The Annals of Applied Statistics
2019, Vol. 13, No. 2, 900–930
https://doi.org/10.1214/18-AOAS1221

LATENT SPACE MODELLING OF MULTIDIMENSIONAL NETWORKS WITH APPLICATION TO THE EXCHANGE OF VOTES IN EUROVISION SONG CONTEST

BY SILVIA D'ANGELO*, THOMAS BRENDAN MURPHY† AND MARCO ALFÒ*

Sapienza, University of Rome and University College Dublin*†

The Eurovision Song Contest is a popular TV singing competition held annually among country members of the European Broadcasting Union. In this competition, each member can be both contestant and jury, as it can participate with a song and/or vote for other countries' tunes. During the years, the voting system has repeatedly been accused of being biased by tactical voting; votes would represent strategic interests rather than actual musical preferences of the voting countries. In this work, we develop a latent space model to investigate the presence of a latent structure underlying the exchange of votes. Focusing on the period from 1998 to 2015, we represent the vote exchange as a multivariate network: each edition is a network, where countries are the nodes and two countries are linked by an edge if one voted for the other. The different networks are taken to be independent replicates of a conditional Bernoulli distribution, with success probability specified as a function of a common latent space capturing the overall relationships among the countries. Proximity denotes similarity, and countries close in the latent space are more likely to exchange votes. If the exchange of votes depends on the similarity between countries, the quality of the competing songs might not be a relevant factor in the determination of the voting preferences, and this would suggest the presence of some bias. A Bayesian hierarchical modelling approach is employed to estimate the parameters, where the probability of a connection between any two countries is a function of their distance in the latent space, network-specific parameters and edge-specific covariates. The estimated latent space is found to be relevant in the determination of edge probabilities, however, the positions of the countries in such space only partially correspond to their actual geographical positions.

1. Introduction. The Eurovision Song Contest is a popular TV show, held since 1956, that takes place every year with participants from the countries members 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

Received March 2018; revised August 2018.
Key words and phrases. Eurovision, latent space models, multidimensional networks.

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 jth country and 0 otherwise, where $i, j = 1, \dots, n$ 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, 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 subgroups, 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 (2016).

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), $\mathbf{Y} = (\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(K)})$, which may be thought of as the realization of a collection of graphs $G = (G^{(1)}, \dots, G^{(K)})$, where $k = 1, \dots, K$ 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 (2017) 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 multi-way arrays and applied low-rank factorization to infer the underlying structure. Durante, Dunson and Vogelstein (2017) proposed a Bayesian nonparametric 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 summarizes the history of the Eurovision Song Contest together with the principal works on the subject (Section 2.1) and presents the analysed data (Section 2.2). Latent space models for network data are introduced in Section 3 and the proposed model is outlined in Section 3.1. Model estimation is discussed in Section 4. Further issues are discussed in Section 5, such as model identifiability (Section 5.1), missing data (Section 5.2) and the introduction of edge-specific covariates (Section 5.3). The application is presented in Section 6 and the results are discussed in Section 7. A large scale simulation study is outlined in Section 8, where also the main findings are reported. Section 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 10.

2. The Eurovision song contest.

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 10 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 10 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 nonobjective (*nondemocratic*) 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 subgroups of countries. The subgroups found are not fully explained by 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

[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.

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. *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 10 for each country was determined by 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 10 favourites of the jury and the 10 favourites of the televoting. That is, in the last two years of the competition, each country could elicit more than 10 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 predetermined 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 10. 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. Rosén (1972) proved that when sampling r units from a population of size[3] n, if $r \ll n$, 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 unweighted. 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 5.2), and this object is indeed a multidimensional network. The following section introduces a general modelling framework for such data.

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

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

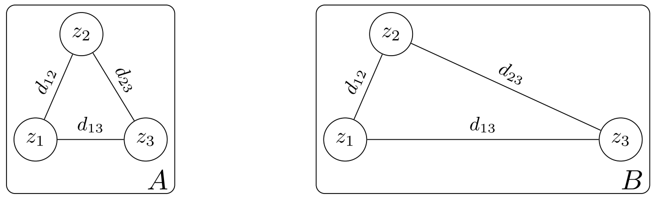


FIG. 1. *Latent space representations of a network with* 3 *nodes*.

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 (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 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 preexisting 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 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 5.2 for details).

The set collection of adjacency matrices will be denoted by $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(K)}\}$. 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 kth 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.

3.1. *The proposed model.* Given the assumptions made in Section 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 kth network of the multiplex is given by

$$\Pr\bigl(y_{ij}^{(k)} = 1 \mid \mathbf{\Omega}^{(k)}, d(z_i, z_j)\bigr) = \frac{\exp\{\alpha^{(k)} - \beta^{(k)} d(z_i, z_j)\}}{1 + \exp\{\alpha^{(k)} - \beta^{(k)} d(z_i, z_j)\}} = p_{ij}^{(k)}, \tag{3.1}$$

where $\mathbf{\Omega} = (\mathbf{\Omega}^{(1)}, \dots, \mathbf{\Omega}^{(K)}) = (\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(K)}, \boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(K)})$ 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 $d(z_i, z_j) = \sum_{l=1}^{p} (z_{il} - z_{jl})^2 = d_{ij}$. The distance matrix will be denoted by $\boldsymbol{D}$. 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 nonsymmetric edge probability 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 kth 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 kth 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

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

and the model for the kth network reduces to a random graph (Erdős and Rényi (1959)) with edge probability $p_{\mathrm{RG}}^{(k)}$. 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 $p_{\mathrm{RG}}^{(k)}$. To counterbalance the effect of the coefficient, the intercept parameter $\alpha^{(k)}$ is bounded as well so that the graph corresponding to $p_{\mathrm{RG}}^{(k)}$ is not disconnected. Indeed, according to the properties of random graphs (Erdős and Rényi (1960)), if $p_{\mathrm{RG}}^{(k)} > \frac{(1-\epsilon)\log(n)}{n}$, 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

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

Thus, the lower bound $\mathrm{LB}(\alpha)$ is independent of $k = 1, \dots, K$, 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 (3.1) would correspond to edge probabilities tending to 0 and numerically undistinguishable. In such a case, two distinct matrices of edge probabilities having elements $p_{ij}^{(k)} \approx 0$ would lead numerically to the same likelihood.

4. Parameter estimation.

4.1. *Likelihood and posterior.* Given the model for the edge probabilities defined in equation (3.1), the likelihood function for the model is a product of $Kn(n-1)$ terms

$$L(\mathbf{\Omega}, \mathbf{D} \mid \mathbf{Y}) = \prod_{k=1}^{K} \prod_{i=1}^{n} \prod_{j \neq i} (p_{ij}^{(k)})^{y_{ij}^{(k)}} (1 - p_{ij}^{(k)})^{1-y_{ij}^{(k)}}, \tag{4.1}$$

and the corresponding log-likelihood is

$$\begin{aligned} \ell(\mathbf{\Omega}, \boldsymbol{D} \mid \mathbf{Y}) &= \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} \ell_{ij}^{(k)} \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} y_{ij}^{(k)} (\alpha^{(k)} \\ &\quad - \beta^{(k)} d_{ij}) - \log(1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij}\}). \end{aligned} \tag{4.2}$$

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: $z_i \sim \mathrm{MV}N_p(\mathbf{0}, \mathbf{I})$. In the present context, the dimension of the multivariate Gaussian is fixed to $p = 2$ (see Section 3).

The parameter space for the intercepts and the coefficients is bounded, as described in paragraph 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: $\alpha^{(k)} \sim N_{[\mathrm{LB}(\alpha),\infty]}(\mu_\alpha, \sigma_\alpha^2)$ and $\beta^{(k)} \sim N_{[0,\infty]}(\mu_\beta, \sigma_\beta^2)$.

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, leading to a hierarchical structure. The prior distributions specified for the nuisance parameters are: $\mu_\alpha | \sigma_\alpha^2 \sim N_{[\mathrm{LB}(\alpha),\infty]}(m_\alpha, \tau_\alpha \sigma_\alpha^2)$, $\sigma_\alpha^2 \sim \mathrm{Inv}\,\chi^2_{\nu_\alpha}$, $\mu_\beta | \sigma_\beta^2 \sim N_{[0,\infty]}(m_\beta, \tau_\beta \sigma_\beta^2)$ and $\sigma_\beta^2 \sim \mathrm{Inv}\,\chi^2_{\nu_\beta}$.

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

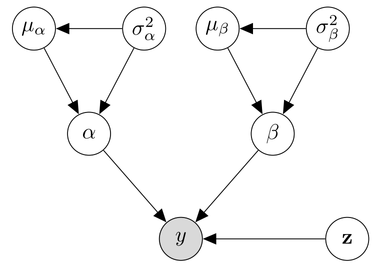


FIG. 2. *Hierarchy structure of the model.*

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

$$\begin{aligned} &P(\boldsymbol{\alpha}, \boldsymbol{\beta}, z, \mu_\alpha, \mu_\beta, \sigma^2_\alpha, \sigma^2_\beta|\mathbf{Y}) \\ &\quad = L(\boldsymbol{\alpha}, \boldsymbol{\beta}, z|\mathbf{Y})\pi(z)\pi(\boldsymbol{\alpha}|\mu_\alpha, \sigma^2_\alpha)\pi(\mu_\alpha|\sigma^2_\alpha, \tau_\alpha) \\ &\qquad \times \pi(\sigma^2_\alpha|\nu_\alpha)\pi(\boldsymbol{\beta}|\mu_\beta, \sigma^2_\beta)\pi(\mu_\beta|\sigma^2_\beta, \tau_\beta)\pi(\sigma^2_\beta|\nu_\beta). \end{aligned} \tag{4.3}$$

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 distributions have been developed and are presented in the Supplementary Naterial (D'Angelo, Murphy and Alfò (2019)), together with the distributions of the nuisance parameters.

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 (4.1) 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 (1998)) 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_α^2 and σ_β^2 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 $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. 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 $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. In a similar fashion, σ_α^2 and σ_β^2 are initialised as the sample variances of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

5. Further issues.

5.1. *Identifiability.* As it can be easily noticed, the likelihood in equation (4.1) is invariant to linear transformations of the scaling coefficients. Indeed, for some constant c,

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

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:

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

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 from a standard Gaussian distribution. This leads to

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

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.

5.2. *The issue of "nonparticipating" 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 an increasing number of participants, preliminary stages had to be introduced to select a smaller subgroup 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 kth 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 10 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 preselection 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 nonparticipant (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

$$\ell(\mathbf{\Omega}, \mathbf{D} \mid \mathbf{Y}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j \neq i} h_{ij}^{(k)} \ell_{ij}^{(k)}, \tag{5.1}$$

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

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 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 a model would be straightforward.

Each covariate is stored in a $n \times n$ matrix, that will be denoted by $\mathbf{X}_f$, where $f = 1, \dots, F$ 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 6). Therefore, the effect associated to each covariate will be characterized in a similar fashion to the scaling coefficients $\beta^{(k)}$, $k = 1, \dots, K$ (see the Supplementary Material). That is, the edge probability in equation (3.1) is modified to the following:

$$\begin{aligned} &P\big(y_{ij}^{(k)} = 1 \mid \alpha^{(k)}, \beta^{(k)}, d_{ij}, \lambda, \mathbf{x}_{ij}\big) \\ &\quad = \frac{\exp\{\alpha^{(k)} - \beta^{(k)} d_{ij} - \sum_{l=1}^{F} \lambda_l x_{ijl}\}}{1 + \exp\{\alpha^{(k)} - \beta^{(k)} d_{ij} - \sum_{l=1}^{F} \lambda_l x_{ijl}\}}. \end{aligned} \tag{5.2}$$

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.

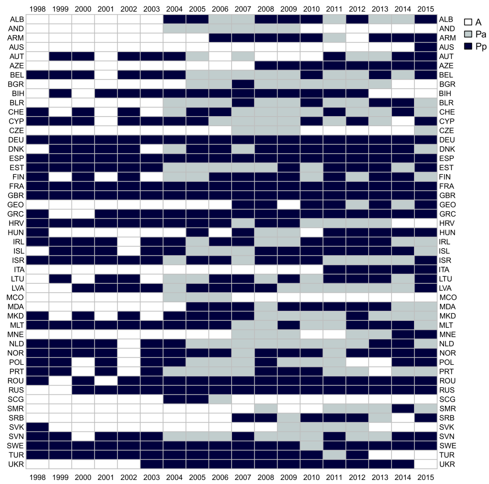
6. The Eurovision song contest data. We considered 18 different editions of the Eurovison Song Contest, from 1998 to 2015. In this period, two major changes have occurred in the structure of the program, due to the growing number of countries willing to participate in the show. First, in 2004, a semi-final 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 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 3 describe some features of the 18 networks. In particular, in Figure 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 3(b) reports the values for the association in the exchanging of votes between two different editions, measured by the index

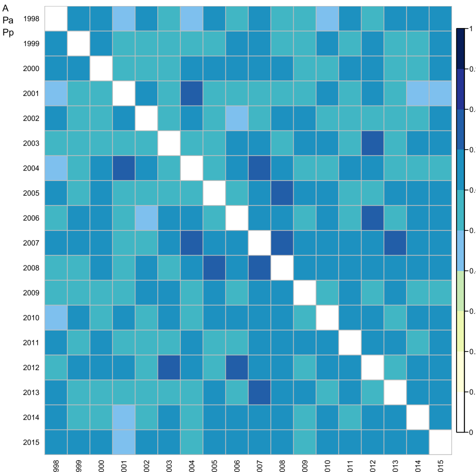
$$A_{(k,l)} = \frac{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = h_{ij}^{(l)} y_{ij}^{(l)})}{\sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} = h_{ij}^{(l)} y_{ij}^{(l)}) + \sum_{i,j}^{n} \mathbf{I}(h_{ij}^{(k)} y_{ij}^{(k)} \neq h_{ij}^{(l)} y_{ij}^{(l)})}.$$

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 3(c) and 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 3(d) is not symmetric and the ith row shows the average number of votes that country i gave to others. Instead, the jth 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.

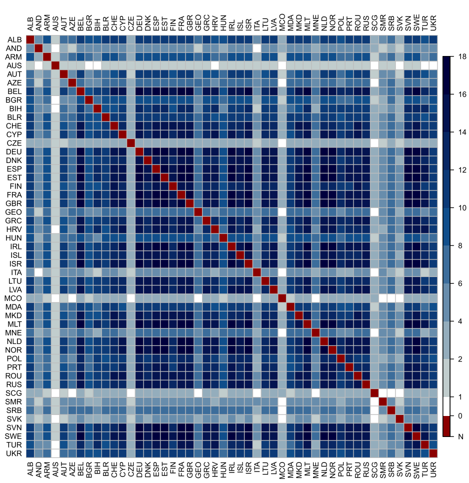
At a second stage, covariates have been included in the analysis, similarly to what has been done by Blangiardo and Baio (2014), Spierdijk and Vellekoop (2006), in order to see whether the vote exchanges in the period could be, at least partially, explained by "cultural" factors. The covariates we considered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



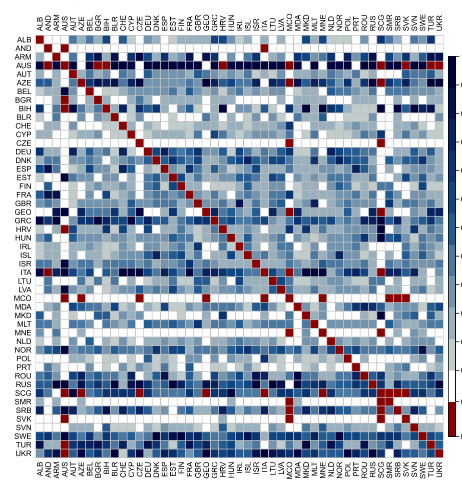
(a) Countries participation by year.



(b) Association between adjacency matrices.



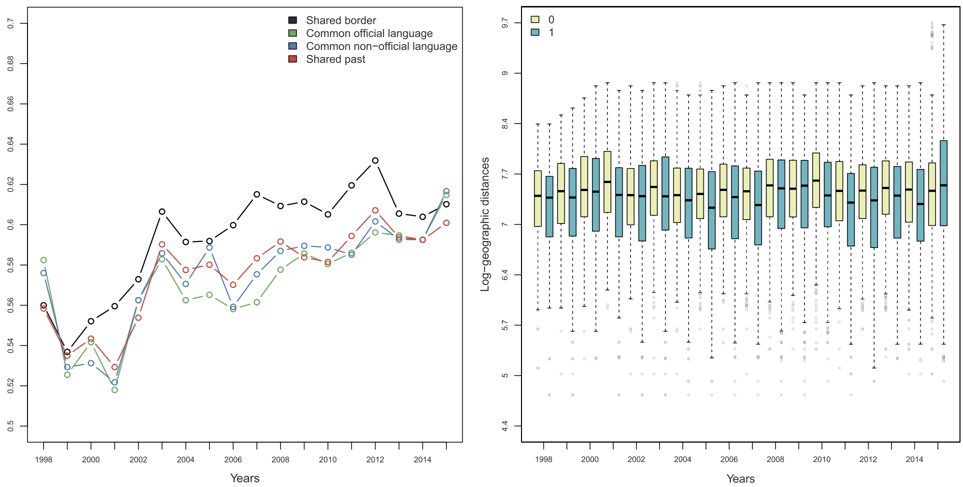
(c) Number of times two countries have jointly participated to the competition, with null values in the diagonal (N).



(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.

FIG. 3. *Eurovision data*: *some exploratory statistics*.

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.



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

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

FIG. 4. *Covariates.*

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 4 describes the association between the covariates and the adjacency matrices. The plot in 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 4(b) reports the boxplots for the couple (adjacency matrix, $\log \mathbf{X}_1$). We can not find, at least visually, 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 subperiods, 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 subperiods 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/.

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 50,000 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)):

$$\mathrm{DIC} = D(\hat{\mathbf{\Theta}}) + 2\big(\bar{D}(\mathbf{\Theta}) - D(\hat{\mathbf{\Theta}})\big),$$

where $D(\mathbf{\Theta}) = -2\log L(\mathbf{\Theta})$ and $\Theta = (\mathbf{\Omega}, \mathbf{D}, \lambda)$. The deviance term $D(\hat{\mathbf{\Theta}})$ is computed with the posterior estimates, while $\bar{D}(\mathbf{\Theta})$ is the mean deviance of the posterior distribution. The best model is the one with the lowest value of the DIC, see

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

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

TABLE 1
DIC values for fitted models

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
DIC	19,584.12	19,494.74	19,461.52	**19,412.12**	23,340.84	23,899.62

Table 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. 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 5, 6 and 7 show the estimates obtained under *model 2* for the latent positions, the distances and the posterior distribution for the parameters of interest. Figure 5(a) reports the posterior means for the country latent coordinates (reported with their ISO3 codes, see the Supplementary Material) together with

TABLE 2
Model parameters: *estimated averages and standard deviations*, 1998–2015

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

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 define one of the major immigrant group in Spain, and Latvians are one of the largest immigrant groups in Ireland. Therefore, some of the *geographical misplacements* within the subgroups 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 6 and 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 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 3. The table confirms what was already visible from Figure 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 9 reports the matrix of the intersections between the sets of neighbours LN_{i,r_i^*} and

[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

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

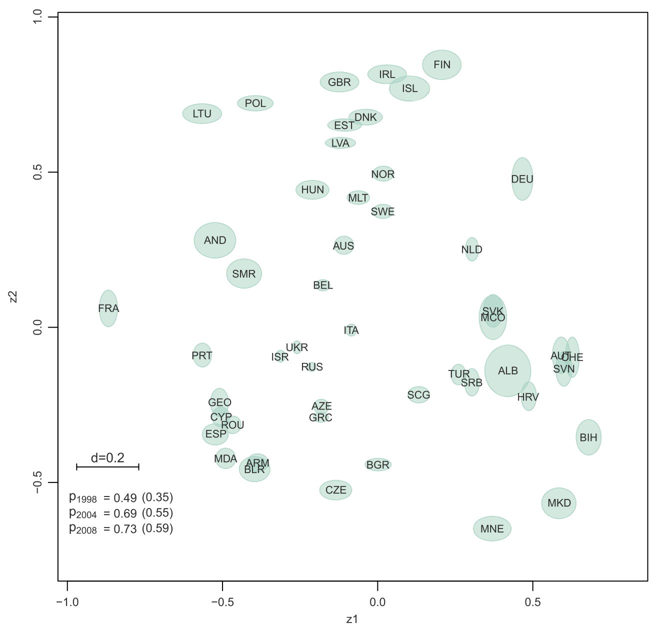


FIG. 5. *Estimated latent positions* 1998–2015. *The legend reports the probabilities corresponding to a distance of* 0.2 *in the latent space*, *years* 1998, 2004 *and* 2008. *The values refer to the case of* $x_{2,ij} = 0$, *within the brackets are reported the values for* $x_{2,ij} = 1$.

CN_{i,r_i^*}, for 1998–2015. Bulgaria, Lithuania and Serbia and Montenegro are the countries that tend to vote more for 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 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 parameters $\beta^{(k)}$ 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,

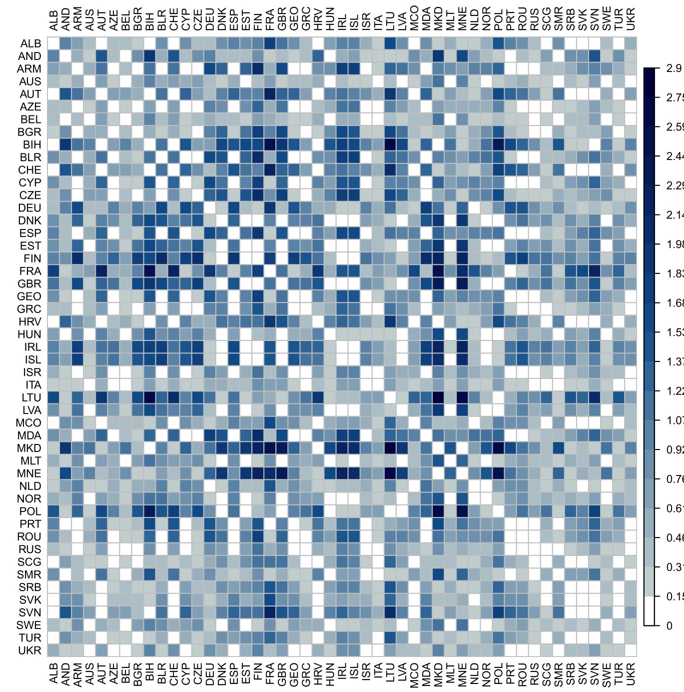


FIG. 6. *Estimated distances between couple of countries*, 1998–2015.

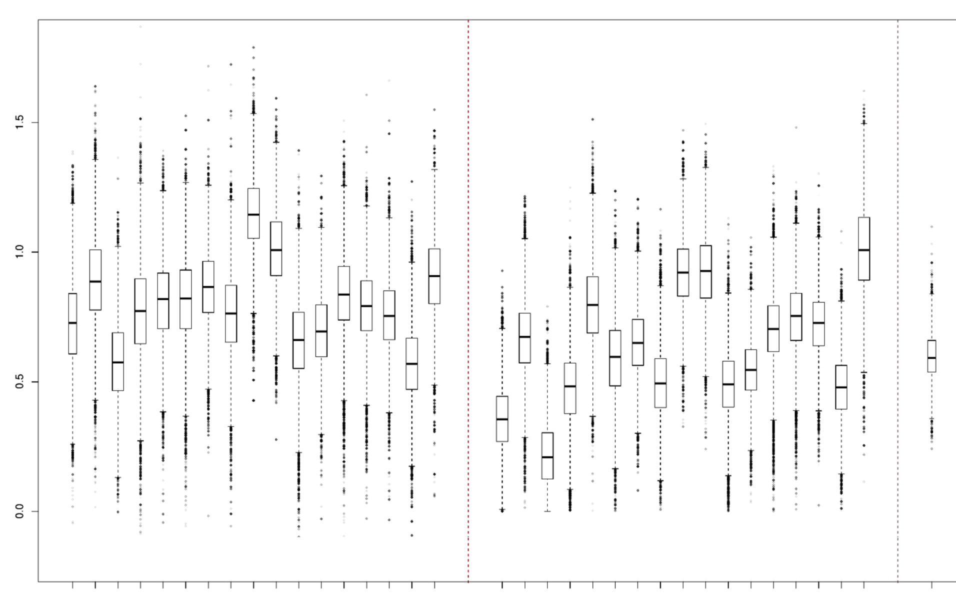
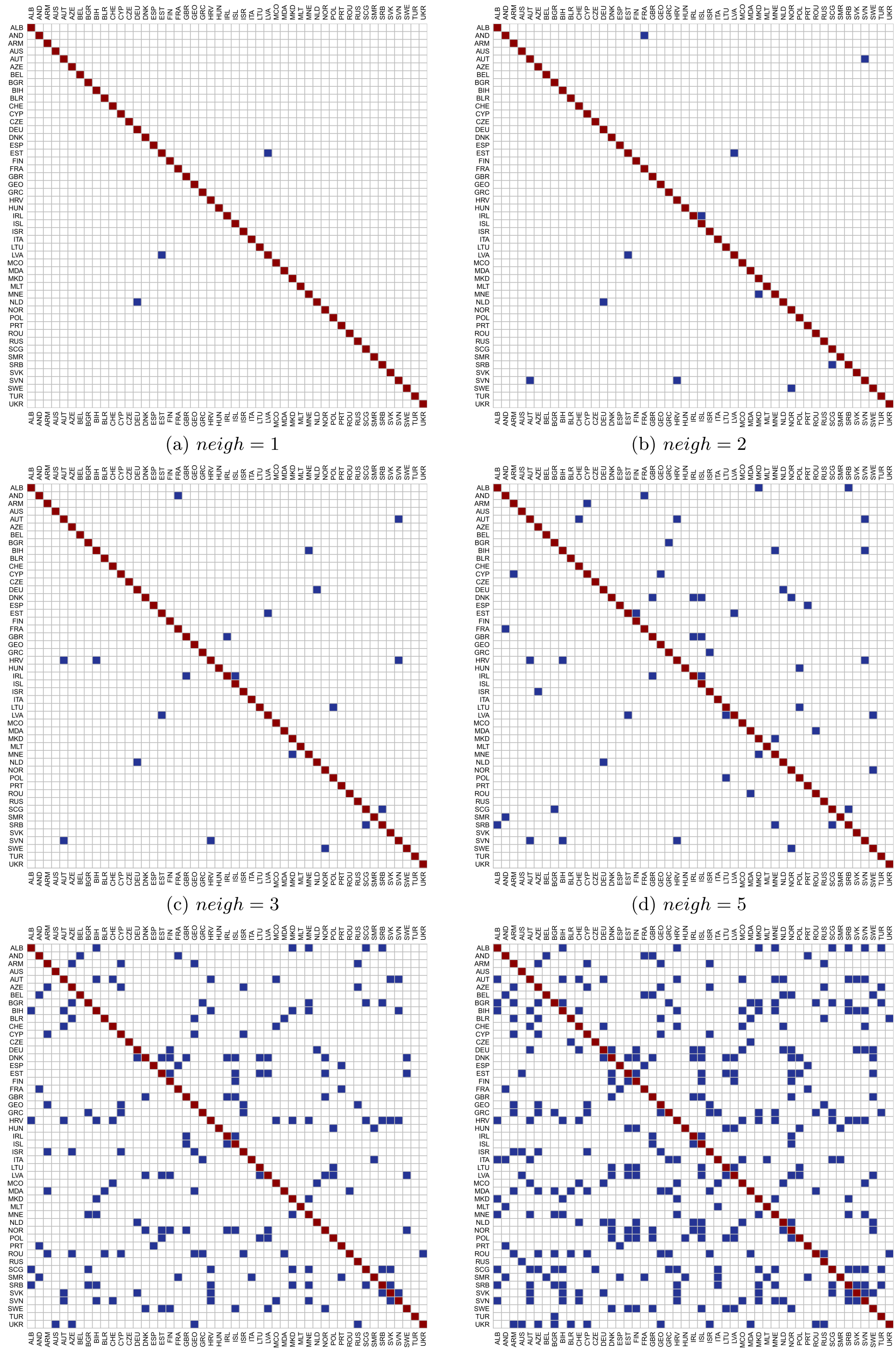


FIG. 7. *Boxplots for model parameter estimates and the coefficient for* $\mathbf{X}_2$, 1998–2015.



(a) $neigh = 1$

(b) $neigh = 2$

(c) $neigh = 3$

(d) $neigh = 5$

(e) $neigh = 10$

(f) $neigh = 15$

FIG. 8. *Intersections of the set of neighbours $LN_{i,r}$ and $GN_{i,r}$*, 1998–2015.

TABLE 3

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

	Q	$r=2$	$r=3$	$r=5$	$r=10$	$r=15$
Average number	0.06	0.11	0.14	0.21	0.37	0.47
Maximum number	1	1	2	3	8	12

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

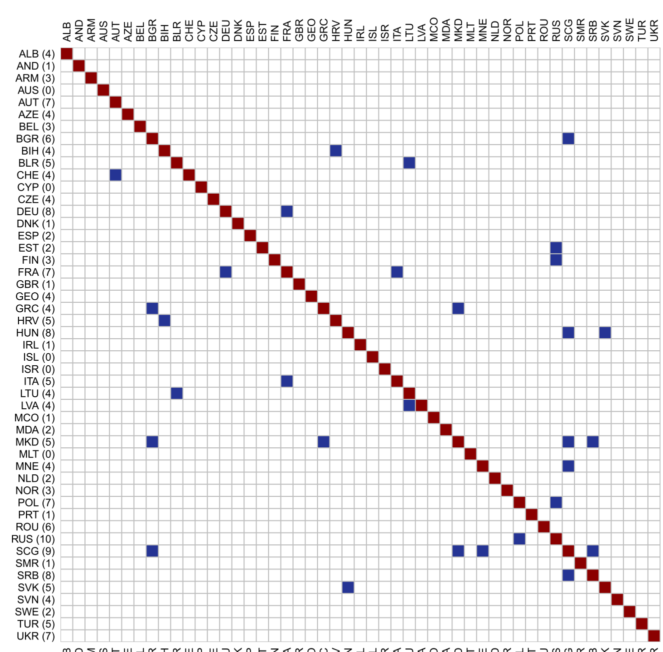


FIG. 9. *Intersection of the set of neighbours LN_{i,r_i^*} and CN_{i,r_i^*}*, 1998–2015. *On the left column, in brackets, are reported the values of r_i^*.*

is different. Indeed, not all the participants in 1998–2015 were also participant in both of the subperiods. Just to give an example, Italy rejoined the competition in 2011, after being absent in the period 1998–2007. The multidimensional networks for the subperiods can not include the same set of nodes of the period 1998–2015. Indeed, countries completely absent from the competition, in a given subperiod, 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 subperiods 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 subgroups observed in Figure 5 are still present. For example, Northern Europe countries tend always to be closer to each other, as do Eastern Europe countries.

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 4.1), with $\sigma_\alpha^2 = \sigma_\beta^2 = 1$, $\mu_\alpha = \mu_\beta = 0$. 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 degrees 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 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 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 4.1), with $\sigma_\alpha^2 = \sigma_\beta^2 = 1$, $\mu_\alpha = \mu_\beta = 0$.

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 4.2 for details). Each model was estimated 10 times, performing 40,000 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 detail 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.

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

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.

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 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 variety 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).

Acknowledgments. We would like to thank the Editor and the referees, whose suggestions and comments helped improve this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Latent space modelling of multidimensional networks with application to the exchange of votes in Eurovision song contest" (DOI: 10.1214/18-AOAS1221SUPP; .pdf). The Supplementary Material (D'Angelo, Murphy and Alfò (2019)) provide with the derivation of the full conditional and proposal distributions used to estimate the model, a table with the ISO3 codes for the countries participating in the Eurovision Song Contest in the period 1998–2015, the results of the analysis for the subperiods 1998–2007 and 2008–2015, the results of the different simulation scenarios and a pseudo-code of the algorithm used for parameter estimation.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed-membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

BLANGIARDO, M. and BAIO, G. (2014). Evidence of bias in the Eurovision song contest: Modelling the votes using Bayesian hierarchical models. *J. Appl. Stat.* **41** 2312–2322. MR3292674

BUTTS, C. T. and CARLEY, K. M. (2005). Some simple algorithms for structural comparison. *Computational and Mathematical Organization Theory* **11** 291–305.

CLERIDES, S. and STENGOS, T. (2006). Love thy neighbour, love thy kin: Strategy and bias in the Eurovision song contest. *Discussion Paper Series of Centre for Economic Policy Research* **5732** 1–28.

D'ANGELO, S., MURPHY, T. B. and ALFÒ, M. (2019). Supplement to "Latent space modelling of multidimensional networks with application to the exchange of votes in Eurovision song contest." DOI:10.1214/18-AOAS1221SUPP.

DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, Chichester. MR1646114

DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J. Amer. Statist. Assoc.* **112** 1516–1530. MR3750873

ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

FENN, D., SULEMANA, O., EFSTATHIOUB, J. and JOHNSON, N. F. (2006). How does Europe make its mind up? Connections, cliques, and compatibility between countries in the Eurovision song contest. *Phys. A* **360** 576–598.

FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple sociometric relations. *J. Amer. Statist. Assoc.* **80** 51-67.

FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

GINSBURGH, V. and NOURY, A. (2008). The Eurovision song contest. Is voting political or cultural?. *European Journal of Political Economy* **24** 41–52.

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

GOLLINI, I. and MURPHY, T. B. (2016). Joint modeling of multiple network views. *J. Comput. Graph. Statist.* **25** 246–265. MR3474046

GREENE, D. and CUNNINGHAM, P. (2013). Producing a unified graph representation from multiple social network views. *Proceedings of the* 5*th Annual ACM Web Science Conference* (*Web-Sci'*13) 118–121.

HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *J. Roy. Statist. Soc. Ser. A* **170** 301–354. MR2364300

HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.* **100** 286–295. MR2156838

HOFF, P. D. (2011). Hierarchical multilinear models for multiway data. *Comput. Statist. Data Anal.* **55** 530–543. MR2736574

HOFF, P. (2015). Dyadic data analysis with amen.

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088

KRIVITSKY, P. N., HANDCOCK, M. S., RAFTERY, A. E. and HOFF, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc. Netw.* **31** 204–213.

LYNCH, K. (2015). Eurovision recognised by Guinness world records as the longest-running annual TV music competition (international). Available at http://www.guinnessworldrecords.com/news/2015/5/eurovision-recognised-by-guinness-world-records-as-the-longest-running-annual-tv-379520.

MANTZARIS, A. V., REIN, S. R. and HOPKINS, A. D. (2018). Examining collusion and voting biases between countries during the Eurovision song contest since 1957. *Journal of Artificial Societies and Social Simulation* **21**.

MURPHY, T. B. (2016). Model-based clustering for network data. In *Handbook of Cluster Analysis*. *Chapman & Hall/CRC Handb*. *Mod*. *Stat*. *Methods* 337–357. CRC Press, Boca Raton, FL. MR3644719

ROBINS, G., PATTISON, P., KALISH, Y. and LUSHER, D. (2007). An introduction to exponential random graph (p*) models for social network. *Soc*. *Netw.* **29** 173–191.

ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I. *Ann*. *Math*. *Stat*. **4** 373–397. MR0321223

SAAVEDRAA, S., EFSTATHIOUA, J. and REED-TSOCHASB, F. (2007). Identifying the underlying structure and dynamic interactions in a voting network. *Phys*. *A* **377** 672–688.

SALTER-TOWNSHEND, M. and MCCORMICK, T. H. (2017). Latent space models for multiview network data. *Ann*. *Appl*. *Stat*. **11** 1217–1244. MR3709558

SALTER-TOWNSHEND, M., WHITE, A., GOLLINI, I. and MURPHY, T. B. (2012). Review of statistical network analysis: Models, algorithms, and software. *Stat*. *Anal*. *Data Min*. **5** 260–264. MR2958152

SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. MR1449742

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J*. *R*. *Stat*. *Soc*. *Ser. B*. *Stat*. *Methodol*. **64** 583–639. MR1979380

SPIERDIJK, L. and VELLEKOOP, M. (2006). Geography, culture, and religion: Explaining the bias in Eurovision song contest voting. *Memorandum*, *Department of Applied Mathematics*, *University of Twente*, *Enschede* **1794** 1–30.

SWEET, T. M., THOMAS, A. C. and JUNKER, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioural Statistics* **38** 295–318.

YAIR, G. (1995). "Unite unite Europe" the political and cultural structures of Europe as reflected in the Eurovision song contest. *Soc*. *Netw.* **17** 147–161.

S. D'ANGELO
M. ALFÒ
DEPARTMENT OF STATISTICAL SCIENCES
SAPIENZA, UNIVERSITY OF ROME
PIAZZALE ALDO MORO 5, ROME
ITALY
E-MAIL: silvia.dangelo@uniroma1.it
marco.alfo@uniroma1.it

T. B. MURPHY
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD, DUBLIN 4
IRELAND
E-MAIL: brendan.murphy@ucd.ie