Computational Statistics and Data Analysis 169 (2022) 107432

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

A latent space model for multilayer network data

Juan Sosa [a,*], Brenda Betancourt [b]

[a] Universidad Nacional de Colombia, Carrera 45 # 26-85, Edificio Uriel Gutiérrez, Bogotá D.C., Colombia
[b] University of Florida, 220 Griffin-Floyd Hall, P.O. Box 118545, Gainesville, FL, 32611-8545, United States of America

ARTICLE INFO

Article history:
Received 17 February 2021
Received in revised form 11 January 2022
Accepted 12 January 2022
Available online 18 January 2022

Keywords:
Bayesian modeling
Cognitive social structures data
Latent space models
Markov chain Monte Carlo
Multilayer network data
Social networks

ABSTRACT

A Bayesian statistical model to simultaneously characterize two or more social networks defined over a common set of actors is proposed. The key feature of the model is a hierarchical prior distribution that allows the user to represent the entire system jointly, achieving a compromise between dependent and independent networks. Among others things, such a specification provides an easy way to visualize multilayer network data in a low-dimensional Euclidean space, generate a weighted network that reflects the consensus affinity between actors, establish a measure of correlation between networks, assess cognitive judgments that subjects form about the relationships among actors, and perform clustering tasks at different social instances. The model's capabilities are illustrated using real-world and synthetic datasets, taking into account different types of actors, sizes, and relations.

1. Introduction

The study of information that emerges from the interconnectedness among autonomous elements in a system (and the elements themselves) is extremely important in the understanding of many phenomena. Structures formed by these elements (individuals or actors) and their interactions (ties or connections), commonly known as networks, are popular in many research areas such as finance (studying alliances and conflicts among countries as part of the global economy), social science (studying interpersonal social relationships and social schemes of collaboration such as legislative cosponsorship networks), biology (studying arrangements of interacting genes, proteins or organisms), epidemiology (studying the spread of a infectious disease), and computer science (studying the Internet, the World Wide Web, and also communication networks), just to mention a few examples, primarily because interactions typically arise under several contexts or points of view.

Relational structures consisting of J types of interactions (layers or views) established over a common set of I actors are regularly referred to as either multilayer or multiview network data, which results in a sequence of J adjacency matrices $\mathbf{Y}_1, \ldots, \mathbf{Y}_J$, with $\mathbf{Y}_j = [y_{i,i',j}]_{i,i'=1,\ldots,I,i\neq i'}$ for $j = 1, \ldots, J$, each having structural zeros along the main diagonal (note that $y_{i,i',j} \equiv y_{i',i,j}$ for undirected relations). This type of data is very frequent nowadays. For instance, the interactions that employees have with others according to their roles in work are not necessarily the same as the interpersonal relationships they build among them; however, the corresponding social structures defined by these two types of relationships may have some characteristics in common. Thus, given the richness of information provided in $\mathbf{Y} = \{\mathbf{Y}_j\}$, our main goal consists of

* Corresponding author.
E-mail address: jcsosam@unal.edu.co (J. Sosa).

https://doi.org/10.1016/j.csda.2022.107432
0167-9473/

modeling dependencies both within and between layers in order to formally test features about the social dynamics in the system.

A very popular statistical model in the literature for a single network is the latent position model given in Hoff et al. (2002). According to this model, interaction probabilities marginally depend on how close or far apart actors are on a latent "social space" (a K-dimensional vector space, typically $\mathbb{R}^K$, in which each individual occupies a fixed position). This formulation is appealing because latent structures based on distances naturally induce transitivity and homophily, which are typical features found in many social networks. Other meaningful advances in latent space models for networks can be found in Nowicki and Snijders (2001), Schweinberger and Snijders (2003), Hoff (2005, 2008, 2009), Handcock et al. (2007), Linkletter (2007), Krivitsky and Handcock (2008), Krivitsky et al. (2009), and Li et al. (2011).

Here, we extend Hoff's latent position model in order to describe the generative process of cross-sectional multilayer network data. The key feature of our model is a hierarchical prior distribution that allows us to characterize the entire system jointly. Such a prior specification is very convenient for multiple reasons. First, the model provides a direct description of actors' roles within and across networks at global and specific levels. Second, it provides the tools for representing several network features effortlessly at any instance. Finally, the proposed framework accounts for dependence structures between layers which is key to perform formal tests about actor and network characteristics.

Perhaps the closest in spirit to our modeling strategy is the work given in Gollini and Murphy (2016) and Salter-Townshend and McCormick (2017). Unlike our approach, Gollini and Murphy (2016) introduce a latent space model assuming that the interaction probabilities in each network view are explained by a unique latent variable. Later, Salter-Townshend and McCormick (2017) consider the same assumption but in the context of a multivariate Bernoulli likelihood, which leads to a clear estimate of interview dependence. Our proposal builds on the latent configuration of these models, by considering a full hierarchical prior specification that provides a parsimonious characterization of actors from many perspectives.

Aside from the previous work, other alternatives for studying multilayer network data have emerged during the last two years from the latent space modeling perspective. In brain connectomics, Durante and Dunson (2018) present a Bayesian nonparametric approach via mixture modeling, which reduces dimensionality and efficiently incorporates network information within each mixture component by leveraging latent space representations. More recently, Wang et al. (2019) propose a method to estimate a common structure and low-dimensional individual-specific deviations from replicated networks, based on a logistic regression mapping combined with a hierarchical singular value decomposition. In turn, D'Angelo and collaborators extend latent space models in other contexts, by considering node-specific effects (D'Angelo et al., 2020a), network-specific parameters and edge-specific covariates (D'Angelo et al., 2019), and finally, a clustering structure in the framework of an infinite mixture distribution (D'Angelo et al., 2020b). Finally, in the field of neuroscience and also under a frequentist paradigm, Wilson et al. (2020) develop a hierarchical latent space model accounting for multiple predictors and specific effects for both individuals and layers. Other important advances from a frequentist point of view are available in Zhang (2020).

Additional work related to cross-sectional multilayer network data includes community detection (e.g., Han et al. (2015), Reyes and Rodriguez (2016), Paul and Chen (2016), Gao et al. (2019), Paez et al. (2019), Chen et al. (2020), Paul et al. (2020a), and Paul et al. (2020b)), and the perception assessment in cognitive social structures (e.g., Swartz et al. (2015), Sewell (2019), and Sosa and Rodriguez (2021)). Finally, from the dynamic point of view, there is a large variety of approaches to modeling network evolution over time (e.g., Durante and Dunson (2014), Hoff (2015), Sewell and Chen (2015, 2016, 2017), Gupta et al. (2018), Kim et al. (2018), Turnbull (2020), Betancourt et al. (2020)).

Our contribution has many folds. In Section 2, we present our proposal for modeling multiple layer network data, including prior elicitation. In Section 3, we discuss the topics of identifiability and model selection for our approach. Next, in Section 4, we provide two illustrations using popular data sets in the literature, for which we develop formal tests involving network correlation and perceptual agreement, as well as a full analysis of the social dynamics. In Section 5, we carry out a cross-validation study on additional datasets in order to test the predictive capabilities of our proposed model. Then, in Section 6, we perform an extensive simulation study in order to investigate the capabilities of our proposal. Finally, concluding remarks and directions for future work are provided in Section 7.

2. Latent space models

Since the foundational work of Hoff et al. (2002), generalized linear mixed models became a popular alternative to model networks. In particular, consider an undirected binary network $\mathbf{Y} = [y_{i,i'}]$ in which the $y_{i,i'}$ s, $i, i' = 1, \ldots, I$, $i < i'$, are assumed to be conditionally independent with interaction probabilities $\vartheta_{i,i'} \equiv \Pr\left(y_{i,i'} = 1 \mid \zeta, \gamma_{i,i'}\right) = \Phi(\zeta + \gamma_{i,i'})$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard Gaussian distribution (other link functions can be considered), ζ is a fixed effect representing the global propensity of observing an edge between actors i and i', and $\gamma_{i,i'}$ is an unobserved dyad-specific random effect representing any additional patterns unrelated to those captured by ζ. Following results in Hoover (1982) and Aldous (1985) (see also Hoff (2008)), it can be shown that if the matrix of random effects $[\gamma_{i,i'}]$ is jointly exchangeable, there exists a symmetric function $\alpha(\cdot, \cdot)$ and a sequence of independent random variables (vectors) $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_I$ such that $\gamma_{i,i'} = \alpha(\boldsymbol{u}_i, \boldsymbol{u}_{i'})$. It is mainly through $\alpha(\cdot, \cdot)$ that we are able to capture relevant features of the network. A number of potential formulations for $\alpha(\cdot, \cdot)$ have been explored in the literature to date; see Minhas et al. (2019) and Sosa and Buitrago (2021) for a review.

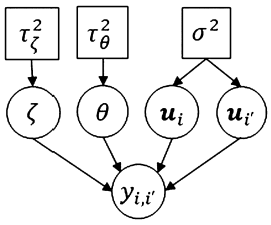


Fig. 1. DAG representation of the LPM for a single network. Circles represent either random variables or random vectors, and the edges convey conditional independence. Squares represent fixed quantities (constants).

In particular, consider the latent position model (LPM) given in Hoff et al. (2002). This model assumes that each actor i has an unknown position $\boldsymbol{u}_i$ in a social space of latent characteristics, typically $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,K}) \in \mathbb{R}^K$, where K is assumed to be known, and that the probability of an edge between two actors may decrease as the latent characteristics of the individuals become farther apart of each other. In this spirit, the latent effects can be specified as $\gamma_{i,i'} = -e^{\theta} \| \boldsymbol{u}_i - \boldsymbol{u}_{i'} \|$, where e^{θ} serves as a weighting factor that regulates the contribution attributed to the latent effects, and therefore, $y_{i,i'} \mid \zeta, \theta, \boldsymbol{u}_i, \boldsymbol{u}_{i'} \overset{\text{ind}}{\sim} \text{Ber}\left(\Phi\left(\zeta - e^{\theta} \| \boldsymbol{u}_i - \boldsymbol{u}_{i'} \|\right)\right)$. In order to perform a fully Bayesian analysis, we must specify a prior distribution on the model parameters; a standard choice that works well in practice consists in setting mutually independent prior distributions, $\zeta \sim \text{N}_1(0, \tau_\zeta^2)$, $\theta \sim \text{N}_1(0, \tau_\theta^2)$, and $\boldsymbol{u}_i \overset{\text{iid}}{\sim} \text{N}_K(\boldsymbol{0}, \sigma^2 \mathbf{I})$, for constants $\tau_\zeta^2, \tau_\theta^2, \sigma^2 > 0$, although other similar formulations are available (e.g., Rastelli et al. (2019)). Thus, the entire model has $IK + 2$ unknown parameters to estimate, namely, $\zeta, \theta, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_I$, associated with the hyperparameters τ_ζ^2, τ_θ^2, and σ^2, which need to be picked sensibly to ensure appropriate model performance. In our experience, letting $\tau_\zeta^2 = \tau_\theta^2 = 3$ and $\sigma^2 = 1/9$ is a reasonable choice; however, other heuristics are possible (e.g., Krivitsky et al. (2009)). Fig. 1 provides a directed acyclic graph (DAG) representation of the LPM for a single network. In the following section we present an extension of this model that is suited for multilayer networks.

2.1. Extension to multilayer networks

Here, we present our approach to simultaneously model a set of $J \geq 2$ undirected binary networks $\mathbf{Y}_1, \ldots, \mathbf{Y}_J$, defined over a common set of I actors, with $\mathbf{Y}_j = [y_{i,i',j}]$ for $j = 1, \ldots, J$. Since each network contains relevant information about a determined aspect of the social dynamics, instead of just fitting independent LPMs to each network, the main idea behind our approach consists of borrowing information across networks by means of a hierarchical prior specification on the interaction probabilities $\vartheta_{i,i',j}$.

Our model is an unequivocal hierarchical extension of the LPM for a single network that accommodates relevant features associated with multilayer network data. Here, we still assume that observations are conditionally independent, $y_{i,i',j} \mid \vartheta_{i,i',j} \overset{\text{ind}}{\sim} \text{Ber}\left(\vartheta_{i,i',j}\right)$, and construct a hierarchical prior on the array $[\vartheta_{i,i',j}]$, by letting

$$\vartheta_{i,i',j} = \Phi\left(\zeta_j - e^{\theta_j} \| \boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \|\right), \tag{1}$$

where the additional index j makes explicit the reference to network j, i.e., ζ_j is the global propensity of observing an edge between actors i and i' in network j, e^{θ_j} is a weighting factor that regulates the contribution attributed to the latent effects in network j, and $\boldsymbol{u}_{i,j} = (u_{i,j,1}, \ldots, u_{i,j,K})$ is the latent position of actor i in network j. In this context, note that the interpretation of the latent structure remains unchanged: if $\boldsymbol{u}_{i,j}$ and $\boldsymbol{u}_{i',j}$ "move away" from each other in the social space, then $\| \boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \|$ increases, and therefore, the probability of observing an edge between actors i and i' in network j may decrease depending on the regularization provided by e^{θ_j}. Here, we explicitly acknowledge that considering latent positions embedded in a communal social space for modeling interactions in a multilayer setting does not constitute a novel feature by itself and it is actually quite reminiscent of previous approaches in the literature (e.g., Gollini and Murphy (2016), Salter-Townshend and McCormick (2017), and D'Angelo et al. (2019)). However, our formulation deviates from others in terms of the form and structure of the sampling distribution as well as the prior specification (see Section 7 for some specific differences). Lastly, since our latent positions are subject-layer specific (again not a new structural modeling feature) as opposed to subject specific, we propose a two-stage hierarchical prior distribution as shown below.

In addition, unlike Gollini and Murphy (2016), Salter-Townshend and McCormick (2017), D'Angelo et al. (2019), and many others, we employ a probit link instead of a logit one, in the specification of the interaction probabilities given in (1). In our experience as well as that of others (e.g., Lofland et al. (2017) and Sosa and Rodriguez (2021)), both probit and logit link functions provide essentially similar results, which in general seems to be the case for univariate binary response models like ours (see Hahn and Soyer (2005) and references therein). Interestingly, based on Hahn and Soyer's analysis, the probit link tends to improve both in-sample and out-of-sample model performance, particularly when multivariate binary response models with random effects are fitted to moderate size datasets; by contrast, the logit link seems preferable for multivariate link models when there are extreme independent variable levels. Nonetheless, adapting our approach to handle a logit link functions is straightforward.

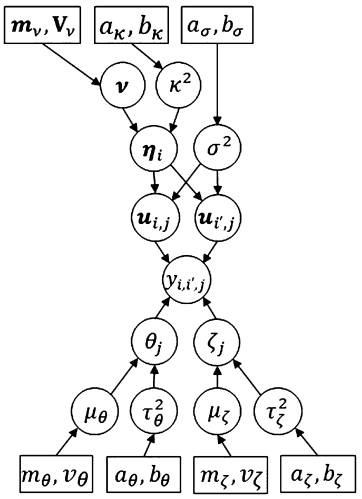


Fig. 2. DAG representation of the MNLPM for multilayer network data.

If mutually independent prior distributions were assigned to each set of ζ_j s, θ_j s, and $\boldsymbol{u}_{i,j}$ s, then such a formulation would be equivalent to fitting independently a LPM to each network. Instead, we consider a hierarchical prior distribution that characterizes the heterogeneity of the model parameters across networks. Our approach parsimoniously places conditionally independent Gaussian priors as follows:

$$\zeta_j \mid \mu_\zeta, \tau_\zeta^2 \overset{\text{iid}}{\sim} \mathsf{N}_1\left(\mu_\zeta, \tau_\zeta^2\right), \;\; \theta_j \mid \mu_\theta, \tau_\theta^2 \overset{\text{iid}}{\sim} \mathsf{N}_1\left(\mu_\theta, \tau_\theta^2\right), \;\; \boldsymbol{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma^2 \overset{\text{ind}}{\sim} \mathsf{N}_K\left(\boldsymbol{\eta}_i, \sigma^2 \mathbf{I}\right). \tag{2}$$

On the one hand, $(\mu_\zeta, \tau_\zeta^2)$ and $(\mu_\theta, \tau_\theta^2)$ parameterize the sampling distributions that describe the heterogeneity across networks in terms of fixed effects $\zeta_1, \ldots, \zeta_J$ and weighting log-factors $\theta_1, \ldots, \theta_J$, respectively. On the other hand, the mean $\boldsymbol{\eta}_i = (\eta_{i,1}, \ldots, \eta_{i,K})$ can be conveniently interpreted as the average "global" position of actor i in relation to the dynamics that define social interactions in the system. Now, we can capture similarities among the observed networks and borrow information across them, mainly by placing a common prior distribution on $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_I$. Thus, we let

$$\boldsymbol{\eta}_i \mid \boldsymbol{\nu}, \kappa^2 \overset{\text{iid}}{\sim} \mathsf{N}_K\left(\boldsymbol{\nu}, \kappa^2 \mathbf{I}\right), \qquad \sigma^2 \sim \mathsf{IG}\left(a_\sigma, b_\sigma\right), \tag{3}$$

in order to characterize between-actor mean sampling variability in a straightforward fashion. Furthermore, note that the sampling variability of the latent positions σ^2 is assumed to be constant across actors and networks. We believe this is a sensible choice because inferences on the latent positions seem to be invariant when we eliminate such an assumption.

Finally, the model is completed by specifying prior distributions in a conjugate fashion on the remaining model parameters:

$$\begin{aligned} &\mu_\zeta \sim \mathsf{N}_1\left(m_\zeta, v_\zeta^2\right), \qquad \mu_\theta \sim \mathsf{N}_1\left(m_\theta, v_\theta^2\right), \qquad \boldsymbol{\nu} \sim \mathsf{N}_K\left(\boldsymbol{m}_\nu, \mathbf{V}_\nu\right), \\ &\tau_\zeta^2 \sim \mathsf{IG}\left(a_\zeta, b_\zeta\right), \qquad \tau_\theta^2 \sim \mathsf{IG}\left(a_\theta, b_\theta\right), \qquad \kappa^2 \sim \mathsf{IG}\left(a_\kappa, b_\kappa\right), \end{aligned} \tag{4}$$

where $a_\sigma, b_\sigma, a_\zeta, b_\zeta, a_\theta, b_\theta, a_\kappa, b_\kappa, m_\zeta, v_\zeta, m_\theta, v_\theta, \boldsymbol{m}_\nu, \mathbf{V}_\nu$ are fixed hyperparameters. Therefore, the full set of model parameters is

$$\boldsymbol{\Upsilon} \equiv \boldsymbol{\Upsilon}_{I,J,K} = \left(\zeta_1, \ldots, \zeta_J, \theta_1, \ldots, \theta_J, \boldsymbol{u}_{1,1}, \ldots, \boldsymbol{u}_{I,J}, \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_I, \sigma, \mu_\zeta, \tau_\zeta, \mu_\theta, \tau_\theta, \boldsymbol{\nu}, \kappa\right),$$

which includes $IK(J+1)+2J+K+6$ unknown quantities to estimate. Fig. 2 shows a DAG representation of our Multilayer Network Latent Position Model (MNLPM) given in (1), (2), (3), and (4). Note the clear hierarchical structure in the model.

This model for multilayer network data is such that the resulting marginal distribution of the data is fully jointly exchangeable, which means that the joint distribution of $\{y_{i,i',j}\}$ is the same as the distribution of $\{y_{\pi_1(i),\pi_2(i'),\pi_3(j)}\}$ only if $\pi_1 = \pi_2$, where π_1 and π_2 are permutations of $[I]$, and π_3 is a permutation of $[J]$. Full joint exchangeability (rather than a weaker form of exchangeability) is particularly attractive in this setting because all indexes i, i' (and potentially j) refer to the same set of actors (Sosa and Rodriguez, 2021).

2.2. Prior elicitation

Once again, careful elicitation of the hyperparameters is key to ensure appropriate model performance since the model is sensitive to this choice. In our experience, the following heuristic procedure produces adequate results for a wide variety

Table 1
Expected value and standard deviation of the prior latent distance $d_{i,i',j} = \|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|$, rate hyperparameters, and variance components for the prior distributions on the variance parameters for $K = 1, \ldots, 6$.

K	$\mathsf{E}(d_{i,i',j})$	$\mathsf{SD}(d_{i,i',j})$	b_ζ	b_θ	b_σ	b_κ	v_ζ	v_θ	v_ν
1	0.377	0.286	1.508	2.448	0.074	0.074	0.868	1.106	0.192
2	0.592	0.308	2.367	1.546	0.074	0.074	1.088	0.879	0.192
3	0.752	0.317	3.009	1.066	0.074	0.074	1.227	0.730	0.192
4	0.885	0.321	3.541	0.740	0.074	0.074	1.331	0.608	0.192
5	1.003	0.323	4.010	0.491	0.074	0.074	1.416	0.496	0.192
6	1.109	0.327	4.434	0.290	0.074	0.074	1.489	0.381	0.192

of multilayer network datasets. In the absence of prior information, we set $m_\theta = m_\zeta = 0$ and $\boldsymbol{m}_{\boldsymbol{v}} = \mathbf{0}$ and $\mathbf{V}_{\boldsymbol{v}} = v_\nu^2\mathbf{I}$ in order to center the model, roughly speaking, around an Erdös-Rényi model Erdös and Rényi (1959), and also ensure that the prior distributions are invariant to rotations of the latent space (see also Section 3.1).

Now, we establish some constraints that allow us to appropriately contrast models constructed with different values of K. Thus, for the prior distributions on the variance parameters, we naturally let $a_\zeta = a_\theta = a_\sigma = a_\kappa = 3$, which leads to a proper prior with finite moments and a coefficient of variation equal to 1. Under this set up, it can be shown that marginally

$$\mathsf{Var}(\zeta_j) = \frac{b_\zeta}{2} + v_\zeta^2 \ , \qquad \mathsf{Var}(\theta_j) = \frac{b_\theta}{2} + v_\theta^2 \ , \qquad \mathsf{Var}(u_{i,j,k}) = \frac{b_\sigma}{2} + \frac{b_\kappa}{2} + v_\nu^2 \ ,$$

each of which we split equally among all terms. First, resembling a regular LPM, we set $\mathsf{Var}(u_{i,j,k}) = 1/9$ a priori, such that $b_\sigma = b_\kappa = 2/27$ and $v_\nu^2 = 1/27$. Then, from a naive application of the delta method we obtain that

$$\mathsf{Var}(\theta_j) \approx 2\log\left(-\frac{\Phi^{-1}(\vartheta_0)}{\mathsf{E}\left(\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|\right)}\right) ,$$

where ϑ_0 is the prior probability of observing an edge between any two actors (which can be tuned to reflect prior information), and $\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \sim \mathsf{N}_K(\mathbf{0}, \frac{2}{9}\mathbf{I})$ as long as $\mathsf{Var}(u_{i,j,k}) = \mathsf{Var}(\eta_{i,k}) = 1/9$. In our experiments, we set $b_\theta = \mathsf{Var}(\theta_j)$ and $v_\theta^2 = \mathsf{Var}(\theta_j)/2$ with $\vartheta_0 = 0.1$. Finally, we set $b_\zeta = \mathsf{Var}(\zeta_j)$ and $v_\zeta^2 = \mathsf{Var}(\zeta_j)/2$ with $\mathsf{Var}(\zeta_j) = 4\mathsf{E}(\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|)$, which allow a wide range of values of ζ_j.

Table 1 displays specific hyperparameter values for $K = 1, \ldots, 6$. In addition, Fig. 3 shows histograms of 10,000 independent realizations from the induced marginal prior distribution of the interaction probabilities, $\vartheta_{i,i',j}$, for several values of K. Note that these distributions are quite similar, exhibiting a mode at $\vartheta_{i,i',j} = 0.1$ (as expected) and then a somewhat uniform behavior with a slight peak towards $\vartheta_{i,i',j} = 1$.

3. Computation

For a given latent dimension K, the posterior distribution $p(\boldsymbol{\Upsilon} \mid \mathbf{Y})$ is explored using Markov chain Monte Carlo methods (MCMC; e.g., Gamerman and Lopes (2006)). The computational algorithm entails a combination of Gibbs sampling and Metropolis-Hastings steps (e.g., Haario et al. (2001)). Details about the MCMC algorithm can be found in the Appendix A. In the following sections we discuss the issues of identifiability and model selection which are common in the latent space modeling framework.

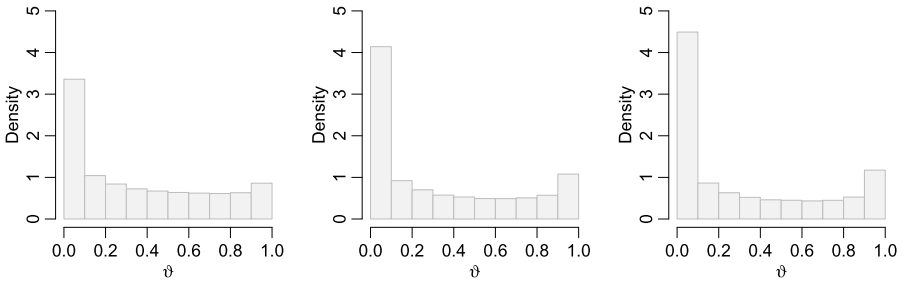
3.1. Identifiability

Our proposed MNLPM inherits the property of invariance to rotations and reflections of the social space from the simple latent space model of Hoff et al. (2002). Indeed, for any $K \times K$ orthogonal matrix $\mathbf{Q}$, the likelihood associated with the reparameterization $\tilde{\boldsymbol{u}}_{i,j} = \mathbf{Q}\boldsymbol{u}_{i,j}$ is independent of $\mathbf{Q}$ since $\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\| = \|\tilde{\boldsymbol{u}}_{i,j} - \tilde{\boldsymbol{u}}_{i',j}\|$. This lack of identifiability does not affect our ability to make inferences on the interaction probabilities $\vartheta_{i,i',j}$ s (which are identifiable). However, it does hinder our ability to provide posterior estimates of latent-position based measures (e.g., network correlation), including the latent positions themselves.

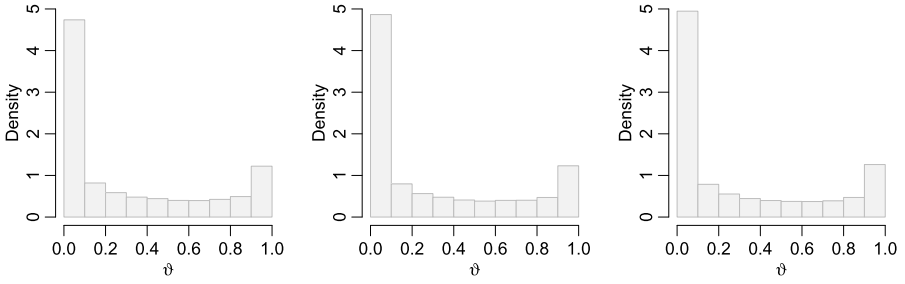
We address this invariance issue using a common parameter expansion used by Hoff et al. (2002) and many others. In particular, the problem of identifiability is addressed through a post-processing step in which the B posterior samples are rotated/reflected to a shared coordinate system. For each sample $\boldsymbol{\Upsilon}^{(b)}$, for $b = 1, \ldots, B$, an orthogonal transformation matrix $\mathbf{Q}^{(b)}$ is obtained by minimizing the Procrustes distance,

$$\tilde{\mathbf{Q}}^{(b)} = \underset{\mathbf{Q} \in \mathcal{S}^K}{\arg\min}\, \mathrm{tr}\left\{\left(\mathbf{E}^{(1)} - \mathbf{E}^{(b)}\mathbf{Q}\right)^\top \left(\mathbf{E}^{(1)} - \mathbf{E}^{(b)}\mathbf{Q}\right)\right\} , \tag{5}$$

where $\mathcal{S}^K$ denotes the set of $K \times K$ orthogonal matrices and $\mathbf{E}^{(b)}$ is the $I \times K$ matrix whose I rows correspond to the transpose of $\boldsymbol{\eta}_1^{(b)}, \ldots, \boldsymbol{\eta}_I^{(b)}$. The minimization problem in (5) can be easily solved using singular value decompositions (e.g.,



(a) $K = 1$ (b) $K = 2$ (c) $K = 3$



(d) $K = 4$ (e) $K = 5$ (f) $K = 6$

Fig. 3. Marginal prior distribution of the interaction probabilities $\vartheta_{i,i',j}$ for $K = 1, \ldots, 6$.

see Borg and Groenen (2005)). Indeed, $\tilde{\mathbf{Q}}^{(b)} = \mathbf{R}^{(b)}\mathbf{L}^{(b)\mathsf{T}}$, where $\mathbf{L}^{(b)}\mathbf{D}^{(b)}\mathbf{R}^{(b)\mathsf{T}}$ is the singular value decomposition of $\mathbf{E}^{(1)\mathsf{T}}\mathbf{E}^{(b)}$. Once the matrices $\tilde{\mathbf{Q}}^{(1)}, \ldots, \tilde{\mathbf{Q}}^{(B)}$ have been obtained, posterior inference for the latent positions of our model is based on the transformed coordinates $\tilde{\boldsymbol{u}}_{i,j}^{(b)} = \tilde{\mathbf{Q}}^{(b)}\boldsymbol{u}_{i,j}^{(b)}$ and $\tilde{\boldsymbol{\eta}}_i^{(b)} = \tilde{\mathbf{Q}}^{(b)}\boldsymbol{\eta}_i^{(b)}$.

3.2. Model selection

In general, setting $K = 2$ for the dimension of the latent space is a sensible choice, as it simplifies the visualization and description of social relationships. However, our objective goes beyond a mere description of multilayer network data, and in consequence the value of K plays a critical role in the results.

The network literature has largely focused on the Bayesian Information Criterion (BIC; e.g., Hoff (2005), Handcock et al. (2007), Airoldi et al. (2009)). However, the BIC is often inappropriate for hierarchical models since the hierarchical structure implies that the effective number of parameters will be typically less than the actual number of parameters in the likelihood. An alternative to the BIC is the Watanabe-Akaike Information Criterion (WAIC; Watanabe (2010), Watanabe (2013), Gelman et al. (2014)),

$$\text{WAIC}(K) = -2\sum_{j}\sum_{i,i':i<i'} \log \mathsf{E}\left[p\left(y_{i,i',j} \mid \boldsymbol{\Upsilon}\right)\right] + 2p_{\text{WAIC}},$$

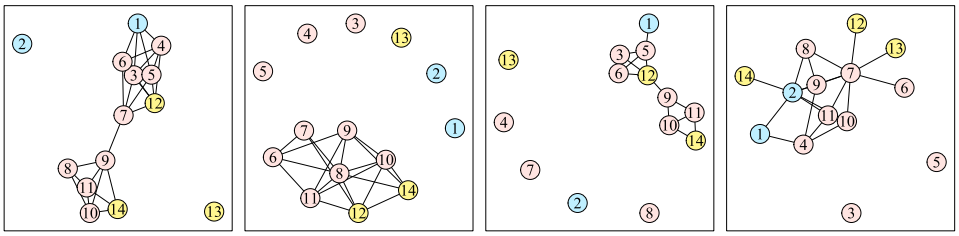
where

$$p_{\text{WAIC}} = 2\sum_{j}\sum_{i,i':i<i'} \left(\log \mathsf{E}\left[p\left(y_{i,i',j} \mid \boldsymbol{\Upsilon}\right)\right] - \mathsf{E}\left[\log p\left(y_{i,i',j} \mid \boldsymbol{\Upsilon}\right)\right]\right),$$

is the model complexity (effective number of parameters), and $\boldsymbol{\Upsilon}$ is the set of model parameters assuming that the dimension of the social space is K. The expected values in these expressions are calculated with respect to the posterior distribution $p(\boldsymbol{\Upsilon} \mid \mathbf{Y})$, and can be approximated by averaging over the MCMC samples $\boldsymbol{\Upsilon}^{(1)}, \ldots, \boldsymbol{\Upsilon}^{(B)}$.

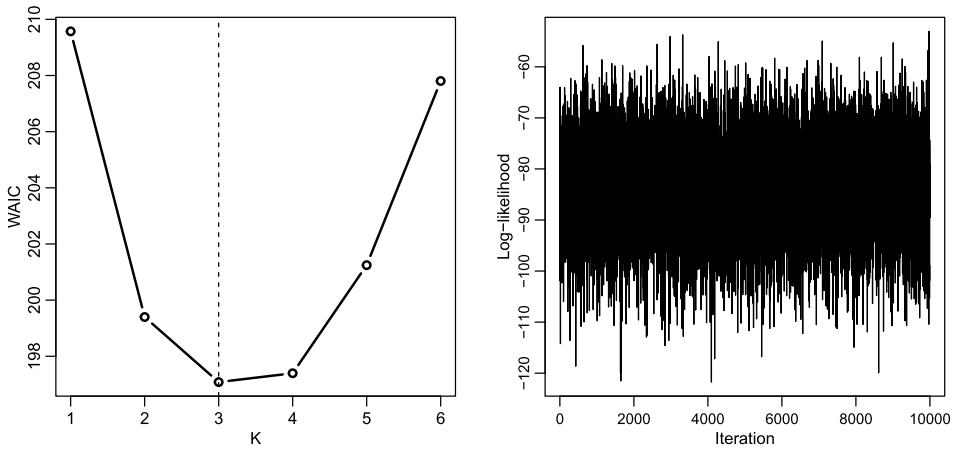
4. Illustrations

In this section, we illustrate and evaluate the performance of our MNLPM using two benchmark data sets: The bank wiring room data of Roethlisberger and Dickson (2003) and the friendship cognitive social structure of Krackhardt (1987). The main contributions involving the characterization of the social dynamics with our MNLPM are formal approaches to measure network correlation and perceptual agreement, which are respectively illustrated with each data set.



(a) Horseplay (b) Arguments (c) Frienship (d) Antagonism

Fig. 4. Visualization of the bank wiring room data. Inspectors are shown in blue, solderers in yellow, and wiremen or assemblers in red. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)



(a) Information criterion (b) Log-likelihood, $K = 3$

Fig. 5. WAIC values to select the latent dimension K of the social space for the Bayesian analysis of the bank wiring room data using our MNLPM, and log-likelihood chain associated with the value of K optimizes the WAIC ($K = 3$).

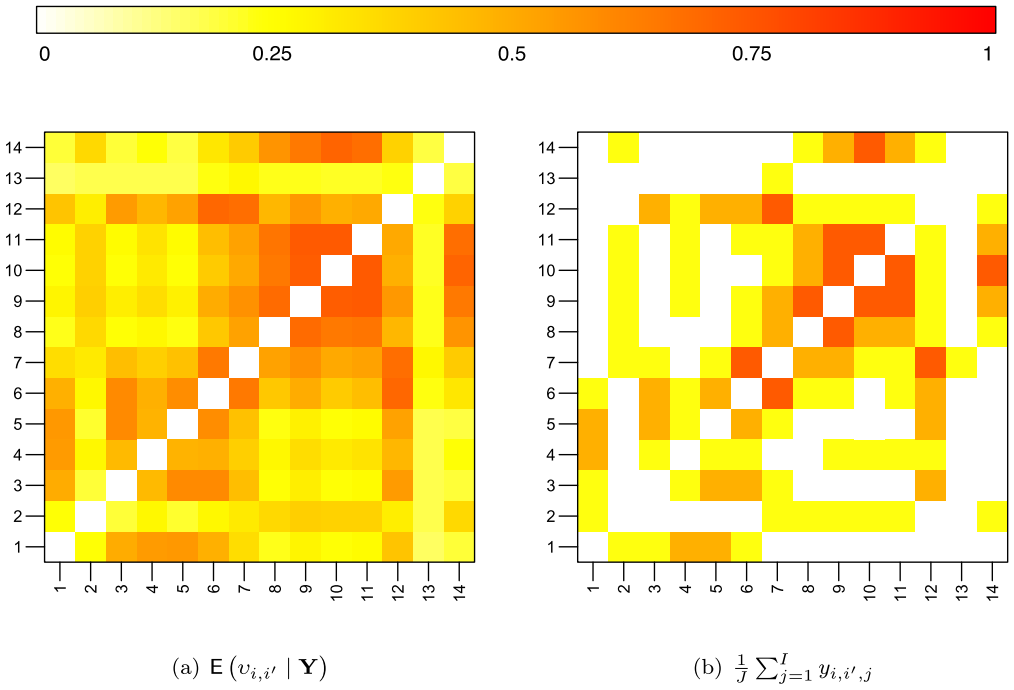
4.1. Bank wiring room data

These are the observational data on $I = 14$ Western Electric (Hawthorne Plant) employees from the bank wiring room presented in Roethlisberger and Dickson (2003). The employees worked in a single room and include two inspectors (actors 1 and 2), three solderers (actors 12, 13, and 14), and nine wiremen or assemblers (actors 3 to 11). The authors gathered data about $J = 4$ symmetric interaction categories including: participation in horseplay (Horseplay, network 1), participation in arguments about open windows (Arguments, network 2), friendship (Friendship, network 3), and antagonistic behavior (Antagonism, network 4). This dataset is considered nowadays as a standard referent to test models for multilayer network data (e.g., Bartz-Beielstein et al. (2014), Liu (2020), and Abdollahpouri et al. (2020)). Fig. 4 displays a visualization of all the relational layers.

We implement our MNLPM for the bank wiring room data following the computational approach described in Section 3. The results presented below are based on $B = 10,000$ samples of the posterior distribution obtained after thinning the original Markov chains every 10 observations and a burn-in period of 100,000 iterations. All chains mix reasonably well. Left panel of Fig. 5 shows the WAIC computed for several MNLPMs fitted using a range of latent dimensions of the social space. This criterion clearly supports $K = 3$ (WAIC $= 197.1$) as an optimal choice, which is the latent dimension we use in all our analyses henceforth. The effective sample sizes of the model parameters following the MCMC algorithm discussed above range from 4,081 to 10,000. Right panel of Fig. 5 displays the log-likelihood chain associated with the latent dimension that optimizes the WAIC, which shows no signs of lack of convergence.

4.1.1. Consensus network

Unlike other models for multilayer network data (e.g., Salter-Townshend and McCormick (2017)), our approach (as well as those by Gollini and Murphy (2016) and D'Angelo et al. (2019), for instance, as a direct consequence of the social space underlying multidimensional relations) provides a straightforward mechanism to construct a "consensus network". Indeed, the average positions $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_I$ can be used to *induce* a weighted network, given by $\upsilon_{i,i'} = \Phi\left(\mu_\zeta - e^{\mu_\theta} \| \boldsymbol{\eta}_i - \boldsymbol{\eta}_{i'} \|\right)$, that "collapses" all the relational layers in a single network by weighting them according to the mean parameters of the



(a) $\mathsf{E}\left(\upsilon_{i,i'} \mid \mathbf{Y}\right)$ (b) $\frac{1}{J}\sum_{j=1}^{J} y_{i,i',j}$

Fig. 6. Consensus network estimates for the bank wiring room data. The left panel provides the posterior mean under our MNLPM, and the right panel shows the proportion of observed links across networks.

hierarchical prior distribution. The consensus network can be very useful when an overall summary of the social dynamics is required.

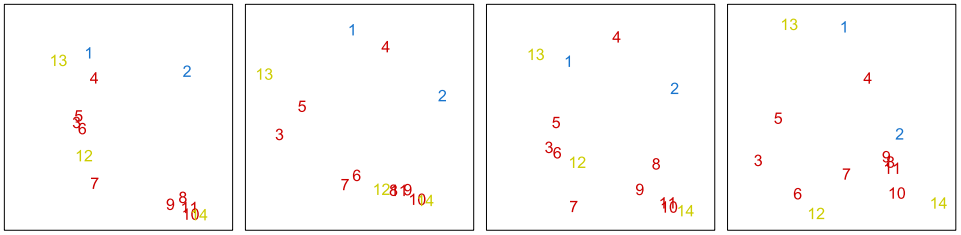
An alternative way to obtain a consensus network involves computing an estimate of the interaction probabilities by modeling an aggregation of the data with a LPM. However, in many situations this approach is not the best course of action. First, such an strategy requires the definition of an aggregate adjacency matrix, say $\tilde{\mathbf{Y}} = [\tilde{y}_{i,i'}]$, defined as $\tilde{y}_{i,i'} = 1$ if $\frac{1}{J}\sum_{j=1}^{J} y_{i,i',j} > \delta_0$, and $\tilde{y}_{i,i'} = 0$ otherwise, where δ_0 is a fixed (but arbitrary) threshold (see Krackhardt (1987) for more aggregations). Any choice of δ_0 may be misleading because it can hide away (highlight) relevant (irrelevant) relationships (e.g., we would face a loss of useful information if $\frac{1}{J}\sum_{j=1}^{J} y_{i,i',j}$ is slightly less than δ_0). On the other hand, fitting a LPM using aggregated data may fail to capture fundamental features of consensus social roles due to the lack of means to account for variability within and between layers. In this spirit, our approach depends on specific parameters for that purpose after differentially modeling each layer and borrowing information across them (hence the pertinence of our hierarchical formulation), which lead us to the direct interpretation of $\upsilon_{i,i'}$ as the global "affinity" or "social distance" between actors i and i' in social space, operationalizing the notion of social roles (for this last interpretation, see Hoff (2005) and Hoff (2009)).

Fig. 6 displays heat-maps for the matrix of posterior means given by $\hat{\upsilon}_{i,i'} = \mathsf{E}\left(\upsilon_{i,i'} \mid \mathbf{Y}\right)$ and the proportion of observed links across networks, $m_{i,i'} = \frac{1}{J}\sum_{j=1}^{J} y_{i,i',j}$. Although the estimate provided by our model is "denser" than the empirical proportion (probably due to shrinkage), note that the estimates are similar. As a matter of fact, the absolute distance between the two matrices, which is bounded by $I(I-1)/2 = 91$, is $\sum_{i:i<i'} |\hat{\upsilon}_{i,i'} - m_{i,i'}| = 17.77$. This also suggests that the model correctly characterizes the data generating process.

4.1.2. Projections in social space

Actor-specific latent positions $\boldsymbol{u}_{1,1}, \ldots, \boldsymbol{u}_{I,J}$ provide a powerful tool for describing social interactions. Fig. 7 shows Procrustes-transformed latent position estimates $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{i,j} \mid \mathbf{Y}\right)$ along the two dimensions with the highest variability for each layer of the system. Even though the social behavior is similar across layers, there are important particularities. First, note that the social dynamics of Horseplay and Friendship are quite similar, except that in Horseplay latent positions seem to be more clustered together. Furthermore, many social patterns are evident. For instance, actors who have close positions in Horseplay and Friendship, typically have distant positions in Antagonism; such an effect is particularly clear among actors 3 and 6, and actors 10 and 14. Lastly, complementing the insights provided by the consensus network, average positions $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_I$ can be also very useful to perform an "global" visualization of the social dynamics.

As pointed out above, latent positions allow us to distinguish groups of actors that fulfill similar social roles. In order to identify such clusters, we can either apply some unsupervised clustering technique (e.g., hierarchical clustering, k-means clustering) or include directly into the model a set of parameters that assign actors to groups. The latter is quite preferable



(a) Horseplay (b) Arguments (c) Frienship (d) Antagonism

Fig. 7. Posterior means of Procrustes-transformed latent positions $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{i,j} \mid \mathbf{Y}\right)$ along the two dimensions with the highest variability, for the bank wiring room data. Inspectors are shown in blue, solderers in yellow, and wiremen or assemblers in red.

since the uncertainty related to the clustering task can be directly quantified along with its relationship to other model parameters (see Section 7 for more details).

4.1.3. Network correlation

Another key feature of the MNLPM is that it *implicitly* allows us to obtain correlation measures between layers as a direct by-product of the model parameterization. Before getting into details, we highlight that, unlike our proposal, the model in Salter-Townshend and McCormick (2017) allows the user to measure network correlation *explicitly* through specific correlation parameters given in the multivariate Bernoulli distribution. Unfortunately, our sampling distribution lacks such parameters. Therefore, we simply employ a well-defined function of our parameter space, by exploiting the arrangement of subject-layer specific latent positions (which are embedded in a hierarchical framework as opposed to be generated independently from each network) to define a measure of network correlation, relying on the premise that networks with matching latent configurations over a shared coordinate system can be regarded as similar. Thus, we define the correlation between layers j and j', for $j, j' = 1, \ldots, J$, as

$$\rho_{j,j'} = \mathsf{cor}(u^*_{1,j}, \ldots, u^*_{I,j}; u^*_{1,j'}, \ldots, u^*_{I,j'}),$$

where $u^*_{i,j}$ is the maximum Procrustes-transformed latent characteristic across latent dimensions of actor i in layer j, i.e., $u^*_{i,j} = \max\{\tilde{u}_{i,j,1}, \ldots, \tilde{u}_{i,j,K}\}$ (alternative definitions for $\rho_{j,j'}$ are possible; e.g., by considering the median instead of the maximum). Moreover, similar arguments can be made using latent distances instead of latent positions alone (see also Salter-Townshend and McCormick (2017) for a brief discussion), and therefore, the correlation between layers j and j' can also be alternatively specified as

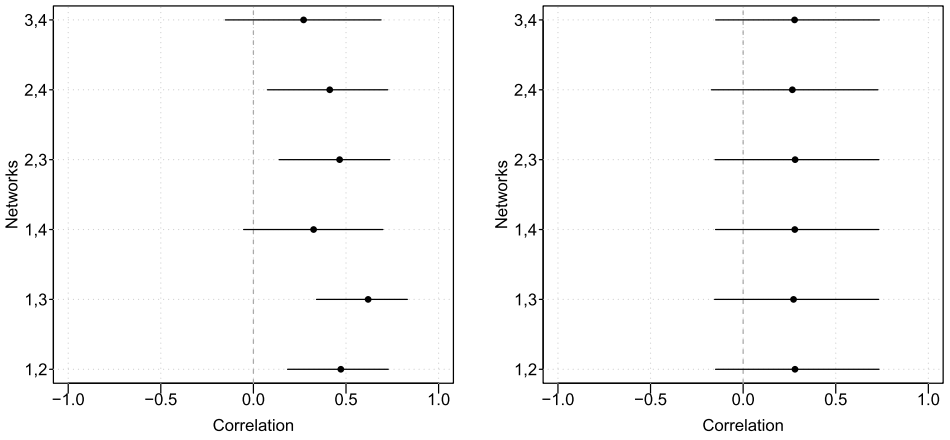
$$\rho_{j,j'} = \mathsf{cor}(d_{1,2,j}, \ldots, d_{I-1,I,j}; d_{1,2,j'}, \ldots, d_{I-1,I,j'}), \tag{6}$$

where $d_{i,i',j} = \| \boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \|$ is the latent distance between actors i and i' in network j. After extensive experimentation, we found that this alternative is preferable because it relies on identifiable quantities (see Section 3.1), and most importantly, it provides more accurate estimates with narrower credible intervals than its counterpart. Almost identical results are found if latent distances in (6) are replaced by either linear predictors or interaction probabilities. Finally, notice that either way, our approach represents the network correlation after accounting jointly for social structure encoded within each layer, thanks to the MNLPM's hierarchical specification.

In order to obtain a credible interval for $\rho_{j,j'}$, we simply compute the corresponding percentiles of $\rho^{(1)}_{j,j'}, \ldots, \rho^{(B)}_{j,j'}$, which are a straightforward byproduct of B samples drawn from the posterior distribution using MCMC methods (recall our discussion in Section 3). On the other hand, notice that a credible interval for $\rho_{j,j'}$ can be used as a formal way to test for statistical significance about the correlation between layers j and j'. As a matter of fact, if such an interval does not include zero, then we would have enough evidence to declare a significant result, since the most plausible values for $\rho_{j,j'}$ indicate that the correlation could be either lower or higher compared to zero.

Left panel in Fig. 8 shows credible intervals along with point estimates for all pairwise network correlations computed using expression (6). We see that all pair of layers are positively correlated, except for horseplay and antagonistic behavior (networks 1 and 4) as well as friendship and antagonistic behavior (networks 3 and 4). Furthermore, a positive correlation is particularly evident between Horseplay and Friendship (networks 1 and 3). Such findings are quite consistent with the social dynamics described above.

Lastly, we test our correlation approach by considering a set of independent networks with no underlying structure. To do so, we independently generate $J = 4$ Erdös-Rényi networks with $I = 14$ actors and interaction probability 0.1, and then fit the MNLPM in order to obtain the pairwise network correlations. Right panel in Fig. 8 present the corresponding inference for these quantities. We see that the correlation between every pair of layers is negligible since the credible intervals are centered at zero, which is the expected behavior for data with no correlation structure.



(a) Bank wiring room data (b) Synthetic data

Fig. 8. 95% credible intervals and posterior means for all pairwise network correlations $\rho_{j,j'}$. Left panel: Bank wiring room data (participation in horseplay, network 1; participation in arguments about open windows, network 2; friendship, network 3; and antagonistic behavior, network 4). Right panel: Synthetic data.

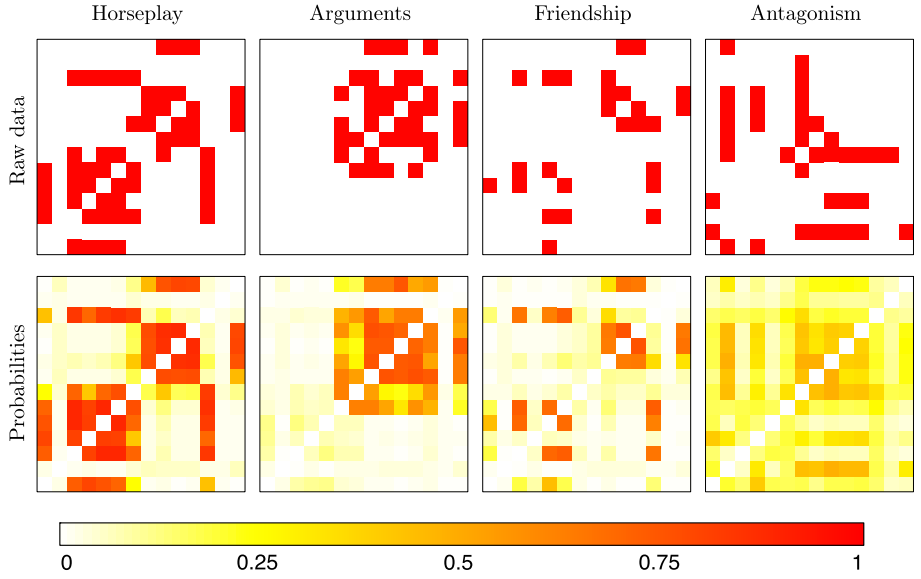


Fig. 9. Multilayer network data $y_{i,i',j}$ and probability of interaction posterior means $\mathsf{E}\left(\vartheta_{i,i',j} \mid \mathbf{Y}\right)$, for the bank wiring room data.

4.1.4. Model fit

We assess our MNLPM fit using both in-sample and out-of-sample metrics. Our in-sample assessment relies on two approaches. First, we compare the observed data $y_{i,i',j}$ against the corresponding probability of interaction posterior means $\mathsf{E}\left(\vartheta_{i,i',j} \mid \mathbf{Y}\right)$. We see in Fig. 9 that point estimates concur with the raw data, which clearly suggests that the model fits the data well in terms of reproducibility.

Now, following Gelman et al. (2013), we further explore the in-sample fit of each model by replicating pseudo-data from the fitted model and calculating a variety of summary statistics for each sample, whose distributions are then compared against their values in the original sample. Fig. 10 shows credible intervals along with point estimates for a set of relevant network measures, including the density, assortativity, and clustering coefficient, among others (see Kolaczyk and Csárdi (2014) for details about these structural summaries). Note that the model appropriately captures these structural features (perhaps with the exception of the assortativity in Antagonism), since observed values belong to the corresponding credible intervals; even most of the estimates virtually coincide with the observed values. Thus, pseudo-data generation also provides evidence of proper in-sample properties in favor of our model. Finally, the out-of-sample predictive performance of the model is presented in Section 5.

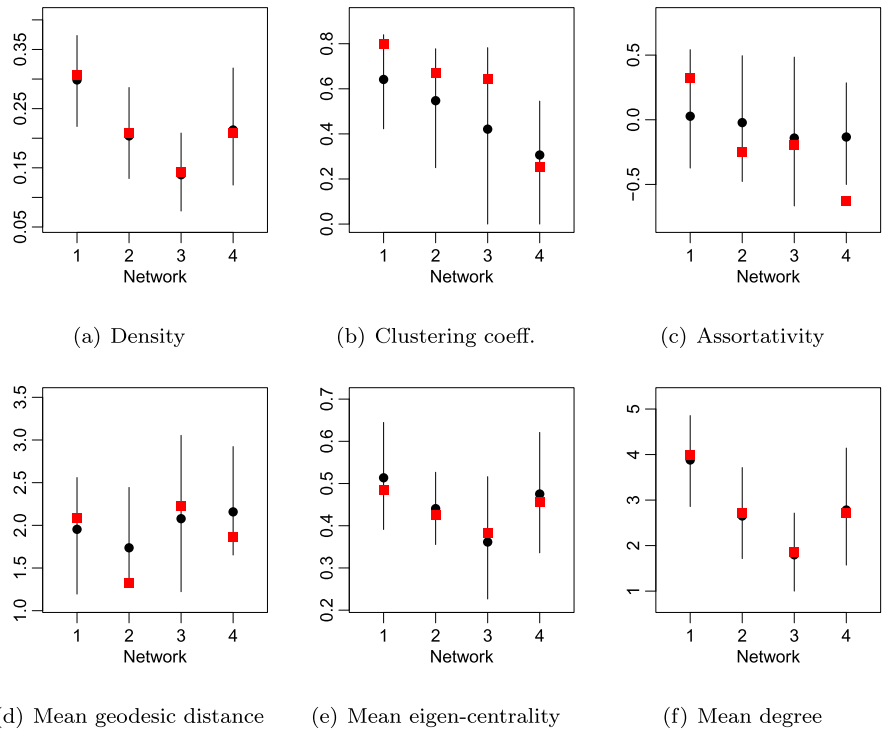


Fig. 10. 95% credible intervals, posterior means (black circle), and observed values (red square) associated with the empirical distribution of a battery of summary statistics, based on 10,000 replicas of the bank wiring room dataset (participation in horseplay, network 1; participation in arguments about open windows, network 2; friendship, network 3; and antagonistic behavior, network 4).

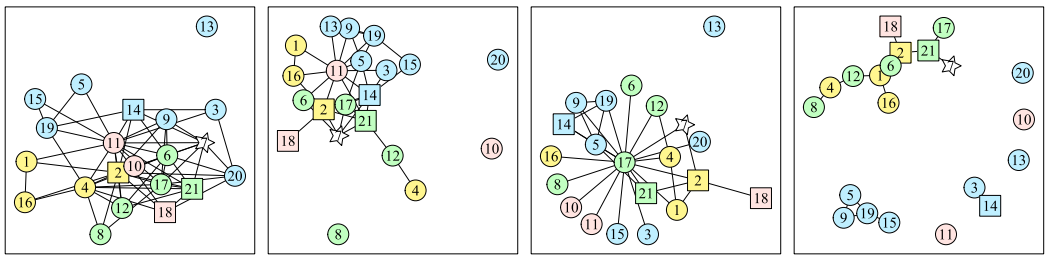
4.2. Friendship data

In this section we develop a formal test to assess the level of "agreement" (as opposed to "accuracy", which requires the definition of an external "gold standard"), between an actor's self perception of their own position in a social environment and that of other actors embedded in the same system (e.g., Swartz et al. (2015), Sewell (2019), Sosa and Rodriguez (2021)). Our approach relies on the hierarchical structure of the MNLPM, which allows us to define a measure of cognitive agreement. The posterior distribution of such a measure makes possible to identify those individuals whose position in the social space agrees with the judgments of other actors.

A cognitive social structure (CSS) is defined by a set of cognitive judgments that subjects form about the relationships among actors (themselves as well as others) who are embedded in a common environment. Hence, each subject reports a full description of the social network structure. We consider a CSS reported by Krackhardt (1987) in which $I = 21$ management personnel in a high-tech machine manufacturing organization were observed in order to evaluate the effects of a recent management intervention program. Each person was asked to fill out a questionnaire indicating not only who he/she believes his/hers friends are, but also his/her perception of others friendships. Thus, we have a collection composed of $J = 21$ undirected binary networks $\mathbf{Y}_1, \ldots, \mathbf{Y}_I$, with $\mathbf{Y}_j = [y_{i,i',j}]$, defined over a common set of $I = 21$ actors, such that $y_{i,i',j} = 1$ if i and i' are friends of each other, and $y_{i,i',j} = 0$ otherwise. Some attribute information about each executive was also available, including corporate level (president, vice-president, or general manager), and department membership (there are four departments labeled from 1 to 4; the CEO is not in any department). Such information can potentially be included in the analysis (see Section 7 for details).

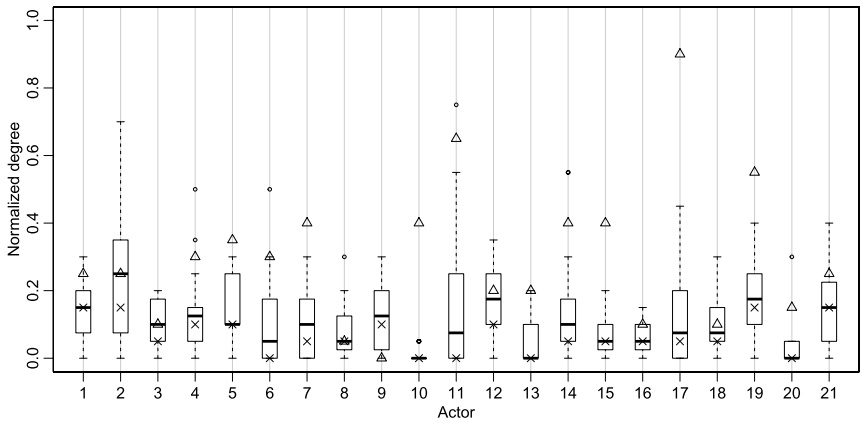
Part of these multilayer network data along with the "consensus" network are represented in Fig. 11. In this context, there is a link present between two actors according to the consensus if at least half of the personnel have reported that link. Note that even though the variability on the perceptions is not negligible, there are some commonalities across networks. For instance, more than half of the management personnel believes that actors 2 and 18 (both vice-presidents), actors 21 (vice-president) and 17 (manager) both in department 2, and actors 14 (vice-president) and 3 (manager) both in department 3, are friends. Furthermore, managers 3, 8, 9, and 20 only report less than six relations each, and manager 9, who only recognizes four friendships, is the only executive that considers himself with no friends. Senior executives (president and vice-presidents) report networks with more than 20 connections each.

Now, we describe the "popularity" of each executive in terms of how connected they are. Top panel in Fig. 12 summarizes the degree distribution of actors across networks. In general, we see that all the executives perceive themselves as more popular than what they actually are according to the general opinion, with the exception of actors 8 and 9. On the other

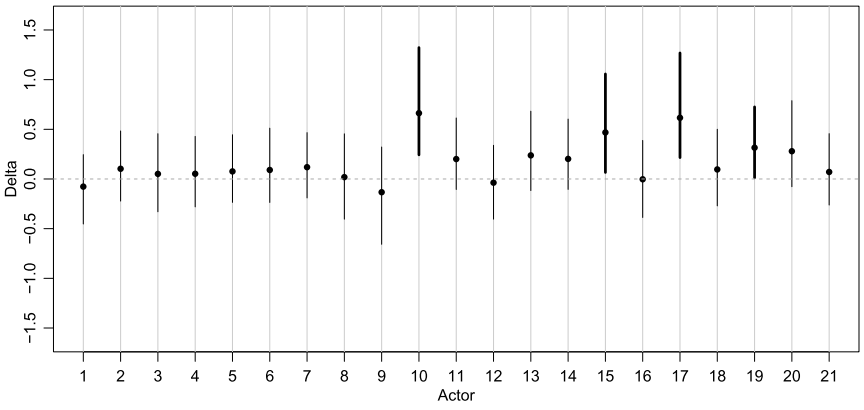


(a) Actor 7 (b) Actor 14 (c) Actor 17 (d) Consensus

Fig. 11. Visualization of some networks in the friendship CSS data corresponding to actors 7, 14, and 17, along with the consensus network. Vertex shape indicates the executive's level in the company (star: president, actor 7; square: vice-presidents, actors 2, 14, 18, 21; and circles: managers), whereas vertex color indicates the executive's department in the company (the president does not belong to any department).



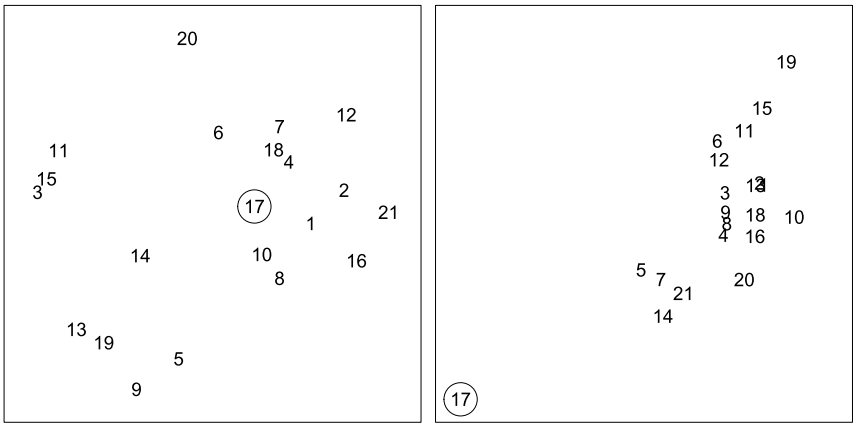
(a) Normalized degree



(b) Assessment index

Fig. 12. Top panel: Normalized degree distribution across networks. The i-th boxplot summarizes the distribution of the degree for all reporters except i, while the self-perceived degree is represented by a triangle ($\triangle$) and the respective degree in the consensus network by a cross ($\times$). Bottom panel: 95% credible intervals and posterior means for the distribution of the personal assessment parameters δ_i. Thicker lines correspond to credible intervals that do not contain zero. The normalized degree centrality is defined as the degree divided by the maximum possible degree.

hand, actors 1, 2, 19, and 21 are perceived by the personnel as the executives with most connections. We also highlight the case of actors 10 and 17 who are highly egocentric in comparison with the consensus; in particular, actor 10 is perceived with no friends at all, but this actor believes just the opposite.



(a) $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{i,17} \mid \mathbf{Y}\right)$ (b) $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{17,i} \mid \mathbf{Y}\right)$

Fig. 13. Posterior means of procrustes-transformed latent positions along the two dimensions with highest variance for actor 17 (circled). Left panel: As perceived by actor 17, $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{i,17} \mid \mathbf{Y}\right)$. Right panel: As perceived by all actors, $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{17,i} \mid \mathbf{Y}\right)$.

Table 2
Multilayer network datasets for which a series of cross-validation experiments are performed using independently fitted LPMs (baseline) and our MNLPM. Note that wiring and tech are widely analyzed in Section 4. Also, micro corresponds to the network data of village number 10.

Acronym	Reference	Actors	Layers	Edges
wiring	Roethlisberger and Dickson (2003)	14	4	79
tech	Krackhardt (1987)	21	21	550
seven	Vickers and Chan (1981)	29	3	222
girls	Steglich et al. (2006)	50	3	119
aarhus	Magnani et al. (2013)	61	5	620
micro	Banerjee et al. (2013)	77	6	903

4.2.1. Perception assessment

We consider the agreement question in which we ask whether an individual's perception of their relationships is the same as the perception that others hold. To answer this question, we define the assessment parameter δ_i, for $i = 1, \ldots, I$, as the difference between subject i's self-assessment and the mean assessment of subject i by others, i.e.,

$$\delta_i = \left\| \tilde{\boldsymbol{u}}_{i,i} \right\| - \left\| \frac{1}{I-1} \sum_{j \neq i} \tilde{\boldsymbol{u}}_{i,j} \right\|,$$

where $\tilde{\boldsymbol{u}}_{i,j}$ is the Procrustes-transformed version of $\boldsymbol{u}_{i,j}$. This quantity is an effort to parametrize the accuracy of self-assessment in perceiving ties.

Bottom panel in Fig. 12 provides credible intervals along with point estimates for the personal assessment parameters $\delta_1, \ldots, \delta_I$, based on $B = 10,000$ samples of the posterior distribution obtained after thinning the original Markov chains every 10 observations and a burn-in period of 100,000 iterations, associated with the value of K that optimizes the WAIC ($K = 6$). We see that most actors have a slightly elevated view of themselves in terms of their capacity to befriend others, whereas very few have a negative view. On the other hand, actors 10, 15, 17, and 19 have a significant inflated perception of their ability to form friendship ties. Note that the results of this test are quite consistent with the exploratory data analysis discussed previously.

Finally, in order to exemplify the social behavior for an unskillful actor in perceiving relations, we show in Fig. 13 Procrustes-transformed latent positions estimates along the two dimensions with highest variance for actor 17 (who clearly has a misleading view of his/hers surroundings according to the test), as perceived by this actor, $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{i,17} \mid \mathbf{Y}\right)$, and as perceived by all actors (including him/herself), $\mathsf{E}\left(\tilde{\boldsymbol{u}}_{17,i} \mid \mathbf{Y}\right)$. These plots are consistent with those from Fig. 12. Actor 17 see him/herself in quite a "central" position of the friendship relations; however, according with the general opinion, actor 17 is clearly isolated from the others. This is again consistent with the test showed in Fig. 12.

5. Prediction

As an additional goodness-of-fit assessment, we carry out cross-validation experiments on several multilayer network datasets (see Table 2) exhibiting different kinds of actors, sizes, and relations. More specifically, we performed a five-fold

Table 3
Mean and SD AUCs corresponding to the prediction of missing links in a series of CV experiments to assess the predictive performance of IFLPM, GMLPM, MNLPM, and LSJM, using each dataset provided in Table 2.

Dataset	Statistic	AUC			
		IFLPM	GMLPM	MNLPM	LSJM
wiring	Mean	0.905	0.823	0.895	0.928
	SD	0.031	0.031	0.029	0.068
tech	Mean	0.804	0.882	0.904	0.814
	SD	0.017	0.025	0.022	0.020
seven	Mean	0.888	0.957	0.958	0.977
	SD	0.006	0.015	0.014	0.016
girls	Mean	0.781	0.879	0.882	0.989
	SD	0.044	0.046	0.052	0.002
aarhus	Mean	0.922	0.926	0.949	0.917
	SD	0.021	0.013	0.009	0.015
micro	Mean	0.773	0.936	0.932	0.925
	SD	0.014	0.012	0.013	0.011

cross-validation (CV) in which five randomly selected subsets of roughly equal size in the dataset are treated as missing and then predicted using the rest of the data. Specifically, the relational values are partitioned into five subsets, and the data in those subsets are predicted after training the model on each of the complements of those subsets (e.g., Kolaczyk and Csárdi (2014)). We carry out such a prediction based on the posterior predictive mean of the missing edges conditional on the training data ignoring any sampling pattern, i.e., unobserved values are considered missing at random (MCA; e.g., Hoff (2008)). The case where the pattern of unobserved edges depends on the unobserved edges themselves is beyond the scope of this paper. Nonetheless, for what it's worth, all models are compared under the same framework. Under a MCA setting, handling missing data is straightforward under any MCMC sampling scheme, since the full conditional distributions for the model parameters remain unchanged using both observed data and current missing data values Rubin (1976). Thus, we adjust our sampling algorithms by introducing an additional step, in which we draw first missing edges from the sampling distribution (given the value of the rest of the parameters at the current iteration), and then, once the missing links have been sampled (and therefore the full dataset is complete), the rest of the algorithm proceeds naturally as before (e.g., Sewell and Chen (2015)).

We summarize our findings in Table 3, where we report the average area under the receiver operating characteristic curve (AUC) for each dataset described in Table 2. Specifically, we report the AUC for the model with the optimal value of K according to the WAIC criteria following the model fitting described above. The values correspond to the prediction of missing links using independently fitted LPMs (IFLPM), our MNLPM, and also, a variant of MNLPM that is very reminiscent of Gollini and Murphy (2016). The latter, referred to as GMLPM, considers unique latent positions with no hierarchical structure in such a way that $\vartheta_{i,i',j} = \Phi(\zeta_j - e^{\theta_j} \| \boldsymbol{u}_i - \boldsymbol{u}_{i'} \|)$. Furthermore, we also take into account the "exact" version of the latent space joint model (LSJM) by Gollini and Murphy (2016) using the `lvm4net` package (Gollini, 2019) that implements a variational EM algorithm Algorithm in order to estimate all the model parameters (we were very careful in tuning both hyperparameters as well as starting values through a comprehensive set of trials, since the algorithm is quite sensitive to these choices). In this context, the AUC is a measure of how well a given model is capable of predicting missing links (higher AUC values are better). We report the AUC for the models with the optimal value of K according to the WAIC criteria. As before, our predictions are based on $B = 10,000$ samples of the posterior distribution obtained after thinning the original Markov chains every 10 observations and a burn-in period of 100,000 iterations.

We see that both LSJM and MNLPM constitute the best alternatives in terms of predictive performance. Specifically, the out-of-sample performance of IFLPM and NMLPM is practically the same for wiring, as well as that of GMLPM and MNLPM for seven, girls, and micro. For all the other datasets, LSJM and MNLPM have a better predictive behavior than their competitors (excluding LSJM in the case of tech). Such an effect is particularly clear when fitting MNLPM as opposed to IFLPM, which provides even more evidence about why considering our hierarchical prior as in MNLPM is beneficial. On the other hand, not surprisingly. These results strongly suggest that the predictive potential as well as the inner flexibility of MNLPM are indeed comparable with or even better than those offered by alternative approaches.

6. Simulation study

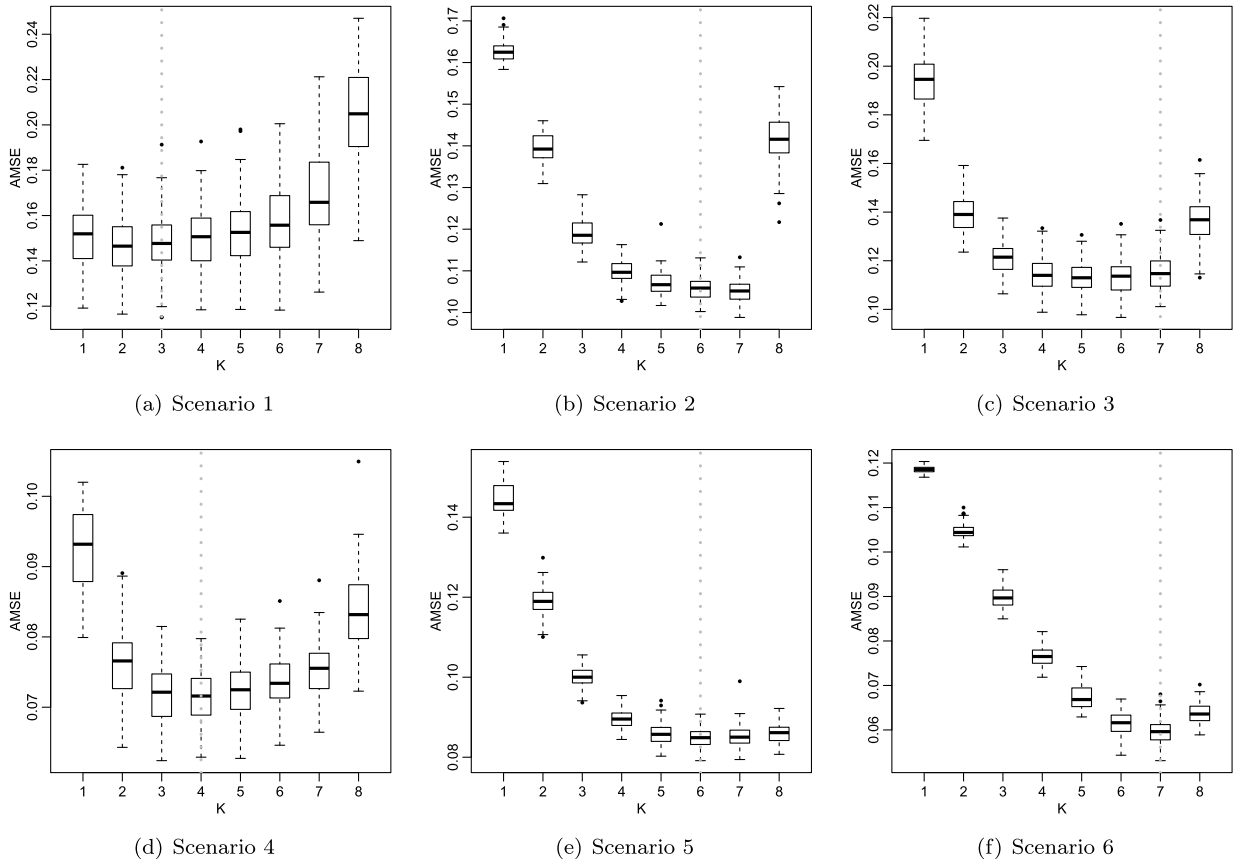
6.1. Selection of the latent dimension

When fitting MNLPM, we treat the latent dimension K as known, but in practice, we typically face the model selection problem (recall our discussion in Section 3.2). Therefore, it is pertinent to study the performance of the WAIC in choosing the optimal value of K as well as the impact of K on model parameter estimation.

To this end, we examine several scenarios. Under the first scenario, 100 synthetic datasets are generated from the sampling distribution given in (1), by letting the number of actors and layers be as in the wiring dataset, and setting the true

Table 4
Percentage of times a latent dimension K is chosen according to the WAIC, and mean running times (in seconds) using a eight core of an AMD A12-9730P processor, for 100 synthetic datasets independently generated from MNLPM, when drawing 10,000 samples of the posterior distribution of MNLPM obtained after a burn-in period of 10,000 iterations, under six simulation scenarios.

Scenario	Seed data	Time	True K	Latent dimension K							
				1	2	3	4	5	6	7	8
1	wiring	130.5	3	1	40	41	13	5	0	0	0
2	tech	819.3	6	0	0	0	0	15	64	21	0
3	seven	255.4	7	0	0	1	3	17	25	54	0
4	girls	583.4	4	1	0	34	39	19	5	2	0
5	aarhus	1339.3	6	0	0	0	0	4	56	40	0
6	micro	2561.9	7	0	0	0	0	0	11	83	6



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

(d) Scenario 4 (e) Scenario 5 (f) Scenario 6

Fig. 14. AMSE empirical distributions for interaction probabilities when fitting MNLPM as described in Table 4, across a range of values of latent dimensions, under six simulation scenarios. The dashed, vertical line indicates the true value of K.

interaction probabilities to the corresponding posterior means obtained in our analysis of this seed dataset in Section 4.1 (notice that in doing so, we are implicitly fixing the true value of K to the optimal value suggested by the information criteria when the model was originally fitted). Similarly, in scenarios 2 to 6, the data are simulated taking as seeds all those datasets provided in Table 2, respectively. Thus, as opposed to assuming that K is known a priori, we use the WAIC to select the optimal value of K to fit MNLPM each time (100 times under each scenario). Then, once MNLPM is fitted using such latent dimension, we compute the average mean square error (AMSE) for the estimates of the interaction probabilities.

Table 4 shows the percentage of times a latent dimension K is labeled as optimal according to the WAIC. In general, we see that the information criteria tend to suggest the right value for K. However, there is a propensity to select a latent dimension either one unit smaller or higher than the true value. Either way, given the results presented in Fig. 14, we believe that the WAIC constitutes a powerful alternative to select a reasonable latent dimension, since there are no apparent differences between the chosen values of K in terms of parameter estimation. Indeed, the values that turned out to be favored by the information criteria minimize AMSE across dimensions.

Table 5
AMSE empirical distribution summaries for interaction probabilities and latent positions, across a range of prior interaction probability values, for 100 synthetic datasets independently generated from MNLPM under fixed latent dimensions, under the same fitting conditions described in Table 4.

Scenario	Seed data	Statistic	Basal probability ϑ_0				
			0.01	0.05	0.10	0.15	0.20
Interaction probabilities							
1	wiring	Mean	0.1482	0.1482	0.1479	0.1485	0.1532
		SD	0.0135	0.0144	0.0144	0.0140	0.0162
2	tech	Mean	0.1065	0.1063	0.1057	0.1055	0.1062
		SD	0.0028	0.0027	0.0028	0.0027	0.0028
3	seven	Mean	0.1144	0.1141	0.1132	0.1134	0.1139
		SD	0.0072	0.0067	0.0069	0.0066	0.0067
4	girls	Mean	0.0715	0.0718	0.0717	0.0716	0.0727
		SD	0.0040	0.0042	0.0042	0.0039	0.0045
5	aarhus	Mean	0.0855	0.0855	0.0854	0.0853	0.0855
		SD	0.0027	0.0024	0.0027	0.0024	0.0028
6	micro	Mean	0.0614	0.0616	0.0614	0.0613	0.0617
		SD	0.0033	0.0033	0.0034	0.0030	0.0029
Latent positions							
1	wiring	Mean	0.1023	0.1029	0.1083	0.1169	0.1919
		SD	0.0238	0.0226	0.0321	0.0473	0.0771
2	tech	Mean	0.2263	0.2254	0.2150	0.2117	0.2313
		SD	0.0314	0.0311	0.0308	0.0296	0.0339
3	seven	Mean	0.6599	0.6502	0.6021	0.6093	0.6314
		SD	0.0643	0.0641	0.0677	0.0883	0.0688
4	girls	Mean	0.6447	0.7803	0.7722	0.7849	0.9058
		SD	0.2458	0.3177	0.3576	0.3020	0.3405
5	aarhus	Mean	0.7482	0.7116	0.7554	0.7968	0.7535
		SD	0.2309	0.2046	0.2487	0.2617	0.2186
6	micro	Mean	0.9695	0.9814	0.9772	0.9726	0.9825
		SD	0.0958	0.0996	0.0817	0.0848	0.0893

6.2. Sensitivity analysis

Our protocol for carrying out a sensible choice of hyperparameters heavily relies on the basal probability ϑ_0, the prior probability of observing an edge between any two actors (recall our discussion in Section 2.2). Here, we consider the effect of varying ϑ_0 on the posterior estimates of interaction probabilities $\{\vartheta_{i,i',j}\}$, latent positions $\{u_{i,j,k}\}$, and network correlations $\{\rho_{j,j'}\}$.

In what follows, we consider again the simulation strategy implemented in our previous experiment, but fixing the latent dimension K to its true value (see column four in Table 4), with the aim of examining a substantial range of ϑ_0 values from low to high (recall we use a parsimonious value of $\vartheta_0 = 0.10$ in all the data analyses provided in Section 4). Therefore, under each scenario, we generate 100 synthetic datasets, and then, MNLPM is fitted using $\vartheta_0 \in \{0.01, 0.05, 0.10, 0.15, 0.20\}$, which allow us to assess the robustness of our model to the choice of ϑ_0. Finally, we complete the assessment by computing the AMSE for the quantities of interest.

Note that our strategy provides us with a sense of how well the model behaves in terms of recovering the quantities of interest because we have knowledge of the "true" model parameters (otherwise we would not have been able to compute the AMSE). Indeed, each of the 1,200 synthetic datasets considered here were generated from setting the model parameters to the corresponding posterior means obtained in our analysis of the seed datasets. Therefore, we do have an indication about how reliable our model parameter estimates are in comparison with the "truth", across a range of different prior specifications.

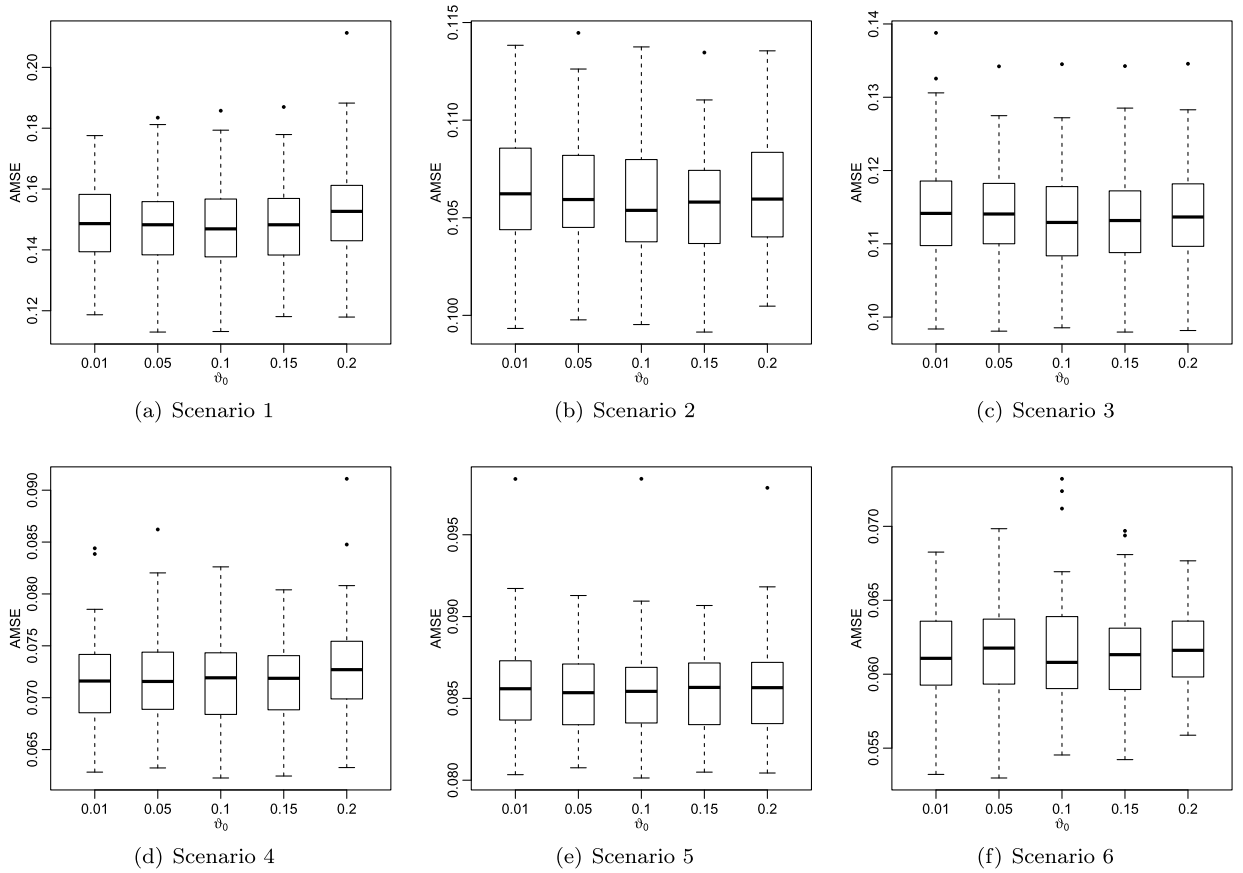
Results are shown in Tables 5 and 6. In general, for each set of parameters, we see that estimates tend to be quite stable, since both the mean AMSE and SD AMSE remain roughly constant across ϑ_0 values. We make this behavior evident graphically in Fig. 15, by displaying the AMSE distribution of the interaction probabilities for every value of ϑ_0 under each scenario. These findings strongly suggest that our approach is quite robust to the choice of ϑ_0, and are also quite comforting, since even non-identifiable quantities present relatively low AMSE values.

7. Discussion

This paper presents a novel approach to modeling multilayer network data with a method that encourages the flow of information across networks, as opposed to an independent characterization of each of them. Our proposal is based on a natural hierarchical extension of a latent space distance model, which provides a direct description of actors' roles within and across networks at global and specific levels. Furthermore, our experiments provide sufficient empirical evidence to establish that our approach is highly competitive in terms of prediction and goodness-of-fit.

Table 6
AMSE empirical distribution summaries for network correlations, across a range of prior interaction probability values, for 100 synthetic datasets independently generated from MNLPM under fixed latent dimensions, under the same fitting conditions described in Table 4.

Scenario	Seed data	Statistic	Basal probability ϑ_0				
			0.01	0.05	0.10	0.15	0.20
Network correlations							
1	wiring	Mean	0.1043	0.1055	0.1037	0.1114	0.1305
		SD	0.0497	0.0493	0.0526	0.0549	0.0647
2	tech	Mean	0.0850	0.0846	0.0888	0.0937	0.0888
		SD	0.0338	0.0325	0.0373	0.0401	0.0352
3	seven	Mean	0.0724	0.0659	0.0495	0.0539	0.0612
		SD	0.0380	0.0307	0.0294	0.0327	0.0333
4	girls	Mean	0.0325	0.0259	0.0305	0.0298	0.0590
		SD	0.0716	0.0140	0.0348	0.0391	0.1324
5	aarhus	Mean	0.0575	0.0545	0.0599	0.0552	0.0591
		SD	0.0507	0.0196	0.0542	0.0211	0.0478
6	micro	Mean	0.0208	0.0223	0.0191	0.0181	0.0179
		SD	0.0157	0.0215	0.0159	0.0123	0.0128



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

(d) Scenario 4 (e) Scenario 5 (f) Scenario 6

Fig. 15. AMSE empirical distributions for interaction probabilities when fitting MNLPM as described in Table 5, across a range of prior interaction probability values, under six simulation scenarios.

From the literature review presented in Section 1, we highlight that a non-negligible number of hierarchical approaches in the context of multilayer network data have been successfully postulated in the past. In particular, our approach may resemble some hierarchical ideas proposed first in the prolific work by D'Angelo and collaborators (e.g., D'Angelo et al. (2020a), D'Angelo et al. (2019), D'Angelo et al. (2020b)). However, there are certainly substantial differences. For instance, in order to make fewer approximation steps in the estimation procedure, Gollini and Murphy (2016) consider the squared Euclidean measure in latent space instead of the Euclidean distance used in Hoff et al. (2002) and in our approach. On the other hand, we simply consider conditionally independent univariate Bernoulli random draws, as opposed to a multivariate

Bernoulli likelihood as in Salter-Townshend and McCormick (2017). Finally, D'Angelo et al. (2019) consider subject-specific latent positions, while we regard them as subject-layer-specific and embedded into a two-stage prior distribution; also, D'Angelo et al. (2019) employ sender and receiver effects regarding to both nodes and networks (either null, constant or variable) in addition to view-specific intercepts, whereas we merely rely on the latter. Thus, we exhort the reader to be aware of such differences (and surely many others not mentioned here), and pursue the model that accommodates better to the problem at hand.

Our MNLPM is susceptible to many generalizations. First, the model can be extended to represent patterns in the data related to known covariates by letting $\vartheta_{i,i',j} = \Phi(\boldsymbol{x}_{i,i'}^{\mathsf{T}}\boldsymbol{\zeta}_j - e^{\theta_j}\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|)$, where $\boldsymbol{x}_{i,i'} = (x_{i,i',1}, \ldots, x_{i,i',P})$, in addition to a global intercept, is a vector of predictors that incorporates known attributes associated with actors i and i', and $\boldsymbol{\zeta}_j = (\zeta_{j,1}, \ldots, \zeta_{j,P})$ is an unknown vector of fixed effects. Furthermore, in order to represent more general combinations of structural equivalence and homophily in varying degrees, it is also possible to consider other types of latent effects as in a factorial model, by letting $\vartheta_{i,i',j} = \Phi(\zeta_j + \boldsymbol{u}_i\boldsymbol{\Lambda}_j\boldsymbol{u}_{i'})$, where $\boldsymbol{\Lambda}_j = \mathsf{diag}(\lambda_{j,1}, \ldots, \lambda_{j,K})$ is a $K \times K$ diagonal matrix. Lastly, the model can also be modified to handle undirected networks by distinguishing latent "sender" positions, $\boldsymbol{u}_{i,j}$, and latent "receiver" positions, $\boldsymbol{v}_{i,j}$, which leads to $\vartheta_{i,i',j} = \Phi(\zeta_j - e^{\theta_j}\|\boldsymbol{u}_{i,j} - \boldsymbol{v}_{i',j}\|)$.

In the same spirit of Green and Hastie (2009), we can also conceive a trans-dimensional version of the model that treats the latent dimension K as a model parameter (as opposed to a fixed pre-specified quantity), which is quite challenging aside from the computational complexity, since in a varying-dimension case it is not clear how to provide a meaningful interpretation to the latent dimensions. Moreover, following Guhaniyogi and Rodriguez (2020), a truncation of a non-parametric process can be also incorporated into the model, but based on the authors' experience, results are likely to be quite similar.

Also, note that social positions might exhibit clustering patterns, which can be modeled directly by considering cluster assignment parameters $\xi_1, \ldots, \xi_I$ into the hierarchical specification of the model through a Categorical-Dirichlet prior (e.g., Handcock et al. (2007), Krivitsky and Handcock (2008), Krivitsky et al. (2009)). Specifically, we can assume that all the actors in the system are clustered into H groups, each of which occupies a position $\boldsymbol{\varphi}_h$ in the social space, $h = 1 \ldots, H$. In this way, we can think of actor i's average position $\boldsymbol{\eta}_i$ as a Normal deviation from the group position to which it belongs, i.e., $\boldsymbol{\eta}_i \mid \{\boldsymbol{\varphi}_h\}, \kappa^2, \xi_i \overset{\text{ind}}{\sim} \mathsf{N}_K(\boldsymbol{\varphi}_{\xi_i}, \kappa^2\mathbf{I})$, where $\xi_i = h$ means that actor i belongs to cluster h. Nonparametric Bayes approaches in the same spirit of Rodriguez (2015) and D'Angelo et al. (2019) are also possible.

Finally, we recommend consider alternative inference methods in order to account for "big networks", which is currently an active research area in computational statistics (e.g., Gollini and Murphy (2016), Ma et al. (2020), Spencer et al. (2020), Aliverti and Russo (2020)).

Acknowledgements

Funding: This work was supported by the National Science Foundation [grant number 2051911].

Appendix A. MCMC algorithm

The posterior distribution is given by:

$$
\begin{aligned}
p(\boldsymbol{\Upsilon} \mid \mathbf{Y}) = &\prod_{j,i<i'} \mathsf{Ber}\left(y_{i,i',j} \mid \Phi\left(\zeta_j - e^{\theta_j}\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|\right)\right) \times \prod_{i,j} \mathsf{N}_K\left(\boldsymbol{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma^2\mathbf{I}\right) \\
&\times \prod_j \mathsf{N}_1\left(\theta_j \mid \mu_\theta, \tau_\theta^2\right) \times \prod_j \mathsf{N}_1\left(\zeta_j \mid \mu_\zeta, \tau_\zeta^2\right) \times \mathsf{N}_K\left(\boldsymbol{\eta}_i \mid \boldsymbol{\nu}, \kappa^2\mathbf{I}\right) \times \mathsf{IG}\left(\sigma^2 \mid a_\sigma, b_\sigma\right) \\
&\times \mathsf{N}_1\left(\mu_\theta \mid m_\theta, v_\theta^2\right) \times \mathsf{IG}\left(\tau_\theta^2 \mid a_\theta, b_\theta\right) \times \mathsf{N}_1\left(\mu_\zeta \mid m_\zeta, v_\zeta^2\right) \times \mathsf{IG}\left(\tau_\zeta^2 \mid a_\zeta, b_\zeta\right) \\
&\times \mathsf{N}_1\left(\boldsymbol{\nu} \mid \boldsymbol{m}_{\boldsymbol{\nu}}, \mathbf{V}_{\boldsymbol{\nu}}\right) \times \mathsf{IG}\left(\kappa^2 \mid a_\kappa, b_\kappa\right).
\end{aligned}
$$

For a given set of fixed hyperparameters, a_σ, b_σ, a_ζ, b_ζ, a_θ, b_θ, a_κ, b_κ, m_ζ, v_ζ, m_θ, v_θ, $\boldsymbol{m}_{\boldsymbol{\nu}}$, $\mathbf{V}_{\boldsymbol{\nu}}$, the algorithm proceeds by generating a new state $\boldsymbol{\Upsilon}^{(b+1)}$ from a current state $\boldsymbol{\Upsilon}^{(b)}$, $b = 1, \ldots, B$, as follows:

1. Sample $\boldsymbol{u}_{i,j}^{(b+1)}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, according to an adaptive Metropolis-Hastings algorithm with the full conditional distribution:

$$
\begin{aligned}
p(\boldsymbol{u}_{i,j} \mid \text{rest}) \propto &\prod_{i':i<i'} \mathsf{Ber}\left(y_{i,i',j} \mid \Phi(\zeta_j - e^{\theta_j}\|\boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j}\|)\right) \\
&\times \prod_{i':i>i'} \mathsf{Ber}\left(y_{i',i,j} \mid \Phi(\zeta_j - e^{\theta_j}\|\boldsymbol{u}_{i',j} - \boldsymbol{u}_{i,j}\|)\right) \\
&\times \mathsf{N}_K(\boldsymbol{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma^2\mathbf{I}).
\end{aligned}
$$

2. Sample $\boldsymbol{\eta}_i^{(b+1)}$, $i = 1, \ldots, I$, from:

$$\boldsymbol{\eta}_i \mid \text{rest} \sim \mathsf{N}_K \left(\left[\frac{1}{\kappa^2} + \frac{J}{\sigma^2} \right]^{-1} \left[\frac{1}{\kappa^2} \boldsymbol{\nu} + \frac{1}{\sigma^2} \sum_j \boldsymbol{u}_{i,j} \right], \left[\frac{1}{\kappa^2} + \frac{J}{\sigma^2} \right]^{-1} \mathbf{I} \right).$$

3. Sample $(\sigma^2)^{(b+1)}$ from:

$$\sigma^2 \mid \text{rest} \sim \mathsf{IG} \left(a_\sigma + \frac{I\,J\,K}{2}, b_\sigma + \frac{1}{2} \sum_{i,j} (\boldsymbol{u}_{i,j} - \boldsymbol{\eta}_i)^\intercal (\boldsymbol{u}_{i,j} - \boldsymbol{\eta}_i) \right).$$

4. Sample $\boldsymbol{\nu}^{(b+1)}$ from:

$$\boldsymbol{\nu} \mid \text{rest} \sim \mathsf{N}_K \left(\left[\mathbf{V}_{\boldsymbol{\nu}}^{-1} + \frac{I}{\kappa^2} \mathbf{I} \right]^{-1} \left[\mathbf{V}_{\boldsymbol{\nu}}^{-1} \boldsymbol{m}_{\boldsymbol{\nu}} + \frac{1}{\kappa^2} \sum_i \boldsymbol{\eta}_i \right], \left[\mathbf{V}_{\boldsymbol{\nu}}^{-1} + \frac{I}{\kappa^2} \mathbf{I} \right]^{-1} \right).$$

5. Sample $(\kappa^2)^{(b+1)}$ from:

$$\kappa^2 \mid \text{rest} \sim \mathsf{IG} \left(a_\kappa + \frac{I\,K}{2}, b_\kappa + \frac{1}{2} \sum_i (\boldsymbol{\eta}_i - \boldsymbol{\theta})^\intercal (\boldsymbol{\eta}_i - \boldsymbol{\theta}) \right).$$

6. Sample $\theta_j^{(b+1)}$, $j = 1, \ldots, J$, according to an adaptive Metropolis-Hastings algorithm with the full conditional distribution:

$$p(\theta_j \mid \text{rest}) \propto \prod_{i,i':i<i'} \mathsf{Ber} \left(\Phi(\zeta_j - e^{\theta_j} \| \boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \|) \right) \times \mathsf{N}_1(\theta_j \mid \mu_\theta, \tau_\theta^2).$$

7. Sample $\mu_\theta^{(b+1)}$ from:

$$\mu_\theta \mid \text{rest} \sim \mathsf{N}_1 \left(\left[\frac{1}{\nu_\theta^2} + \frac{J}{\tau_\theta^2} \right]^{-1} \left[\frac{m_\theta}{\nu_\theta^2} + \frac{\sum_j \theta_j}{\tau_\theta^2} \right], \left[\frac{1}{\nu_\theta^2} + \frac{J}{\tau_\theta^2} \right]^{-1} \right).$$

8. Sample $(\tau_\theta^2)^{(b+1)}$ from:

$$\tau_\theta^2 \mid \text{rest} \sim \mathsf{IG} \left(a_\theta + \frac{J}{2}, b_\theta + \frac{1}{2} \sum_j (\theta_j - \mu_\theta)^2 \right).$$

9. Sample $\zeta_j^{(b+1)}$, $j = 1, \ldots, J$, according to an adaptive Metropolis-Hastings algorithm with the full conditional distribution:

$$p(\zeta_j \mid \text{rest}) \propto \prod_{i,i':i<i'} \mathsf{Ber} \left(\Phi(\zeta_j - e^{\theta_j} \| \boldsymbol{u}_{i,j} - \boldsymbol{u}_{i',j} \|) \right) \times \mathsf{N}_1(\zeta_j \mid \mu_\zeta, \tau_\zeta^2).$$

10. Sample $\mu_\zeta^{(b+1)}$ from:

$$\mu_\zeta \mid \text{rest} \sim \mathsf{N}_1 \left(\left[\frac{1}{\nu_\zeta^2} + \frac{J}{\tau_\zeta^2} \right]^{-1} \left[\frac{m_\zeta}{\nu_\zeta^2} + \frac{\sum_j \zeta_j}{\tau_\zeta^2} \right], \left[\frac{1}{\nu_\zeta^2} + \frac{J}{\tau_\zeta^2} \right]^{-1} \right).$$

11. Sample $(\tau_\zeta^2)^{(b+1)}$ from:

$$\tau_\zeta^2 \mid \text{rest} \sim \mathsf{IG} \left(a_\zeta + \frac{J}{2}, b_\zeta + \frac{1}{2} \sum_j (\zeta_j - \mu_\zeta)^2 \right).$$

Appendix B. Notation

The cardinality of a set A is denoted by $|A|$. If P is a logical proposition, then $\mathbf{1}\{\mathrm{P}\} = 1$ if P is true, and $\mathbf{1}\{\mathrm{P}\} = 0$ if P is false. $\lfloor x \rfloor$ denotes the floor of x, whereas $[n]$ denotes the set of all integers from 1 to n, i.e., $\{1, \ldots, n\}$. The Gamma function is given by $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, \mathrm{d}u$.

Matrices and vectors with entries consisting of subscripted variables are denoted by a boldfaced version of the letter for that variable. For example, $\boldsymbol{x} = (x_1, \ldots, x_n)$ denotes an $n \times 1$ column vector with entries $x_1, \ldots, x_n$. We use $\mathbf{0}$ and $\mathbf{1}$ to

denote the column vector with all entries equal to 0 and 1, respectively, and $\mathbf{I}$ to denote the identity matrix. A subindex in this context refers to the corresponding dimension; for instance, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The transpose of a vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^{\mathsf{T}}$; analogously for matrices. Moreover, if $\mathbf{X}$ is a square matrix, we use $\text{tr}(\mathbf{X})$ to denote its trace and $\mathbf{X}^{-1}$ to denote its inverse. The norm of $\boldsymbol{x}$, given by $\sqrt{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}}$, is denoted by $\|\boldsymbol{x}\|$.

References

Abdollahpouri, A., Salavati, C., Arkat, J., Tab, F.A., Manbari, Z., 2020. A multi-objective model for identifying valuable nodes in complex networks with minimum cost. Clust. Comput. 23 (4), 2719–2733.

Airoldi, E., Blei, D., Fienberg, S., Xing, E., 2009. Mixed membership stochastic blockmodels. In: Advances in Neural Information Processing Systems, pp. 33–40.

Aldous, D.J., 1985. Exchangeability and related topics. In: École d'Été de Probabilités de Saint-Flour XIII—1983. Springer, pp. 1–198.

Aliverti, E., Russo, M., 2020. Stratified stochastic variational inference for high-dimensional network factor model. arXiv preprint. arXiv:2006.14217.

Banerjee, A., Chandrasekhar, A., Duflo, E., Jackson, M., 2013. The diffusion of microfinance. Science 341 (6144).

Bartz-Beielstein, T., Branke, J., Filipič, B., Smith, J., 2014. In: Parallel Problem Solving from Nature–PPSN XIII: 13th International Conference, Ljubljana, Slovenia, September 13-17, 2014, Proceedings, vol. 8672. Springer.

Betancourt, B., Rodriguez, A., Boyd, N., 2020. Modelling and prediction of financial trading networks: an application to the New York mercantile exchange natural gas futures market. J. R. Stat. Soc., Ser. C, Appl. Stat. 69 (1), 195–218.

Borg, I., Groenen, P., 2005. Modern Multidimensional Scaling: Theory and Applications. Springer Science & Business Media.

Chen, S., Liu, S., Ma, Z., 2020. Global and individualized community detection in inhomogeneous multilayer networks. arXiv preprint. arXiv:2012.00933.

D'Angelo, S., Murphy, T.B., Alfò, M., 2019. Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. Ann. Appl. Stat. 13 (2), 900–930.

D'Angelo, S., Alfò, M., Fop, M., 2020b. Model-based clustering for multivariate networks. arXiv preprint. arXiv:2001.05260.

D'Angelo, S., Alfò, M., Murphy, T.B., 2020a. Modeling node heterogeneity in latent space models for multidimensional networks. Stat. Neerl. 74 (3), 324–341.

Durante, D., Dunson, D., 2014. Nonparametric Bayes dynamic modelling of relational data. Biometrika 101 (4), 883–898.

Durante, D., Dunson, D., 2018. Bayesian inference and testing of group differences in brain networks. Bayesian Anal. 13 (1), 29–58.

Erdös, P., Rényi, A., 1959. On random graphs. Publ. Math. 6 (290–297), 5.

Gamerman, D., Lopes, H., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. CRC Press.

Gao, L., Witten, D., Bien, J., 2019. Testing for association in multi-view network data. arXiv preprint. arXiv:1909.11640.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. CRC Press.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24 (6), 997–1016.

Gollini, I., 2019. lvm4net: latent variable models for networks. r package version 0.3. https://CRAN.R-project.org/package=lvm4net.

Gollini, I., Murphy, T., 2016. Joint modeling of multiple network views. J. Comput. Graph. Stat. 25 (1), 246–265.

Green, P., Hastie, D., 2009. Reversible jump mcmc. Genetics 155 (3), 1391–1403.

Guhaniyogi, R., Rodriguez, A., 2020. Joint modeling of longitudinal relational data and exogenous variables. Bayesian Anal. https://doi.org/10.1214/19-BA1160.

Gupta, S., Sharma, G., Dukkipati, A., 2018. A generative model for dynamic networks with applications. arXiv preprint. arXiv:1802.03725.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7 (2), 223–242.

Hahn, E.D., Soyer, R., 2005. Probit and logit models: differences in the multivariate realm. J. R. Stat. Soc. B, 1–12.

Han, Q., Xu, K., Airoldi, E., 2015. Consistent estimation of dynamic and multi-layer block models. In: International Conference on Machine Learning, pp. 1511–1520.

Handcock, M., Raftery, A., Tantrum, J., 2007. Model-based clustering for social networks. J. R. Stat. Soc., Ser. A, Stat. Soc. 170 (2), 301–354.

Hoff, P., 2009. Multiplicative latent factor models for description and prediction of social networks. Comput. Math. Organ. Theory 15 (4), 261–272.

Hoff, P., 2015. Multilinear tensor regression for longitudinal relational data. Ann. Appl. Stat. 9 (3), 1169.

Hoff, P.D., 2005. Bilinear mixed-effects models for dyadic data. J. Am. Stat. Assoc. 100 (469), 286–295.

Hoff, P.D., 2008. Modeling homophily and stochastic equivalence in symmetric relational data. In: Advances in Neural Information Processing Systems, pp. 657–664.

Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. J. Am. Stat. Assoc. 97 (460), 1090–1098.

Hoover, D.N., 1982. Row-column exchangeability and a generalized model for probability. In: Exchangeability in Probability and Statistics, pp. 281–291.

Kim, B., Lee, K., Xue, L., Niu, X., 2018. A review of dynamic network models with latent variables. Stat. Surv. 12, 105.

Kolaczyk, E.D., Csárdi, G., 2014. Statistical Analysis of Network Data with R, Vol. 65. Springer.

Krackhardt, D., 1987. Cognitive social structures. Soc. Netw. 9 (2), 109–134.

Krivitsky, P., Handcock, M., 2008. Fitting latent cluster models for networks with latentnet. J. Stat. Softw. 24 (5).

Krivitsky, P.N., Handcock, M.S., Raftery, A.E., Hoff, P.D., 2009. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. Soc. Netw. 31 (3), 204–213.

Li, W.-J., Yeung, D.-Y., Zhang, Z., 2011. Generalized latent factor models for social network analysis. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI). Barcelona, Spain.

Linkletter, C., 2007. Spatial process models for social network analysis. Ph.D. thesis. Citeseer.

Liu, Q., 2020. In: The 10th International Conference on Computer Engineering and Networks, Vol. 1274. Springer Nature.

Lofland, C., Rodríguez, A., Moser, S., 2017. Assessing differences in legislators' revealed preferences: a case study on the 107th us senate. Ann. Appl. Stat. 11 (1), 456–479.

Ma, Z., Ma, Z., Yuan, H., 2020. Universal latent space model fitting for large networks with edge covariates. J. Mach. Learn. Res. 21 (4), 1–67.

Magnani, M., Micenkova, B., Rossi, L., 2013. Combinatorial analysis of multiple networks. arXiv preprint. arXiv:1303.4986.

Minhas, S., Hoff, P.D., Ward, M.D., 2019. Inferential approaches for network analysis: amen for latent factor models. Polit. Anal. 27 (2), 208–222.

Nowicki, K., Snijders, T., 2001. Estimation and prediction for stochastic blockstructures. J. Am. Stat. Assoc. 96 (455), 1077–1087.

Paez, M., Amini, A., Lin, L., 2019. Hierarchical stochastic block model for community detection in multiplex networks. arXiv preprint. arXiv:1904.05330.

Paul, S., Chen, Y., 2016. Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. Electron. J. Stat. 10 (2), 3807–3870.

Paul, S., Chen, Y., et al., 2020a. A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. Ann. Appl. Stat. 14 (2), 993–1029.

Paul, S., Chen, Y., et al., 2020b. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. Ann. Stat. 48 (1), 230–250.

Rastelli, R., Maire, F., Friel, N., 2019. Computationally efficient inference for latent position network models. arXiv preprint. arXiv:1804.02274.

Reyes, P., Rodriguez, A., 2016. Stochastic blockmodels for exchangeable collections of networks. arXiv preprint. arXiv:1606.05277.

Rodriguez, A., 2015. A Bayesian nonparametric model for exchangeable multinetwork data based on fragmentation and coagulation processes. Tech. Rep., Technical report. University of California, Santa Cruz.
Roethlisberger, F., Dickson, W., 2003. Management and the Worker, Vol. 5. Psychology Press.
Rubin, D.B., 1976. Inference and missing data. Biometrika 63 (3), 581–592.
Salter-Townshend, M., McCormick, T.H., 2017. Latent space models for multiview network data. Ann. Appl. Stat. 11 (3), 1217.
Schweinberger, M., Snijders, T., 2003. Settings in social networks: a measurement model. Sociol. Method. 33 (1), 307–341.
Sewell, D., 2019. Latent space models for network perception data. Netw. Sci. 7 (2), 160–179.
Sewell, D., Chen, Y., 2015. Latent space models for dynamic networks. J. Am. Stat. Assoc. 110 (512), 1646–1657.
Sewell, D., Chen, Y., 2016. Latent space models for dynamic networks with weighted edges. Soc. Netw. 44, 105–116.
Sewell, D., Chen, Y., 2017. Latent space approaches to community detection in dynamic networks. Bayesian Anal. 12 (2), 351–377.
Sosa, J., Buitrago, L., 2021. A review of latent space models for social networks. Rev. Colomb. Estad. 44 (1), 171–200.
Sosa, J., Rodriguez, A., 2021. A latent space model for cognitive social structures data. Soc. Netw. 65 (1), 85–97.
Spencer, N.A., Junker, B., Sweet, T.M., 2020. Faster mcmc for Gaussian latent position network models. arXiv preprint. arXiv:2006.07687.
Steglich, C., Snijders, T.A., West, P., 2006. Applying siena. Methodology 2 (1), 48–56.
Swartz, T., Gill, P., Muthukumarana, S., 2015. A Bayesian approach for the analysis of triadic data in cognitive social structures. J. R. Stat. Soc., Ser. C, Appl. Stat. 64 (4), 593–610.
Turnbull, K., 2020. Advancements in latent space network modelling. Ph.D. thesis. Lancaster University.
Vickers, M., Chan, S., 1981. Representing Classroom Social Structure. Victoria Institute of Secondary Education, Melbourne.
Wang, L., Zhang, Z., Dunson, D., 2019. Common and individual structure of brain networks. Ann. Appl. Stat. 13 (1), 85–112.
Watanabe, S., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11 (Dec), 3571–3594.
Watanabe, S., 2013. A widely applicable Bayesian information criterion. J. Mach. Learn. Res. 14 (Mar), 867–897.
Wilson, J.D., Cranmer, S., Lu, Z.-L., 2020. A hierarchical latent space network model for population studies of functional connectivity. Comput. Brain Behav., 1–16.
Zhang, X., 2020. Statistical analysis for network data using matrix variate models and latent space models. Ph.D. thesis. The University of Michigan.