^{(K)})$ are network-specific vector parameters and d_{ij} is the squared Euclidean distance between node i and node j in the p -dimensional latent space.

According to the node-specific behaviour, we may define three different scenarios for each effect:

- *Null* (N): The nodes do not have any specific propensity to send and/or receive links; the edge probabilities reduce to a function of the distances in the latent space alone. This scenario is parametrized as $\theta_i^{(k)} = 0$ and/or $\gamma_i^{(k)} = 0, \forall i, k$;
- *Constant* (C): The nodes exhibit a different propensity in sending/receiving links, but such propensities are constant over the considered views. This second scenario is parametrized by fixing $\theta_i^{(k)} = \theta_i$ and/or $\gamma_i^{(k)} = \gamma_i, \forall i, k$;
- *Variable* (V): The propensity of nodes to send and/or receive links varies across the views analysed. This last scenario may be parametrized considering $\theta_i^{(k)}$ or $\gamma_i^{(k)}, \forall i, k$.

Note that we assume the same type of effect (null, constant or variable) for all the views in the multiplex. 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 7. Table 1 presents a schematic taxonomy of the 9 different models arising from the different assumptions upon the sender and the receiver effects.

		Receiver effect		
		0	γ_j	$\gamma_j^{(k)}$
Sender effect	0	NN	NC	NV
	θ_i	CN	CC	CV
	$\theta_i^{(k)}$	VN	VC	VV

Table 1: The class of models defined by the different assumptions on the sender/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 $\Phi = (\Phi^{(1)}, \dots, \Phi^{(K)})$, with generic element defined by

$$\left[\phi_{ij}^{(k)}\right] = \left[g(\theta_i^{(k)}, \gamma_j^{(k)})\right], \quad g(\theta_i^{(k)}, \gamma_j^{(k)}) = \begin{cases} 1 & \text{if both effects are absent,} \\ \theta_i^{(k)} & \text{if only the sender effect is present,} \\ \gamma_j^{(k)} & \text{if only the receiver effect is present,} \\ \frac{\theta_i^{(k)} + \gamma_j^{(k)}}{2} & \text{if both effects are present.} \end{cases} \quad (2)$$

Last, we can define the function $f(\cdot)$ as follows:

$$f(\alpha^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}) = f(\alpha^{(k)} \phi_{ij}^{(k)}) = \alpha^{(k)} g(\theta_i^{(k)}, \gamma_j^{(k)}) \quad (3)$$

Equations (1-3) 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

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

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, in fact $f(\alpha^{(k)}\phi_{ij}^{(k)}) = \alpha^{(k)}g(\theta_i^{(k)}, \gamma_j^{(k)}) \leq \alpha^{(k)}$. Recalling the definition of $\phi_{ij}^{(k)}$ in equation (2), we may also notice that the sender and receiver effects can be considered as relative quantities. Indeed, inactive (or unpopular) nodes will have an effect value close to -1 , while active (or popular) nodes will have an effect value close to 1 . Obviously, the combined effect $\phi_{ij}^{(k)}$ varies in the same range. Bounding these effects in the interval $[-1, 1]$ allows for easy interpretation of the differences in levels of activity (or popularity) across nodes. The “actual” scale of the sender/receiver effects (their impact on the edge probabilities) in each network is controlled by the $\alpha^{(k)}$ parameter. Its role can be compared to that of a variance parameter controlling the scale for $\gamma_j^{(k)}$ and $\theta_i^{(k)}$. However, avoiding the use of variance parameters for the node-specific effects reduces the number of model parameters and hence model complexity, without affecting flexibility in modelling the edge probabilities.

We must also notice that if we allow $\phi_{ij}^{(k)}$ to be negative and we do consider negative intercepts $\alpha^{(k)}$, a fundamental problem arises. In fact, if we take $\alpha^{(k)} < 0$ and 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 proportional to edge probabilities. For this reason, we bound the intercept to be non-negative, $\alpha^{(k)} \geq 0 = LB(\alpha)$, $k = 1, \dots, K$. This constraint does not alter the interpretation of the intercept and other model parameters. Indeed, fixing a lower bound for the intercept does not imply a lower bound for the edge probabilities, as the impact of the distances in the latent space may decrease substantially this value.

From equation (3), we may also notice that when no sender/receiver effects are present (scenario NN in Table 1), the edge probability in (1) reduces to the model specification introduced by D’Angelo et al. (2018), for which inference procedures have already been provided. In the next section, we will focus on those scenarios that include at least one of such effects. For these models, the edge probability formula presented in equation (1) can be rewritten as:

$$p_{ij}^{(k)} = \frac{\exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij}\}}{1 + \exp\{\alpha^{(k)}\phi_{ij}^{(k)} - \beta^{(k)}d_{ij}\}} \quad (4)$$

3 Estimation

We propose a hierarchical Bayesian approach to perform inference on the latent space model proposed in Section 2. The (log-)likelihood can be derived from equation (4),

$$\ell(\alpha, \beta, \Phi, D \mid \mathbf{Y}) = \sum_{k=1}^K \sum_{\substack{i=1 \\ j \neq i}}^n \ell_{ij}^{(k)} = \sum_{k=1}^K \sum_{\substack{i=1 \\ j \neq i}}^n y_{ij}^{(k)} \log(p_{ij}^{(k)}) + (1 - y_{ij}^{(k)}) \log(1 - p_{ij}^{(k)}). \quad (5)$$

The prior distributions for model parameters are specified as follows:

$$\beta^{(k)} \sim N_{(0, \infty)}(\mu_\beta, \sigma_\beta^2), \quad \alpha^{(k)} \sim N_{(0, \infty)}(\mu_\alpha, \sigma_\alpha^2), \quad \mathbf{z}_i \sim \text{MVN}_p(0, I), \quad \gamma_j^{(k)}, \theta_i^{(k)} \sim \text{Unif}(-1, 1).$$

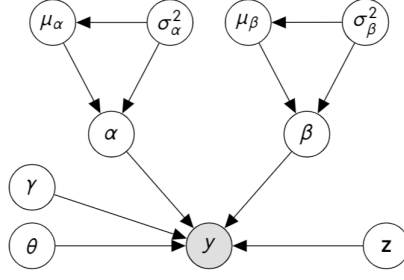


Figure 1: Hierarchy structure for the models.

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

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

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 $\beta^{(k)} \geq 0, k = 1, \dots, K$ is imposed according to the assumption that edge probabilities are inversely proportional to the distance between nodes in the latent space. For further discussion on the choice of the prior distributions for the model parameters, we refer to [D’Angelo et al. \(2018\)](#). A schematic representation of the hierarchical structure of the proposed model is displayed in Figure 1.

3.1 Identifiability

As discussed in [D’Angelo et al. \(2018\)](#), one out of the K parameters $\alpha^{(k)}$ and $\beta^{(k)}$ must be fixed to ensure parameter identifiability for the basic latent space model. The corresponding network is then referred to as the “reference” network. In the present context, fixing these two parameters is not enough, as the multiplicative effect of $\phi_{ij}^{(k)}$ may 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. More in details, when the effect is variable, we propose to choose as reference, in each network, the i^{th} node with the highest observed out-degree (or in-degree) and fix $\theta_i^{(k)} = 1$ (respectively $\gamma_i^{(k)} = 1$). Instead, when the effect is constant, we propose to select the i^{th} node with highest observed mean out-degree (or mean in-degree) and fix $\theta_i = 1$ (respectively $\gamma_i = 1$). Fixing the sender/receiver parameter for a specific 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 deriving precise point estimates for such effects.

3.2 MCMC algorithm

We propose an MCMC algorithm to estimate parameters for the latent space model proposed in section 2. Full conditionals for the nuisance parameters are available in closed form, while Metropolis-Hastings steps are needed for the other parameters, the latent coordinates and the nuisance parameters. Full conditional distributions for the nuisance parameters, proposal distributions for the network-specific parameters and the latent coordinates are described in the supplementary material, together with proposal distributions for the sender/receiver effects.

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 (5) 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 proposed algorithm scales quadratically with the number of nodes n and linearly with the number of networks.

Latent coordinates are initialized via multidimensional scaling on the average geodesic distances calculated over the different networks. Squared Euclidean distances are then computed on the starting latent positions, and are used to perform a logistic regression for the adjacency matrix binary entries to get starting values for the intercept and the coefficient parameters $\alpha^{(k)}$ and $\beta^{(k)}$: Sender and receiver parameters are initialized with 0 starting values.

Edge specific covariates, either constant or variable across the networks can be easily incorporated in the proposed model 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 presented in the supplementary materials refer to the most general case of missing data and edge-specific covariates.

4 Undirected networks

In the particular case of undirected networks, out-degrees and in-degrees are identical for each node, $\sum_i y_{ij}^{(k)} = \sum_i y_{ji}^{(k)} \forall i = 1, \dots, n$. The framework proposed in section 2 can be easily modified to deal with undirected multidimensional networks by imposing the constrain: $\theta_i^{(k)} = \gamma_i^{(k)} = \delta_i^{(k)}$. According to such assumption, the edge probability equation (4) can be rewritten as:

$$P(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, \delta_i^{(k)}, \delta_j^{(k)}, d_{ij}) = \frac{\exp\{\alpha^{(k)}g(\delta_i^{(k)}, \delta_j^{(k)}) - \beta^{(k)}d_{ij}\}}{1 + \exp\{\alpha^{(k)}g(\delta_i^{(k)}, \delta_j^{(k)}) - \beta^{(k)}d_{ij}\}} \quad (6)$$

If not null, the effect $\delta_i^{(k)}$ can be either variable across the different networks, $\delta_i^{(k)}$, or constant, $\delta_i^{(k)} = \delta_i$, for $k = 1, \dots, K$ and $i = 1, \dots, n$. We provide the reader with the proposal distributions used to estimate these model parameters in the supplementary material.

5 Simulations

We defined a simulation study to evaluate the performance of the proposed estimation procedure, with a particular focus on the behaviour of model parameter estimates for models NC, NV, CC, CV and VV. These five models are chosen to investigate the properties of the estimators when the true scenarios refer to different numbers of parameters. For each of the simulated multiplexes (10 for each model), we fit the “true” model, that is, the model a given multiplex was simulated from. A first scenario, **B1**, refers to multiplexes with size

$n = 50, K = 5$. A second larger one, **B2**, has size similar to those of the multiplex that we analyse in the application, see section 6, namely $n = 50, K = 10$.

In both these simulation scenarios, we set $\alpha^{(1)} = 2$ and $\beta^{(1)} = 1$. The prior parameters are $\nu_\alpha = \nu_\beta = 3$, $m_\alpha = 2$, $m_\beta = 0$, $\tau_\alpha = \tau_\beta = (K - 1)/K$. The dimension of the latent space is set to $p = 2$ for parsimony and graphical analysis purposes, as it is often the case with latent space models. Further work in this field is needed to develop criteria for the choice of p , to better drive the decision on such parameter. Small variations of these values have been also tried and did not affect the results. Also, there are no missing edges in the simulated data. The MCMC algorithm run for 800000 iterations with a burn in of 15000. MCMC diagnostics are reported in the Appendix. Table 2 shows the average values of the distance correlation between the true and the estimated edge probabilities, for all the different networks composing the multiplexes in scenarios **B1** and **B2** and the average AUC values (Area Under the Curve). 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 2 reports the average values of the Procrustes correlation between the simulated and the estimated latent space coordinates. As for the edge probabilities, we notice 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 choice is used as we are mainly interested in recovering the nodes ordering, with respect to the effect estimates. Indeed, sender and receiver parameters vary in the relatively small interval $(-1, 1)$, and the exact numerical values may not be of interest. Table 3 reports, for each network in the simulated multiplexes, the average values of the Spearman correlation coefficient between the simulated and the estimated receiver effect, $\gamma_i^{(k)}$. These values are always much greater than 0.5, with a couple of exceptions for some networks drawn from models with a higher complexity, CV and VV. However, as we comment in Table 2, this does not impact the recovering of edge probabilities. Table 4 reports the average values of the Spearman correlation coefficient between the simulated and the estimated sender effect, $\theta_i^{(k)}$. The behaviour of these estimates complies with those of the receiver effect previously discussed, proving that the effects, when both present, are estimated with similar quality.

We refer to the supplementary materials for the $\alpha^{(k)}$ and the $\beta^{(k)}$ estimates. The intercepts are recovered within a 95% credible interval, with two exceptions which occur when the simulated values are “extreme”. However, the ordering between the different intercepts in a given multiplex is always well 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 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 for all the parameters are quite hard to recover, due to the large number of parameters in the models and the “multiplicative” parametrization adopted. 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. All simulations are ran on a Intel(R) Core(TM) i7-8565U @1.80GHz computer and we refer to the supplementary materials for information on the computational times. The supplement also provides information on Average Expected Sample Size (ESS) and acceptance probability values for model parameters and latent coordinates. Although some of the acceptance probability values are

small, the simulation study presented in this Section showed that parameters and latent coordinates are still quite well recovered using the proposed algorithm, under all different sender/receiver effect specifications.

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	AUC	PC
B1	NC	0.86	0.84	0.86	0.87	0.85	-	-	-	-	-	0.83	0.94
	NV	0.87	0.87	0.87	0.85	0.86	-	-	-	-	-	0.82	0.92
	CC	0.89	0.89	0.90	0.91	0.87	-	-	-	-	-	0.80	0.96
	CV	0.84	0.81	0.83	0.84	0.85	-	-	-	-	-	0.79	0.90
	VV	0.83	0.81	0.80	0.81	0.81	-	-	-	-	-	0.77	0.93
B2	NC	0.89	0.84	0.83	0.83	0.81	0.92	0.84	0.81	0.81	0.82	0.87	0.85
	NV	0.83	0.80	0.79	0.76	0.81	0.78	0.88	0.71	0.75	0.75	0.84	0.93
	CC	0.83	0.87	0.91	0.84	0.90	0.91	0.91	0.84	0.86	0.91	0.78	0.90
	CV	0.75	0.86	0.87	0.76	0.86	0.88	0.87	0.86	0.83	0.87	0.85	0.91
	VV	0.73	0.72	0.82	0.85	0.88	0.79	0.79	0.79	0.86	0.84	0.76	0.90

Table 2: *Simulation study. Average distance correlation between the simulated and the estimated edge-probabilities. Columns AUC and PC show, respectively, the average AUC (Area Under the Curve) value and the average Procrustes correlation between the simulated and the estimated latent space coordinates. Standard deviations associated with the different estimates were always lower than 0.10.*

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
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: *Simulation study. Average Spearman correlation index between the simulated and the estimated receiver effects. Standard deviations associated with the different estimates were always lower than 0.07, with the exception of the VV model, where they were lower than 0.12.*

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
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 4: *Simulation study. Average Spearman correlation between the simulated and the estimated sender effects, by simulation scenario and true model structure. Standard deviations associated with the different estimates were always lower than 0.07, with the exception of the VV model, where they were lower than 0.12.*

5.1 An heuristic procedure for model selection

In Section 2 we have proposed a class of models for multidimensional networks, with 9 different specifications for the sender/receiver effects. In general, the issue of model selection can be addressed in two different ways. A first approach is that of an expert who, based on some prior knowledge on the data, suggests which particular model should be used. A second, more common, approach, is that of choosing the “best” model using some selection criteria. In the present context, the estimation of the nine models, on a specific observed multiplex may require substantial (computational) time, especially when the size of the multiplex is large. Hence, it could be convenient 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 some summary statistics calculated on the observed multidimensional network data. The idea of using summary statistics to aid the model selection was introduced by [Hunter et al. \(2008\)](#). In that work, the authors proposed a graphical goodness of fit procedure to compare structural statistics of the observed network with the corresponding statistics on networks simulated from a given model. Although this procedure allows to compare different models without the need to fit them, it relies on graphical comparisons, which may be subjective or difficult when the number of models to compare is large. A more automated model selection procedure is discussed by [Pudlo et al. \(2015\)](#) in the context of ABC algorithms. There, model selection is rephrased as a classification problem, with random forests used to predict, prior to fitting, which model would be the most appropriate for the data.

We propose to summarize observed multiplex data using in-degrees and out-degrees correlations. The general idea is that of computing the correlations among out-degrees/in-degrees. This 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 denotes the number of nodes and K the number of networks. Then, the matrices c_{s_k} and c_{r_k} , of dimension $K \times K$, contain the values of the correlation between the sender/receiver effects across views. Instead, the matrices c_{s_i} and c_{r_i} , of dimension $n \times n$, include the correlations between the sender/receiver effects across 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 use such quantities to choose which type of model has to be estimated, among those proposed in Section 2. The idea is that 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 different. 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 effects, that is, null and variable. The variability between nodes correlations (for which the standard deviations serve as proxies) may be used to discriminate between different model structures, referring to null, constant or variable node-specific effects. Formally, the proposed procedure works as described below. To study the behaviour of this approach, given an observed multidimensional network with n nodes and K views, we propose to simulate T multiplexes from equation 4, for each one of the nine models. For each simulated dataset, the correlation summary statistics $(\bar{c}_{s_k}, \bar{c}_{r_k}, \bar{c}_{s_i}, \bar{c}_{r_i}, sd(c_{s_k}), sd(c_{r_k}), sd(c_{s_i}), sd(c_{r_i}))$ are computed. Such summary statistics are later employed to train a Linear Discriminant Analysis (LDA) classifier, using *mclust* (Scrucca et al. 2016). The proposed procedure has been tested for multidimensional networks with $n = (50, 75, 100)$ nodes and $K = (3, 5, 10)$ networks, with $T = 5000$. Table 5 reports the cross validation error for the classifier, for different n and K values. The classification errors are quite low in all scenarios, and they decrease with increasing K . Indeed, having multiple replicates of the networks may help distinguishing the type of network/node-specific effects. To further test the proposed procedure, we have used the estimated classifier to predict the model-type of other 5000 multidimensional networks, simulated independently from those used to train the classifier. Table 6 reports the average accuracy of the classifier on test data over different (n, K) values, for each one of the models. The accuracy values on the diagonal show that the proposed procedure has a larger discriminant power among the competing models. Also, when it fails to select the right model, it proposes a model that is “near” to the “true” one. For example, when multiplexes are simulated from a *VV* model, the procedure recovers the true model 96% of the times, the *CC* model 2% and the *CV-VC* models 2% of the times. Using such an heuristic procedure to choose among the models can help reducing drastically the computing effort. Moreover, the proposed framework allows the implementation of dimension-specific classifier, as the scheme and the corresponding classification directly depend on the size of the observed multidimensional network, n, K .

$n = 50$			$n = 75$			$n = 100$		
$K = 3$	$K = 5$	$K = 10$	$K = 3$	$K = 5$	$K = 10$	$K = 3$	$K = 5$	$K = 10$
0.152	0.059	0.024	0.076	0.030	0.016	0.051	0.023	0.013

Table 5: Cross validation error for the classifier, for different n and K values.

		Predicted								
		NN	CN	NC	CC	VN	NV	VC	CV	VV
Class	NN	93 (2) %	1 (0) %	0 (0) %	6 (1) %	0 (0) %	0 (0) %	0 (0) %	0 (0) %	0 (0) %
	CN	0 (0) %	97 (4) %	0 (0) %	2 (2) %	0 (0) %	0 (0) %	0 (0) %	1 (1) %	0 (0) %
	NC	0 (0) %	0 (0) %	97 (3) %	2 (2) %	0 (0) %	0 (0) %	1 (1) %	0 (0) %	0 (0) %
	CC	6 (2) %	1 (1) %	1 (1) %	88 (7) %	0 (0) %	0 (0) %	1 (1) %	1 (2) %	2 (1) %
	VN	0 (0) %	0 (0) %	0 (0) %	0 (0) %	97 (3) %	0 (0) %	3 (2) %	0 (0) %	0 (0) %
	NV	0 (0) %	0 (0) %	0 (0) %	0 (0) %	0 (0) %	97 (3) %	0 (0) %	3 (2) %	0 (1) %
	VC	0 (0) %	0 (0) %	0 (1) %	1 (2) %	3 (2) %	0 (0) %	95 (6) %	0 (0) %	1 (2) %
	CV	0 (0) %	0 (1) %	0 (0) %	1 (2) %	0 (0) %	3 (2) %	0 (0) %	95 (6) %	1 (2) %
	VV	0 (0) %	0 (0) %	0 (0) %	2 (1) %	0 (0) %	0 (0) %	1 (2) %	1 (2) %	96 (5) %

Table 6: Average accuracy of the classifier on the test data over different (n, K) values, for each one of the nine models. Standard deviations are reported in brackets.

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 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 node-specific effects, 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 (countries with no links) and to guarantee an easy and feasible representation of the results, we focus on a subgroup of 64 countries, see the supplementary materials. To define such a sub-sample, 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 $n = 64$ 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 associations¹ 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. As the observed out-degrees and in-degrees present a strong association (see the Supplementary materials), the data are a good candidate to test the proposed model. The heuristic procedure described in Section 5.1 suggested to use the CN model with probability 0.98, and the *VC* and *CN* models both with probability 0.01. Hence, we fit a model with constant sender effects, using the MCMC algorithm described in section 3.2, where the number of iterations and the hyper parameters are fixed as in the simulation setting (Section 5). We set $p = 2$, the dimension of the latent space, both for graphical reasons and to compare the estimated with the geographical coordinates. The choice of the number of dimensions in the latent spaces is further validated through AUC for out-of-sample edge prediction. More in details, we fit CN models with different number of latent dimensions to the FAO multiplex data and compute median AUC values for out-of-sample edge prediction (across different views of the FAO multiplex). Such values are $\{0.87, 0.87, 0.86\}$, respectively for $p = \{2, 3, 4\}$. These results suggest that the choice of $p = 2$ can be appropriate for the data at hand as increasing the value of p only increments model complexity, not predictions accuracy. Moreover, Spearman correlation coefficient values computed between couples of estimated sender effects for these models range in $(0.94, 0.97)$, showing stability with respect to different latent space dimensions specifications. Indeed, the proposed model is able to disentangle between reciprocal relations between the nodes, captured by the latent space, and node-specific ones.

Figure 2(left plot) reports the estimated sender effects and the observed mean out-degrees, for the 64 countries considered. Countries corresponding to high estimated sender effects may be considered as “top exporters”. “Top exporting” countries are to be interpreted as those countries that tend to export fruit to a large group of trading partners, conditional on the latent distances to other nodes. Three out of the first four “top exporting” countries are European: Italy, Netherland and Spain. Netherlands (NLD) 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)(FreshPlaza 2015). Contrary to Netherlands, Spain and Italy directly grow most of the fruits they export. Just to give an example, Canary Islands are major pineapples producers. Also, Spain in 2017 became the world’s largest watermelon exporting country (FreshPlaza 2018). These three European countries show good agreement between the estimated sender effects and the observed mean out-degrees, see Figure 2(left plot). On the contrary, New Zealand, which is estimated to be the third “top exporter”, has a moderate mean out-degree. Such mismatch between New Zealand’s estimated sender effect and mean-out degree may be explained looking at its position in the latent space, in Figure 2(right plot). Indeed, New Zealand is estimated to be quite far from the countries it trades with, therefore the “residual” contribution given by the sender effect to the edge probability needs to be really high. In other words, a large value of the sender effect accounts for the “distance” New Zealand has to travel to export its goods. The estimated latent coordinates for the 64 countries are presented in Figure 3, which additionally displays the geographic coordinates of such countries, for comparison. The estimated latent coordinates do not match the geographical ones; indeed, the Procrustes correlation value between the two sets is

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

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

not high (0.53). The estimated latent space is characterized by a large number of Asian countries placed in the right hand side of the space and the majority of the European countries in the left hand part. United States are placed in between Asiatic and Oceanic countries. Indeed, most of these countries are leading suppliers of fruits for the United States, thanks to established or pending free trade agreements (Johnson 2016). Comparing the estimated distances with the number and types of reciprocal ties² exchanged between the nodes, we find that the estimated latent space well depicts symmetric relations in the networks. Indeed, nodes that often exchange reciprocal ties are placed very close to each other in the latent space, while those that hardly ever connect are the ones furthest apart from one another. For more details on the matter, we refer to the supplementary materials. Moreover, the latent space is estimated to be always relevant in the determination of the edge probabilities, and to have similar effect in the different networks. Indeed, Table 7 shows that the estimated coefficients $\beta^{(k)}$ range in (0.76, 2.52). The same Table displays the estimates of the intercepts in the different networks, with all of the intercepts corresponding, at baseline $d_{ij} = 0$, to high edge probability values (greater than 0.9). Figure 4 represents the estimated probabilities in the fruit networks, given some values of the distances and of the sender effect. In particular, the probabilities are computed for the first (0.385), second (1.274) and third (2.281) quartiles of the distances, and for $\theta_i = (-0.5, 0, 0.5, 1)$. The plot in Figure 4 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 sender effect and the network-specific parameters, $\alpha^{(k)}$ and $\beta^{(k)}$.

		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
α	mean	2.00	1.96	2.51	2.49	2.63	1.80	2.26	2.43	1.78	2.59
	sd	-	0.12	0.13	0.15	0.12	0.12	0.13	0.14	0.15	0.14
β	mean	1.00	1.37	0.76	1.04	1.12	1.70	1.46	1.13	2.52	1.42
	sd	-	0.06	0.05	0.05	0.07	0.08	0.07	0.06	0.12	0.07

Table 7: Averages and standard deviations of the estimated posterior distributions for the intercept and the scale coefficient parameters in the fruit networks.

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. The shared latent space assumption can be quite appealing in the presence of some sort of “homogeneity” in the type of relations collected in a multiplex, for example with countries trading similar goods, as it allows to easily summarize multiple views and visualize similarities in the data. Note however that such assumption may not be appropriate for multidimensional networks formed by views inhomogeneous in the relation types, as for example prey-predator networks or networks depicting opposite relations. The proposed models should

²For each network k and for each pair of nodes (i, j) , we can classify ties as: reciprocal and null if $(y_{ij}^{(k)} = y_{ji}^{(k)} = 0)$, asymmetric if $(y_{ij}^{(k)} \neq y_{ji}^{(k)})$ and reciprocal and positive if $(y_{ij}^{(k)} = y_{ji}^{(k)} = 1)$.

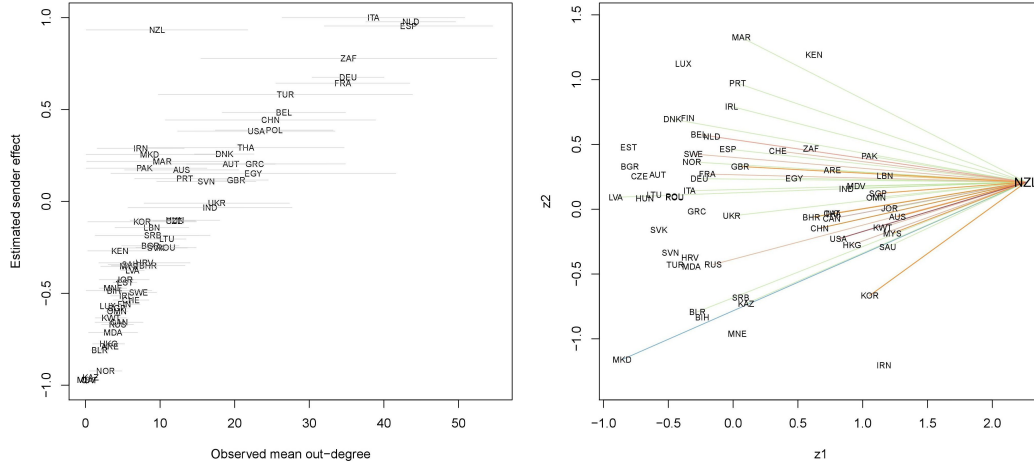


Figure 2: Fruit multiplex. Estimated sender parameters and observed mean out-degrees. The grey lines represent out-degree standard deviations (left figure). Estimated latent space. The segments link New Zealand to the countries it exports to, with different colors corresponding to different fruit markets (right figure).

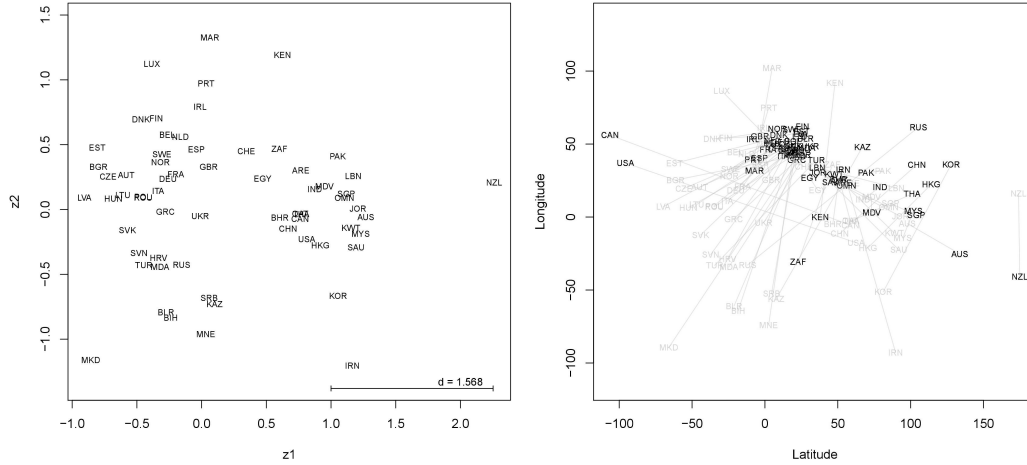


Figure 3: Fruit multiplex. Estimated latent coordinates. The segment represented at the bottom of the plot displays the median value of the estimated distances distribution. The average standard deviation associated with each coordinate is 0.11 (left figure). Geographical coordinates. Estimated latent coordinates are reported in grey in the background, for comparison purposes (right figure).

be extended to account for such types of multiplex data. An example of a feasible extension would rely on clusters of “homogeneous” networks (in terms of type of relation), each of which could be then represented by the same shared 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

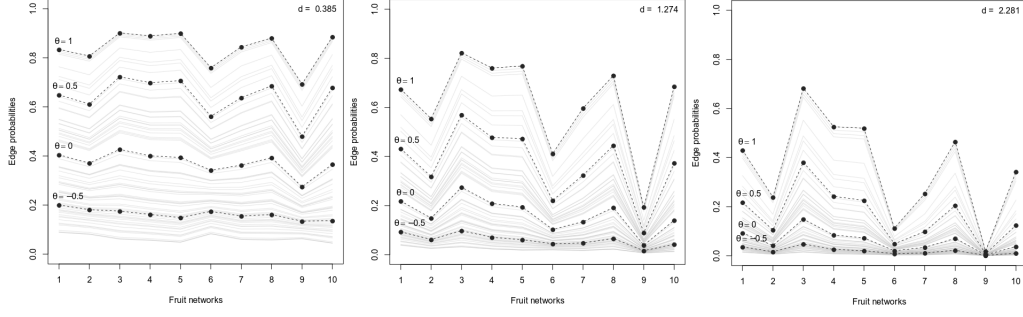


Figure 4: Fruit networks. Estimated probabilities in the multiplex, for different values of the estimated distances (first quartile, median and third quartile) and for $\theta_i = (-0.5, 0, 0.5, 1)$. The grey lines in the background represent the country-specific edge probabilities, obtained considering the estimated sender parameters.

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 $(1, 3)$. 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 allows to bypass a model selection step. A prior 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.

Further, the models were designed to disentangle symmetric and asymmetric relationships in the networks, where these two types of structure are captured respectively by the latent space and the node-specific effects. An illustrative application to FAO trade data regarding different fruit trades has been presented, where the method was able to uncover trade patterns and shared similarities among different fruit markets. Pairs of nodes strongly exhibiting positive reciprocity in the multiplex resulted to be those placed closer to each other in the latent space, showing that the latter is able to capture symmetric structures in the data. Moreover, the estimated node-specific effect values were shown to be consistent under different latent space dimensions specifications, proving that the proposed model is able to well separate asymmetric from symmetric relational structures. 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 and the heuristic model selection procedure are incorporated in the *R* package *spaceNet* (D'Angelo & Fop 2018), available on *CRAN*.

References

CBI (2015), ‘Fresh fruit and vegetables in Europe’.

URL: https://www.cbi.eu/sites/default/files/market_information/researches/trade-statistics-europe-fresh-fruit-vegetables-2015.pdf

D’Angelo, S. & Fop, M. (2018), *spaceNet: Latent Space Models for Multidimensional Networks*. R package version 1.2.

URL: <https://CRAN.R-project.org/package=spaceNet>

D’Angelo, S., Murphy, T. & Alfö, M. (2018), ‘Latent space modeling of multidimensional networks with application to the exchange of votes in Eurovision Song Contest’, *Annals of applied statistics (to appear)*. *arXiv:1803.07166 [stat.AP]*.

Durante, D. & Dunson, D. (2016), ‘Locally adaptive dynamic networks’, *Annals of Applied Statistics* **10**(4), 2203–2232.

Durante, D., Dunson, D. & Vogelstein, J. (2017), ‘Nonparametric Bayes modeling of populations of networks’, *Journal of the American Statistical Association* **112**(520), 1516–1530.

Durante, D., Mukherjee, N. & Steorts, R. (2017), ‘Bayesian learning of dynamic multilayer networks’, *Journal of Machine Learning Research* **18**, 1–29.

FAOStat (2013), ‘Food and agriculture data’.

URL: <http://www.fao.org/faostat/en>

Fosdick, B. & Hoff, P. (2014), ‘Separable factor analysis with applications to mortality data’, *The Annals of Applied Statistics* **8**(1), 120–147.

FreshPlaza (2015), ‘Netherlands world’s biggest (re-)exporter for 11 fruit and veg products’.

URL: [http://www.freshplaza.com/article/135083/Netherlands-worlds-biggest-\(re-\)exporter-for-11-fruit-and-veg-products](http://www.freshplaza.com/article/135083/Netherlands-worlds-biggest-(re-)exporter-for-11-fruit-and-veg-products)

FreshPlaza (2018), ‘Spain became the world’s largest watermelon exporter in 2017’.

URL: <http://www.freshplaza.com/article/196648/Spain-became-the-worlds-largest-watermelon-exporter-in-2017>

Gollini, I. & Murphy, T. (2016), ‘Joint modeling of multiple network views’, *Journal of Computational and Graphical Statistics* **25**(1), 246–265.

Hoff, P. (2003), Random effects models for network data, in ‘Dynamic social network modeling and analysis’, Vol. 126, Breiger, R. and Carley, K. and Pattinson, P. (Eds.), pp. 302–322.

Hoff, P. (2005), ‘Bilinear mixed-effects models for dyadic data’, *Journal of the American Statistical Association* **100**(469), 286–295.

Hoff, P. (2011), ‘Hierarchical multilinear models for multiway data’, *Computational Statistics and Data Analysis* **55**, 530–543.

- Hoff, P. (2015), ‘Multilinear tensor regression for longitudinal relational data’, *The Annals of Applied Statistics* **9**(3), 1169–1193.
- Hoff, P., Raftery, A. & Handcock, M. (2002), ‘Latent space approaches to social network analysis’, *Journal of the American Statistical Association* **97**(460), 1090–1098.
- Holland, P. & Leinhardt, S. (1981), ‘An exponential family of probability distributions for directed graphs’, *Journal of the American Statistical Association* **76**(373), 33–50.
- Hunter, D., Goodreau, S. & Handcock, M. (2008), ‘Goodness of fit of social network models’, *Journal of the American Statistical Association* **103**(481), 248–258.
- Johnson, R. (2016), ‘The U.S. trade situation for fruit and vegetable products’.
URL: <https://fas.org/sgp/crs/misc/RL34468.pdf>
- Krivitsky, P., Handcock, M., Raftery, A. & Hoff, P. (2009), ‘Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models’, *Social Networks* **31**(3), 204–213.
- Pudlo, P., Marin1, J., Estoup, A., Cornuet, J., Gautier, M. & Robert, C. (2015), ‘Reliable ABC model choice via random forests’, *Bioinformatics* **32**(6), 859–866.
- Salter-Townshend, M. & McCormick, T. (2017), ‘Latent space models for multiview network data’, *Annals of Applied Statistics* **11**(3), 1217–1244.
- Schweinberger, M. & Snijders, T. (2003), ‘Settings in social networks: A measurement model’, *Sociological Methodology* **33**, 307–341.
- Scrucca, L., Fop, L., Murphy, T. & Raftery, A. (2016), ‘mclust 5: clustering, classification and density estimation using Gaussian finite mixture models’, *The R Journal* **8**(1), 205–233.
- Sewell, D. & Chen, Y. (2015), ‘Latent space models for dynamic networks’, *Journal of the American Statistical Association* **110**(512), 1646–1657.
- Sewell, D. & Chen, Y. (2016), ‘Latent space models for dynamic networks with weighted edges’, *Social Networks* **44**, 105–116.
- Signorelli, M. & Wit, E. (2020), ‘Model-based clustering for populations of networks’, *Statistical Modelling (in press)* .
- van Duijn, M., Snijders, T. & Zijlstra, B. (2004), ‘ p_2 : a random effects model with covariates for directed graphs’, *Statistica Neerlandica* **58**, 234–254.
- Zhou, J., Bhattacharya, A., Herring, A. & Dunson, D. (2015), ‘Bayesian factorization of big sparse tensors’, *Journal of the American Statistical Association* **110**(512), 1562–1576.

A 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 8: Fao data: country ISO3 codes.



Figure 5: 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 6 represents the boxplots of the estimated distances in the latent space and the median reciprocity values for couples of nodes in the FAO multiplex. Given an adjacency matrix $Y^{(k)}$, we define the reciprocity index for the couple of nodes (i, j) as:

$$rec_{i,j} = y_{ij}^{(k)} + y_{ji}^{(k)} = \begin{cases} 0 & \text{if the relation between the nodes is symmetric and null} \\ 1 & \text{if the relation between the nodes is asymmetric} \\ 2 & \text{if the relation between the nodes is symmetric and positive} \end{cases}$$

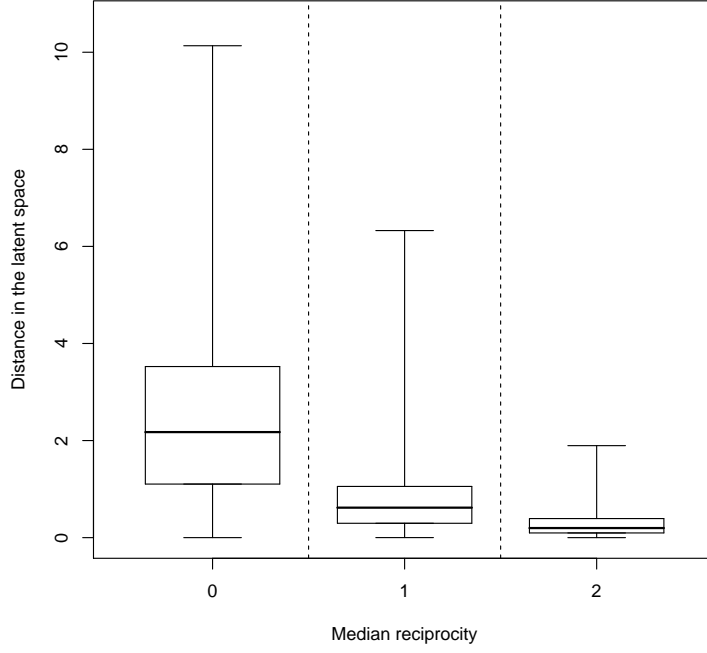


Figure 6: FAO multidimensional network. Boxplots of the estimated distances in the latent space and the median reciprocity values for couples of nodes.

Given the reciprocity index definition, we can see that the estimated latent space well captures symmetric relations between nodes in different views of the FAO multidimensional network. Indeed, nodes that often exchange reciprocal ties, $\text{median}(\text{rec}_{i,j}) = 2$, are placed very close to each other in the latent space, while those that hardly ever connect, $\text{median}(\text{rec}_{i,j}) = 0$, are the ones furthest apart from one another.

B Estimation: proposal and full conditional distributions

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

$$\begin{aligned}
\log\left(P(\alpha, \beta, \theta, \gamma, \mathbf{z}, \mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2 | \mathbf{Y})\right) &\propto \\
&\sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i} h_{ij}^{(k)} \left[y_{ij}^{(k)} (\alpha^{(k)} \phi_{ij}^{(k)} - \beta^{(k)} d_{ij} - \sum_{f=1}^F \lambda_f x_{ijf}) - \log\left(1 + \exp\{\alpha^{(k)} \phi_{ij}^{(k)} - \beta^{(k)} d_{ij} - \sum_{f=1}^F \lambda_f x_{ijf}\}\right) \right] \\
&- \frac{1}{2} \left\{ \sum_{i=1}^n z_i^2 + \frac{\sum_{k=1}^K (\alpha^{(k)} - \mu_\alpha)^2}{\sigma_\alpha^2} + \frac{\sum_{k=1}^K (\beta^{(k)} - \mu_\beta)^2}{\sigma_\beta^2} + K \log(\sigma_\alpha^2) + K \log(\sigma_\beta^2) + \log(\tau_\alpha \sigma_\alpha^2 \tau_\beta \sigma_\beta^2) \right. \\
&+ \frac{\mu_\alpha^2}{\tau_\alpha \sigma_\alpha^2} + \frac{\mu_\beta^2}{\tau_\beta \sigma_\beta^2} + \frac{1}{\sigma_\alpha^2} + \frac{1}{\sigma_\beta^2} + \sum_{f=1}^F \left(\frac{(\lambda_f - \mu_{\lambda_f})^2}{\sigma_{\lambda_f}^2} + \log(\sigma_{\lambda_f}^2) + \log(\tau_\lambda \sigma_{\lambda_f}^2) + \frac{(\mu_{\lambda_f} - \mu_\lambda)^2}{\tau_\lambda \sigma_{\lambda_f}^2} + \frac{1}{\sigma_{\lambda_f}^2} \right) \Big\} \\
&+ \left(-\frac{\nu_\alpha}{2} - 1 \right) \log(\sigma_\alpha^2) + \left(-\frac{\nu_\beta}{2} - 1 \right) \log(\sigma_\beta^2) + \sum_{f=1}^F \left(-\frac{\nu_{\lambda_f}}{2} - 1 \right) \log(\sigma_{\lambda_f}^2),
\end{aligned} \tag{7}$$

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

B.1 Nuisance parameters

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

$$\sigma_\alpha^2 | \alpha, \mu_\alpha, \tau_\alpha, \nu_\alpha, K \sim \text{Inv}\Gamma(r_\alpha, R_\alpha); \quad \sigma_\beta^2 | \beta, \mu_\beta, \tau_\beta, \nu_\beta, K \sim \text{Inv}\Gamma(r_\beta, R_\beta),$$

with parameters:

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

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

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

B.2 Latent positions

The proposal distribution for the i^{th} latent coordinate is derived from the log-posterior distribution of the model in equation 7. The logarithmic term in the log-likelihood, which is indeed a LSE function, is approximated with its lower bound.

$$\tilde{z}_i | \mathbf{Y}, \alpha, \theta_i, \gamma, \beta, \lambda, \mathbf{x}_i, \mathbf{D}, K \sim N(\mu_{\tilde{z}_i}, \sigma_{\tilde{z}_i}^2),$$

where

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

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

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

B.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:

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

with

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

B.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:

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

where

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

B.5 Coefficient parameters (covariates)

The proposal distribution for λ_f is:

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

where

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

B.6 Sender and receiver parameters

To define proposal distributions for the sender and receiver parameter, we consider only the relevant parts of the posterior distribution. In particular, we derive them from the log-likelihood of the model, that is, the first line of the log-posterior distribution presented in equation 7,

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

B.6.1 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 $\theta_i^{(k)t} \approx \theta_i^{(k)t-1}$:

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

with $x = 1$ when only one type of effect is present and $x = 0.5$ when both types are.

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

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

where

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

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

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

where

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

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.

B.6.2 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, trivial, 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:

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

where

$$\begin{aligned} \mu_{\tilde{\gamma}_j^{(k)t}} &= \left[\alpha^{(k)} x \sum_{i \neq j} h_{ij}^{(k)} \left(y_{ij}^{(k)} - \frac{\exp(\alpha^{(k)} \phi_{ij}^{(k)t-1} - \beta^{(k)} d_{ij})}{1 + \exp(\alpha^{(k)} \phi_{ij}^{(k)t-1} - \beta^{(k)} d_{ij})} \right) \right] \sigma_{\tilde{\gamma}_j^{(k)t}}^2 + \gamma_j^{(k)t-1} \\ \sigma_{\tilde{\gamma}_j^{(k)t}}^2 &= \left[\left(\alpha^{(k)} x \right)^2 \sum_{i \neq j} h_{ij}^{(k)} \left(\frac{\exp(\alpha^{(k)} \phi_{ij}^{(k)t-1} - \beta^{(k)} d_{ij})}{\left(1 + \exp(\alpha^{(k)} \phi_{ij}^{(k)t-1} - \beta^{(k)} d_{ij}) \right)^2} \right) \right]^{-1}, \end{aligned}$$

and if the effect is constant:

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

where

$$\begin{aligned}\mu_{\tilde{\gamma}_j^t} &= \left[\sum_{k=1}^K \alpha^{(k)} x \sum_{i \neq j} h_{ij}^{(k)} \left(y_{ij}^{(k)} - \frac{\exp(\alpha^{(k)} \phi_{ij}^{t-1} - \beta^{(k)} d_{ij})}{1 + \exp(\alpha^{(k)} \phi_{ij}^{t-1} - \beta^{(k)} d_{ij})} \right) \right] \sigma_{\tilde{\gamma}_j^t}^2 + \gamma_j^{t-1} \\ \sigma_{\tilde{\gamma}_j^t}^2 &= \left[\sum_{k=1}^K \left(\alpha^{(k)} x \right)^2 \sum_{i \neq j} h_{ij}^{(k)} \left(\frac{\exp(\alpha^{(k)} \phi_{ij}^{t-1} - \beta^{(k)} d_{ij})}{\left(1 + \exp(\alpha^{(k)} \phi_{ij}^{t-1} - \beta^{(k)} d_{ij}) \right)^2} \right) \right]^{-1}.\end{aligned}$$

B.6.3 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 (see the paper), 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:

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

where

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

and if the effect is constant:

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

where

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

C Scenario I: simulation results

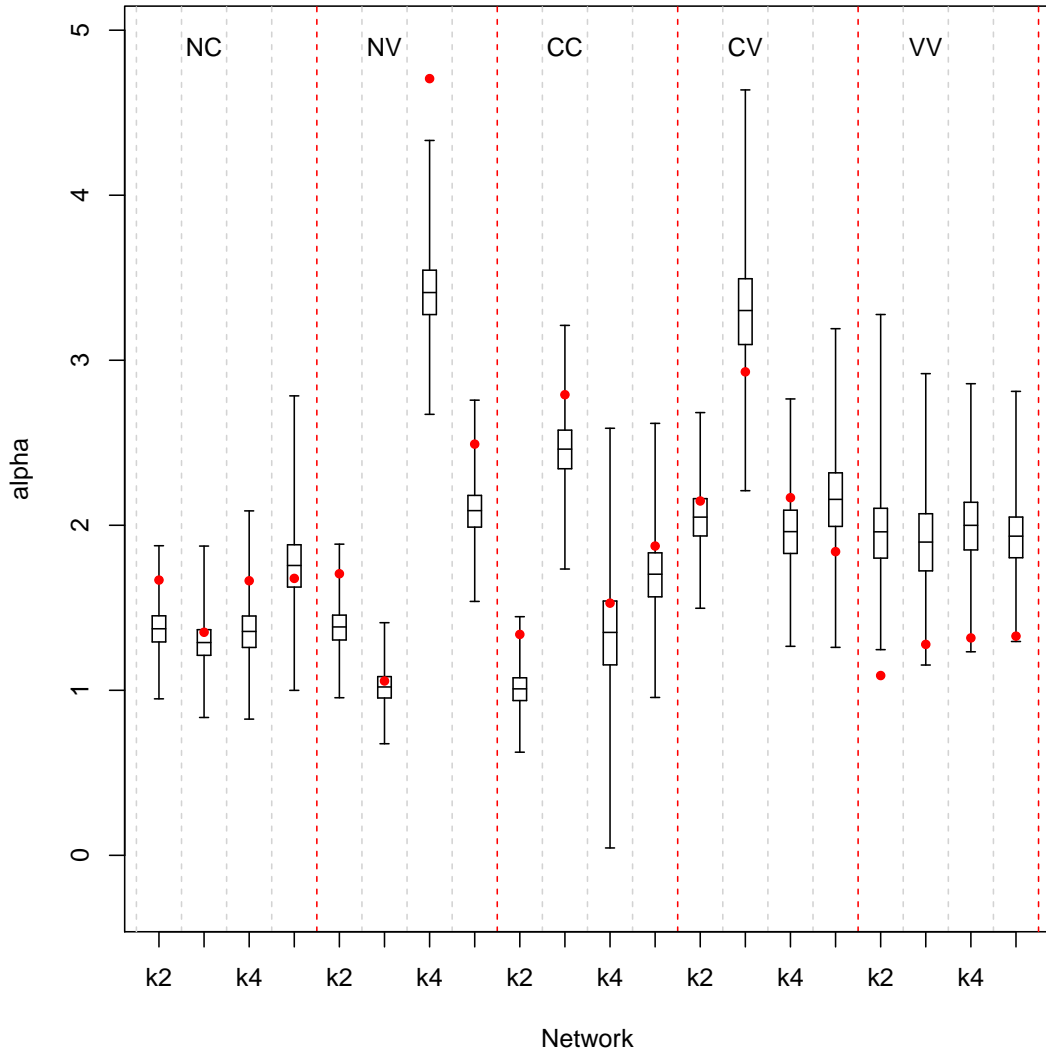


Figure 7: Boxplots of the estimated posterior distributions for the intercept parameters $\alpha^{(k)}$. Red dots indicate the true, simulated, values of the intercepts.

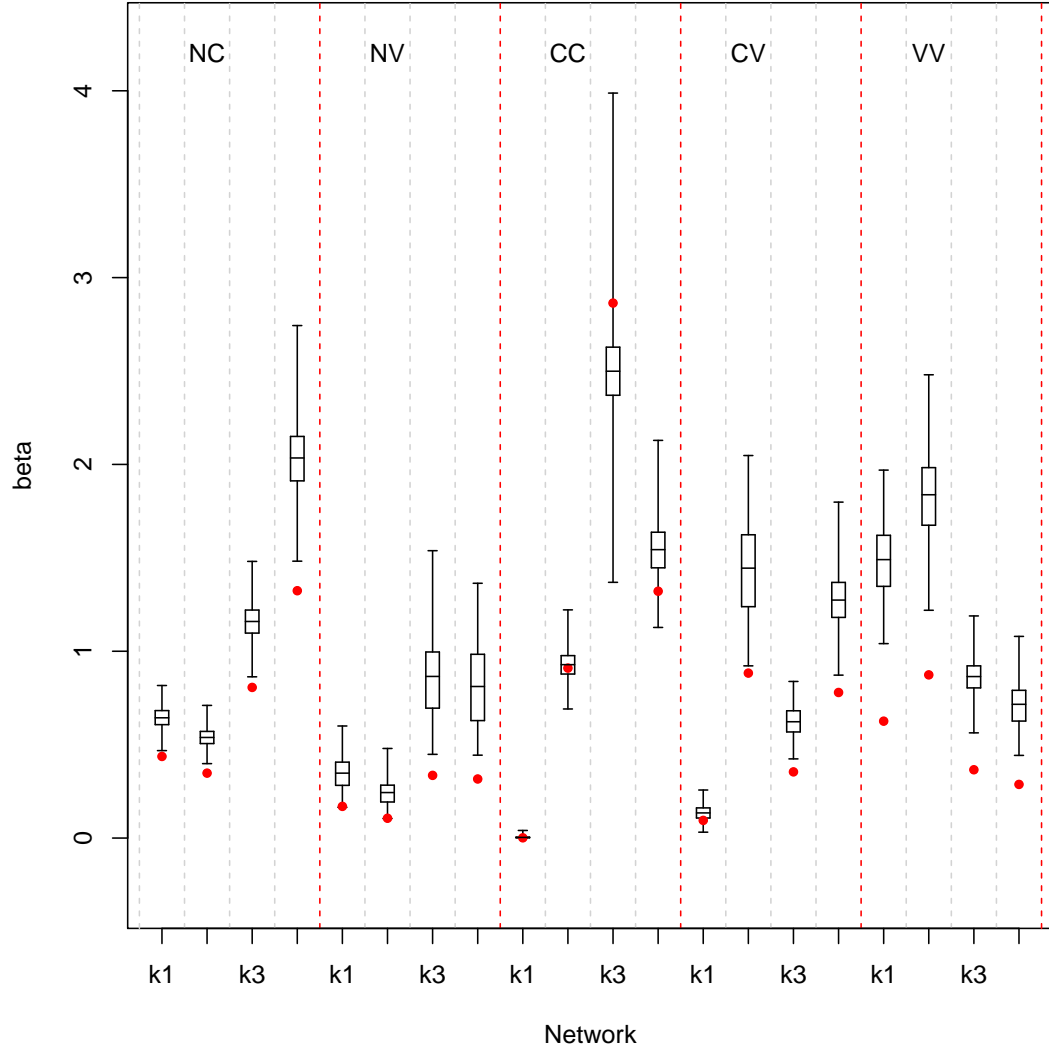


Figure 8: Boxplots of the estimated posterior distributions for the coefficient parameters $\beta^{(k)}$. Red dots indicate the true, simulated, values of the coefficients.

D MCMC diagnostics and computational times

		$\overline{ESS(\alpha)}$	$\overline{ESS(\beta)}$	$\overline{ESS(\gamma)}$	$\overline{ESS(\theta)}$	$\overline{ESS(z)}$	$\overline{AP(\alpha)}$	$\overline{AP(\beta)}$	$\overline{AP(\gamma)}$	$\overline{AP(\theta)}$	$\overline{AP(z)}$
B1	NC	1682	1588	9646	-	8005	0.04	0.04	0.80	-	0.05
	NV	1382	1388	21701	-	3045	0.03	0.03	0.91	-	0.04
	CC	1328	1245	2824	2892	2332	0.04	0.04	0.93	0.93	0.04
	CV	993	1007	6809	6712	17725	0.04	0.04	0.92	0.90	0.06
	VV	971	983	7497	7464	34506	0.05	0.05	0.96	0.95	0.04
B2	NC	4178	4695	16311	-	70864	0.09	0.09	0.73	-	0.07
	NV	1268	1492	24592	-	11013	0.05	0.05	0.92	-	0.04
	CC	1301	989	1202	1362	8499	0.04	0.04	0.93	0.93	0.04
	CV	3051	3038	7678	8127	93389	0.10	0.10	0.97	0.94	0.07
	VV	1023	1336	6438	57616	59805	0.06	0.06	0.98	0.96	0.04
FAO data	CN	3213	2988	-	15372	42827	0.03	0.03	-	0.76	0.04

Table 9: Average ESS (Expected Sample Size) and AP (acceptance probability) values for the model parameters and the latent coordinates in the simulation study and the FAO data application.

		$K = 5$		$K = 10$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$
NC	30.17	218.71	71.32	449.19	
NV	49.66	357.21	95.08	480.17	
CC	50.30	381.23	96.35	498.36	
CV	51.17	392.08	98.78	536.99	
VV	52.13	398.75	105.48	573.57	

Table 10: Average computational times in seconds for 10 iterations of the proposed MCMC algorithm for multidimensional network data with $n = (50, 100)$ and $K = (5, 10)$ (ran on an Intel(R) Core(TM) i7-8565U @1.80GHz computer).