

A Latent Shrinkage Position Model for Binary and Count Network Data^{*†}

Xian Yao Gwee[‡], Isobel Claire Gormley[§], and Michael Fop[¶]

Abstract. Interactions between actors are frequently represented using a network. The latent position model is widely used for analysing network data, whereby each actor is positioned in a latent space. Inferring the dimension of this space is challenging. Often, for simplicity, two dimensions are used or model selection criteria are employed to select the dimension, but this requires choosing a criterion and the computational expense of fitting multiple models. Here the latent shrinkage position model (LSPM) is proposed which intrinsically infers the effective dimension of the latent space. The LSPM employs a Bayesian nonparametric multiplicative truncated gamma process prior that ensures shrinkage of the variance of the latent positions across higher dimensions. Dimensions with non-negligible variance are deemed most useful to describe the observed network, inducing automatic inference on the latent space dimension. While the LSPM is applicable to many network types, logistic and Poisson LSPMs are developed here for binary and count networks respectively. Inference proceeds via a Markov chain Monte Carlo algorithm, where novel surrogate proposal distributions reduce the computational burden. The LSPM’s properties are assessed through simulation studies, and its utility is illustrated through application to real network datasets. Open source software assists wider implementation of the LSPM.

Keywords: Bayesian nonparametrics, latent position model, multiplicative gamma process prior, adaptive sampler, network data.

1 Introduction

Network data are a collection of interconnected objects which are usually represented formally using graph theory. The objects are often called nodes while their connections to each other are called edges. Living in an increasingly connected world, network data have garnered increased interest in recent years. Social network data are a well known example where friendships (Liu and Chen, 2021), competitions (D’Angelo et al., 2019), companies (Ryan et al., 2017), households (Fosdick et al., 2019), and other social relations are exhibited between individuals (or actors) and can be modelled to understand how people interact in different situations. Beyond that, the network-based perspective

^{*}This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6049.

[†]Isobel Claire Gormley and Michael Fop have contributed equally to this work.

[‡]School of Mathematics and Statistics, University College Dublin (UCD), Ireland, xian-yao.gwee@ucdconnect.ie

[§]School of Mathematics and Statistics, University College Dublin (UCD), Ireland, claire.gormley@ucd.ie

[¶]School of Mathematics and Statistics, University College Dublin (UCD), Ireland, michael.fop@ucd.ie

has found useful applications in complex systems across various fields including computational biology such as protein-protein interactions (Chen et al., 2019); neuroscience on the topological properties of brain connectomes (Yang et al., 2020); economy on understanding trades between countries (Sajedianfard et al., 2021); education on online problem based learning (Saqr and Alamro, 2019); and public health on studying the spread of infectious disease (Jo et al., 2021).

There are many types of models for network data. Most are variants of the random graph model (Erdős and Rényi, 1959; Gilbert, 1959), the stochastic blockmodel (Holland et al., 1983), or the latent position model (LPM, Hoff et al., 2002). The LPM has received much attention as it provides a meaningful visualisation of the data and rich qualitative information (Ma et al., 2020; Tafakori et al., 2021). The LPM postulates that the nodes have positions in a latent space, and that the observed edge formation process is explained through the distance between the nodes' latent positions. In this way, important features such as transitivity, reciprocity, and homophily are easily accounted for in the model, which captures local and global structures (Rastelli et al., 2016; Kim et al., 2018; Smith et al., 2019).

The number of dimensions of the latent space, p , in the LPM is usually unknown and needs to be inferred from the data. Typically, p is fixed at two for easy visualisation and interpretation (Sewell and Chen, 2016; Zhang et al., 2020; Liu and Chen, 2021). However, this is an arbitrary choice and may result in an incomplete or overly complex description of the network. The estimation of p has received some attention in the literature. Attempts to tackle the dimension selection issue have used model selection tools such as a variant of the well-known Bayesian information criterion (BIC, Handcock et al., 2007), the Akaike information criterion (Gormley and Murphy, 2010; Sewell, 2021), the deviance information criterion (Friel et al., 2016), and the Watanabe-Akaike information criterion (Ng et al., 2021; Sosa and Betancourt, 2022). Models with different numbers of dimensions are fitted and then compared using the chosen model selection criterion. However, the selected criterion may not be formally correct for choosing p (Handcock et al., 2007). Furthermore, fitting a range of models, each with a different value of p , can incur a large computational cost and restrict the set of models considered.

An alternative approach to the dimension selection problem is to have an automatic process that infers the optimal dimension of the latent space from the data. This type of process has been used in areas such as factor analysis using shrinkage priors (Bhattacharya and Dunson, 2011; Durante, 2017; Murphy et al., 2020), or spike and slab distributions (Legramanti et al., 2020). In network analysis, within a stochastic block-model setting, automatic dimension estimation has been used in Yang et al. (2020) in a frequentist framework while Passino and Heard (2020) used a Bayesian framework.

In the context of the LPM, a number of proposals for automatic inference of the dimension of the latent space using shrinkage priors have been suggested. Rastelli (2018) develops an approach to automatically determine the number of dimensions by creating a finite mixture of unidimensional LPMs and utilising a Dirichlet shrinkage prior on the mixing proportions. The key idea is to intentionally overfit the number of dimensions by considering a relatively large number of mixture components and then empty the superfluous components using the shrinkage prior during inference. The approach shows

good performance in recovering the true dimension with small networks but tends to overestimate the number of dimensions when the number of nodes is large. Also, the approach is less interpretable than the original LPM as the Euclidean distance between nodes across dimensions has no relevant meaning. Durante and Dunson (2014) employ the shrinkage prior of Bhattacharya and Dunson (2011) when considering the dimension of the latent space in the projection model formulation of the LPM for a dynamic binary network; the distance model formulation of the LPM is considered here. Durante et al. (2017) use a similar shrinkage approach but in the context of a population of networks.

Here the *latent shrinkage position model* (LSPM), which facilitates automatic inference on the dimensionality of the latent space, is introduced. The LSPM is a Bayesian nonparametric model that theoretically allows infinitely many dimensions. This is established through a shrinkage process prior inspired by Bhattacharya and Dunson (2011). The key idea is that the variance of the latent positions on each dimension becomes increasingly small as the number of dimensions grows. The informative, effective dimensions are those that have non-negligible variance in the latent positions while the uninformative, negligible dimensions will have latent position variances that tend to zero. With the LSPM, the need to select a model selection criterion is obviated, only a single model needs to be fitted and the LPM’s ease of interpretation is retained. In addition, posterior uncertainty concerning the latent space dimension is automatically accounted for. The LSPM builds on Durante and Dunson (2014) and Durante et al. (2017) by (a) considering binary and count valued networks with the development of the logistic LSPM and the Poisson LSPM respectively, (b) introducing a truncated version of the Bhattacharya and Dunson (2011) multiplicative gamma process prior and (c) proposing a computationally efficient Markov chain Monte Carlo inference scheme through the use of surrogate proposal distributions.

The remainder of this article is structured as follows: Section 2 describes the LPM and the proposed LSPM. Section 3 outlines the inferential process. Section 4 contains simulation studies exploring the performance of the LSPM in terms of inferring the number of effective dimensions and estimating parameters across realistic settings. Section 5 applies the proposed LSPM to a range of networks, having different edge types and characteristics. Section 6 concludes the article and discusses some potential extensions. R (R Core Team, 2022) code with which all results presented herein were produced is freely available from the [lspm](#) GitLab repository.

2 The latent shrinkage position model

2.1 The latent position model

Network data typically take the form of an $n \times n$ adjacency matrix, \mathbf{Y} , where n is the number of observed nodes, with entries $y_{i,j}$ denoting the relationship between node i and node j . To probabilistically model the presence or absence of an edge between two nodes, the latent position model (LPM, Hoff et al., 2002) assumes that the observed edge formation process can be explained in terms of the nodes’ latent positions in a p -dimensional latent space. Under the LPM, edges are assumed independent, conditional

on the latent positions of the nodes. Self-loops are not permitted and thus the diagonal elements of \mathbf{Y} are zero. The sampling distribution is

$$\mathbb{P}(\mathbf{Y} \mid \alpha, \mathbf{Z}) = \prod_{i \neq j} \mathbb{P}(y_{i,j} \mid \alpha, \mathbf{z}_i, \mathbf{z}_j),$$

where \mathbf{Z} is the $n \times p$ matrix of latent positions with \mathbf{z}_i denoting the latent position of node i , while α is a global parameter that captures the overall connectivity level in the network. A generalised linear model (McCullagh and Nelder, 1998) can be used to model the probability of a range of edge types between nodes as a function of their latent positions. Hoff et al. (2002) propose the distance model formulation of the LPM by considering the Euclidean distance between \mathbf{z}_i and \mathbf{z}_j , the latent locations of nodes i and j respectively, i.e.,

$$g[\mathbb{E}(y_{i,j})] = \alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2, \quad (1)$$

where g is an appropriate link function. As in Gollini and Murphy (2016) and D'Angelo et al. (2019), here the distance is taken to be the squared Euclidean distance. This model is particularly suitable for undirected networks or directed networks that exhibit strong reciprocity. The projection model formulation is an alternative LPM that considers the angle between nodes; Salter-Townshend et al. (2012) provide a comprehensive review.

The LPM has been predominantly used to model binary valued networks via the use of a logit link function

$$\log \frac{q_{i,j}}{1 - q_{i,j}} = \alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2,$$

where $q_{i,j}$ is the probability of forming an edge between node i and node j . The probability of forming an edge between nodes when the distance between them is zero is determined by α . There have been various extensions to this model, for example, by including covariates (Hoff et al., 2002) and random effects (Hoff, 2003), sender and receiver effects (Hoff, 2005), social reach (Sewell and Chen, 2015), and stochastic block-models (Ng et al., 2021). While feasible, such extensions are not considered here in the interest of simplicity.

In addition to binary valued networks, various link functions within the generalised linear model framework can be utilised to model alternative edge types. This includes, but is not limited to, the use of a Poisson distribution (Hoff, 2005) for modelling count edges, and a truncated normal distribution (Sewell and Chen, 2016) and an exponential distribution (Rastelli, 2018) for non-negative continuous edges. In addition to binary edges, the LPM for count edges is considered here where

$$\log(\lambda_{i,j}) = \alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2,$$

where $\lambda_{i,j}$ is the rate parameter of a Poisson distribution between node i and j . A similar interpretation to the binary case applies here where α is the rate for forming an edge between a pair of nodes when the distance between them is zero.

Beyond different edge types, the LPM has been extended, for example, to facilitate clustering of nodes in a network (Handcock et al., 2007; Gormley and Murphy, 2010),

to model dynamic networks (Sewell and Chen, 2015), and to model multiple networks (Gollini and Murphy, 2016; Salter-Townshend and McCormick, 2017; D’Angelo et al., 2019). While extending the LPM has received much attention, attempts to objectively infer the latent space dimension have been few; subjectively choosing $p = 2$ or choosing between a set of possible dimensions via a selected model selection criterion are typical approaches. Here, the LSPM is proposed to address the inference of p in a Bayesian nonparametric framework using a shrinkage prior.

2.2 The latent shrinkage position model

The latent shrinkage position model (LSPM) is a LPM which employs a multiplicative truncated gamma process (MTGP) prior allowing the model to intrinsically infer the number of effective dimensions from the network data, where “effective dimensions” means the dimensions necessary to fully describe the network. The MTGP prior builds on the multiplicative gamma process (MGP) prior that originates from the infinite factor model literature (Bhattacharya and Dunson, 2011), where the MGP prior allows an infinite number of factors whose loadings are shrunk towards zero as the factor number increases. Legramanti et al. (2020) modified the MGP by employing a spike and slab distribution structure that increases the prior mass on the spike for later factors. Other popular shrinkage priors include the Dirichlet-Laplace (Bhattacharya et al., 2015), generalized double Pareto (Armagan et al., 2013), and the horseshoe (Carvalho et al., 2010) priors.

Under the LPM, the prior on the latent positions is typically assumed to be a zero centred Gaussian with equal variance across each of the p dimensions. Under the LSPM, the latent positions are assumed to have a zero centred Gaussian distribution with diagonal precision matrix $\mathbf{\Omega}$, whose entries ω_ℓ denote the precision of the latent positions in dimension ℓ , for $\ell = 1, \dots, \infty$. The LSPM employs a MTGP prior on the precision parameters: the latent dimension h has an associated shrinkage strength parameter δ_h , where the cumulative product of δ_1 to δ_ℓ gives the precision ω_ℓ . An unconstrained gamma prior is assumed for the shrinkage strength parameter for the first dimension, while a truncated gamma distribution is assumed for the remaining dimensions to ensure shrinkage. Specifically, for $i = 1, \dots, n$

$$\mathbf{z}_i \sim \text{MVN}(\mathbf{0}, \mathbf{\Omega}^{-1}), \quad \omega_\ell = \prod_{h=1}^{\ell} \delta_h \quad \text{for } \ell = 1, \dots, \infty, \quad (2)$$

$$\delta_1 \sim \text{Gam}(a_1, b_1 = 1), \quad \delta_h \sim \text{Gam}^T(a_2, b_2 = 1, c_2 = 1) \quad \text{for } h > 1.$$

Here a_1 and b_1 are the shape and rate parameters of the gamma prior on the shrinkage parameter of the first dimension, while a_2 is the shape parameter, b_2 is the rate parameter, and c_2 is the left truncation point (here set to 1) of the truncated gamma prior for dimensions $h > 1$. This MTGP prior results in an increasing precision and therefore a shrinking variance in the higher dimensions of the latent space.

While Bhattacharya and Dunson (2011) state that ω_ℓ stochastically increases under the restriction $a_2 > 1$, Durante (2017) shows that this does not guarantee shrinkage,

and indeed that shrinkage does not occur when $a_1 = 1$ and $a_2 = 1.1$. While the reader is referred to Durante (2017) for guidelines to study the behavior of the MGP under all possible combinations of a_1 and a_2 , Durante (2017) theoretically and empirically suggests that using $a_1 = 2$ and $a_2 = 3$ induces posterior distributions with the desired characteristics, with improved performance when the true number of dimensions is low. Although the ω_ℓ will be stochastically increasing under such hyperparameter settings, this sometimes poses difficulties in the MCMC algorithm used for inference. In particular, $\delta_h < 1$ can be accepted in some cases, which increases the latent positions' variance on dimension h rather than decreasing it. To tackle this, here a truncated gamma distribution bounded between 1 and ∞ is assumed for the second and higher dimensions to ensure shrinkage across these dimensions. Nevertheless, the shrinkage parameter on the first dimension does not require the use of a truncated gamma distribution allowing it to have an unconstrained range and allowing the first dimension to encode as much information as required. The variance of the first dimension therefore determines the maximum possible variance for higher dimensions.

Empirically (see Supplementary Material (Gwee et al., 2023), Appendix C), while the use of the MGP rather than the MTGP prior tends to make little difference in terms of inferring the posterior mode of the number of dimensions, the MGP does tend to result in more diffuse posteriors, with the MTGP being more precise in its inference of the dimension and the variance parameters. Moreover, the MTGP prior is faithful to the inherent model principle of shrinking variance at higher dimensions. Furthermore, the MTGP prior induces an intuitively appealing decrease in the order of importance of subsequent dimensions, as it is typical of dimension reduction methods.

Under this MTGP prior, the LSPM is nonparametric with infinitely many dimensions, where unnecessary higher dimensions' variances are increasingly shrunk towards zero. Dimensions that have variance very close to zero will then have little to no meaningful information encoded in them as the distances between nodes will be close to zero. Thus, the effective dimensions are those in which the variance is non-negligible.

2.3 Properties of the latent shrinkage position model

Given the use of the multiplicative truncated gamma process prior, it is of interest to explore the behaviour of distances in the latent space imposed by the LSPM. The full derivations for this section are given in the Supplementary Material in Appendix A.

The expected squared distance between nodes i and j within the ℓ -th dimension given its precision ω_ℓ is

$$\mathbb{E}[(z_{i\ell} - z_{j\ell})^2 \mid \omega_\ell] = 2 \left(\frac{b_1}{a_1 - 1} \right) \left[\frac{\Gamma(a_2 - 1, 1)}{\Gamma(a_2, 1)} \right]^{\ell-1}, \quad (3)$$

where Γ is the upper incomplete gamma function. Since $\Gamma(a_2 - 1, 1) < \Gamma(a_2, 1)$, any dimension higher than $\ell = 1$ is expected to have a smaller distance. Increasing a_1 will result in the biggest decrease in all of the expected distances for $\ell = 1, \dots, \infty$ as later dimensions are based on previous dimension(s). Figure 1, in which a_1 is set to 2,

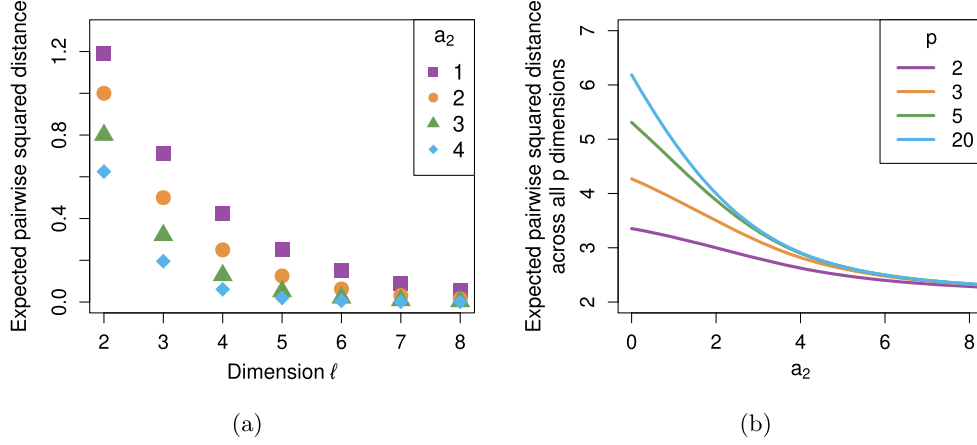


Figure 1: (a) The expected pairwise squared distance for each dimension ℓ for different values of a_2 ; (b) The expected pairwise squared distance in the latent space for different values of a_2 and different truncation levels p .

illustrates the typical behaviour of the expected squared distance between a pair of nodes. Increasing a_2 will decrease the gamma function ratio in (3) and thus decrease the expected squared distance as shown in Figure 1(a). For fixed a_1 and b_1 , the minimum decrease in expected squared distance is $\lim_{a_2 \rightarrow 0} \frac{\Gamma(a_2-1,1)}{\Gamma(a_2,1)} = 0.68$ which means the expected distance contribution from each higher dimension will be at least 32% less than the previous dimension.

The expected squared distance between nodes i and j in the latent space is the geometric sum of (3) across all dimensions $\ell = 1, \dots, p < \infty$, i.e.

$$\mathbb{E}[(\mathbf{z}_i - \mathbf{z}_j)^2 \mid \omega_1, \dots, \omega_p] = 2 \left(\frac{b_1}{a_1 - 1} \right) \left[\frac{1 - \left(\frac{\Gamma(a_2-1,1)}{\Gamma(a_2,1)} \right)^p}{1 - \frac{b_2 \Gamma(a_2-1,1)}{\Gamma(a_2,1)}} \right].$$

Figure 1(b) shows the behaviour of $\mathbb{E}[(\mathbf{z}_i - \mathbf{z}_j)^2 \mid \omega_1, \dots, \omega_p]$ under different truncation levels while varying the hyperparameter a_2 . As a_2 increases, the expected squared distance contribution from the higher dimensions gets increasingly small and encourages a smaller number of effective dimensions, which is notable at $a_2 > 6$, showing very small distance contributions from higher latent dimensions. The positive limit as $a_2 \rightarrow \infty$, alluded to in Figure 1(b), is due to the contribution of the expected squared distance from the first dimension.

3 Inference

The joint posterior distribution of the LSPM is

$$\mathbb{P}(\alpha, \mathbf{Z}, \boldsymbol{\delta} \mid \mathbf{Y}) \propto \mathbb{P}(\mathbf{Y} \mid \alpha, \mathbf{Z}) \mathbb{P}(\alpha) \mathbb{P}(\mathbf{Z} \mid \boldsymbol{\delta}) \mathbb{P}(\boldsymbol{\delta})$$

where $\mathbb{P}(\mathbf{Z} \mid \boldsymbol{\delta})$ and $\mathbb{P}(\boldsymbol{\delta})$ are the prior distributions outlined in Section 2.2; for α , a $N(\mu_\alpha = 0, \sigma_\alpha^2 = 9)$ non-informative prior is assumed throughout, where μ_α is the mean and σ_α^2 is the variance. Markov chain Monte Carlo (MCMC) is employed here to draw samples from the joint posterior distribution. Although the MTGP prior is nonparametric, it is not implementable in practice without setting a truncation level, p , on the number of dimensions fitted. Thus, an adaptive Metropolis-within-Gibbs sampler is used to dynamically shrink or augment the truncation level (see Section 3.2). Details of the derivations of the full conditional distributions of the latent positions and parameters are given in the Supplementary Material, in Appendix B.

The MCMC algorithm proceeds by iterating the following steps for $s = 1, \dots, S$ where s denotes the current iteration and S is the total number of iterations.

1. Sample a value $\check{\mathbf{Z}}$ from the $\text{MVN}(\mathbf{Z}^{(s)}, k\boldsymbol{\Omega}^{-1(s)})$ proposal distribution where k is a step size factor. Accept $\check{\mathbf{Z}}$ as $\mathbf{Z}^{(s+1)}$ with probability

$$\frac{\mathbb{P}(\mathbf{Y} \mid \check{\mathbf{Z}}, \alpha^{(s)}) \text{MVN}(\check{\mathbf{Z}}, k\boldsymbol{\Omega}^{-1(s)})}{\mathbb{P}(\mathbf{Y} \mid \mathbf{Z}^{(s)}, \alpha^{(s)}) \text{MVN}(\mathbf{Z}^{(s)}, k\boldsymbol{\Omega}^{-1(s)})}$$
, otherwise set $\mathbf{Z}^{(s+1)} = \mathbf{Z}^{(s)}$.
2. Sample a value $\check{\alpha}$ from an informed Gaussian proposal distribution and accept $\check{\alpha}$ as $\alpha^{(s+1)}$ following the Metropolis-Hastings (M-H) acceptance ratio. Details of this informed proposal distribution are given in Section 3.1.
3. Sample $\delta_1^{(s+1)}$ from

$$\text{Gam}\left(\frac{np}{2} + a_1, b_1 + \frac{1}{2} \sum_{i=1}^n \sum_{\ell=1}^p \prod_{m=2}^\ell \delta_m^{(s)} [z_{i\ell}^{(s+1)}]^2\right),$$
 where $z_{i\ell}$ is the latent position of node i in dimension ℓ .
4. Sample $\delta_h^{(s+1)}$ for $h = 2, \dots, p$ from

$$\text{Gam}^T\left(\frac{n(p-h+1)}{2} + a_2, b_2 + \frac{1}{2} \sum_{i=1}^n \sum_{\ell=h}^p \prod_{m=1, m \neq h}^\ell \delta_m^{(s^*)} [z_{i\ell}^{(s+1)}]^2, 1\right),$$
 where $s^* = s + 1$ for $m < h$ and $s^* = s$ for $m > h$.
5. Calculate $\omega_\ell^{(s+1)}$ by taking the cumulative product of $\delta_1^{(s+1)}$ to $\delta_\ell^{(s+1)}$.

Since the likelihood function considers the Euclidean distances between the latent positions, it is invariant to rotation, reflection, or translation of the latent positions. Thus, to ensure valid posterior inference, as in Gormley and Murphy (2010), here a Procrustean transformation is performed which translates, reflects and rotates the configurations of the latent positions $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(S)}$ to be as similar as possible to a reference configuration $\check{\mathbf{Z}}$. This reference configuration is the configuration with the highest log-likelihood in the burn-in period of the MCMC chain. While this choice is arbitrary, it is somewhat irrelevant as the configuration is used only to identify the model.

As inference from the MCMC algorithm is sensitive to its initial values, they are initialized using the following approach. When $s = 0$:

1. Calculate the geodesic distances between the nodes in the network. The geodesic distance (Kolaczyk and Csárdi, 2020) is the shortest path between two nodes, measured as the number of edges that must be traversed to move from one node to the other.

2. Apply classical multidimensional scaling (MDS) (Cox and Cox, 2001) to the geodesic distances of the network.
3. K-means clustering (Wu, 2012) is applied to the eigenvalues from the MDS and the number of eigenvalues in the smallest cluster is used as the initial truncation level, p_0 .
4. Set $\mathbf{Z}^{(s)}$ to be the set of $n \times p_0$ positions from the resulting MDS coordinates.
5. Fit a standard regression model, with regression coefficients α and β , to the vectorised adjacency matrix, depending on the edge type, e.g.,
 - (a) for binary edges, fit a logistic regression model where $\log \text{odds}(q_{i,j} = 1) = \alpha - \beta \|\mathbf{z}_i^{(s)} - \mathbf{z}_j^{(s)}\|_2^2$ to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$;
 - (b) for count valued edges, fit a Poisson regression model where $\log(\lambda_{i,j}) = \alpha - \beta \|\mathbf{z}_i^{(s)} - \mathbf{z}_j^{(s)}\|_2^2$, to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$.
6. Set $\alpha^{(s)} = \hat{\alpha}$. As the LSPM model (e.g., (1)) implicitly constrains $\beta = 1$, centre and rescale the latent positions by setting $\mathbf{Z}^{(s)} = \sqrt{|\hat{\beta}|} \tilde{\mathbf{Z}}^{(s)}$, where $\tilde{\mathbf{Z}}^{(s)}$ is the mean-centred matrix of the initial latent positions.
7. Obtain $\omega_\ell^{(s)}$ for $\ell = 1, \dots, p_0$ by calculating the precision empirically from $\mathbf{Z}^{(s)}$.
8. Set $\delta_1^{(s)} = \omega_1^{(s)}$ and calculate $\delta_h^{(s)} = \frac{\omega_h^{(s)}}{\omega_{h-1}^{(s)}}$ where $h = 2, \dots, p_0$.

3.1 An informed proposal distribution for α

To improve mixing and the speed of convergence of the Markov chain, an informed proposal distribution is used for α in the Metropolis-Hastings algorithm. This proposal distribution has parameters that are updated as the chain progresses, ensuring that it shadows the target distribution well. This is achieved by approximating a non-linear term in the loglikelihood function using a quadratic Taylor expansion, similar to Gormley and Murphy (2010). Details of the proposal distribution for the Poisson LSPM are provided below; the derivation of the informed proposal distribution for the logistic LSPM is given in the Supplementary Material in Appendix B.2.

For the Poisson LSPM, the loglikelihood is

$$\log L(\mathbf{Y} | \mathbf{Z}, \alpha) = \sum_{i \neq j} [(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) y_{i,j}] - \sum_{i \neq j} [\exp(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2)] - \sum_{i \neq j} \log[y_{i,j}!],$$

the second term of which can be approximated by a quadratic Taylor expansion around $\alpha^{(s)}$ to give

$$- \sum_{i \neq j} [\exp(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2)] \approx - \sum_{i \neq j} [\exp(\alpha^{(s)} - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2)]$$

$$\begin{aligned}
& - (\alpha - \alpha^{(s)}) \sum_{i \neq j} [\exp(\alpha^{(s)} - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2)] \\
& - 0.5(\alpha - \alpha^{(s)})^2 \sum_{i \neq j} [\exp(\alpha^{(s)} - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2)].
\end{aligned}$$

Substituting this expression and completing the square gives an approximated likelihood function that is quadratic in α , which, when combined with the normal prior on α , suggests the use of a normal proposal distribution $N(\bar{\mu}_{\alpha,C}, \bar{\sigma}_{\alpha,C}^2)$ with mean

$$\bar{\mu}_{\alpha,C} = \alpha^{(s)} + \bar{\sigma}_{\alpha,C}^2 \left[\sum_{i \neq j} y_{i,j} - \sum_{i \neq j} [\exp(\alpha^{(s)} - \|\mathbf{z}_i^{(s)} - \mathbf{z}_j^{(s)}\|_2^2)] + \frac{1}{\sigma_\alpha^2} (\mu_\alpha - \alpha^{(s)}) \right],$$

and variance

$$\bar{\sigma}_{\alpha,C}^2 = \left[\sum_{i \neq j} [\exp(\alpha^{(s)} - \|\mathbf{z}_i^{(s)} - \mathbf{z}_j^{(s)}\|_2^2)] + \frac{1}{\sigma_\alpha^2} \right]^{-1}.$$

The parameters of this informed proposal distribution depend on the current state of the chain and thus are automatically updated each iteration. This maintains the approximation of the target distribution across all iterations of the algorithm. A step size factor that multiplies the variance parameter of the proposal distribution is introduced to assist in achieving satisfactory acceptance and further improve mixing. In terms of computational gain when compared to using a random walk proposal distribution, in the case of a network with $n = 50$ and 2 latent dimensions, for example, using the informed proposal distribution reduced the thinning required by up to 25%.

3.2 An adaptive MCMC sampler to infer the latent dimension

Inference through MCMC for the LSPM is implemented using an adaptive sampler. The sampler shrinks or augments the truncation level p , in order to have a finite number of active dimensions in each iteration of the MCMC chain. A similar sampler is used in Bhattacharya and Dunson (2011) and Murphy et al. (2020). At the s -th iteration, adaptation occurs with probability $\mathbb{P}(s) = \exp(-\kappa_0 - \kappa_1 s)$, with $\kappa_0 \geq 0$ and $\kappa_1 > 0$ chosen so that adaptation occurs often at the beginning of the chain but decreases exponentially fast as the chain grows. Here, $3 \leq \kappa_0 \leq 5$ and $\kappa_1 = 3 \times 10^{-5}$ were found to be useful and are used across the studies that follow. Adaptation only occurs after the burn-in period, in order to ensure targeting of the posterior distribution.

At an adaptation step, when $p > 1$, a reduction in the truncation level is based on a criterion which considers the cumulative proportion of variance that the dimensions $\ell = 1, \dots, p-1$ contain. If the dimensions up to the ℓ -th cumulatively contain at least a proportion ϵ_1 of the total variance, then the dimensions from $\ell + 1$ to p add little information, and p is reduced to ℓ . In practice, $\epsilon_1 = 0.9$ has been found to work well. When the adaptation criterion for reducing p is not met, an increase in p is then considered by examining δ_p^{-1} and a threshold ϵ_2 . If $\delta_p^{-1} > \epsilon_2$, a new dimension is added, by sampling δ_{p+1} from the MTGP prior and the latent positions from a univariate zero mean Normal

distribution with induced variance ω_{p+1} . Under the MTGP, $\delta_p^{-1} < 1$ and when $\delta_p^{-1} \approx 1$ shrinkage in the p th dimension is weak and more dimensions may be required to fully describe the data. Therefore we consider $\epsilon_2 = 0.9$ which was found to work well in practice.

When $p = 1$, only an increase in p is possible. In this case, we consider the proportion of latent positions that have absolute deviation from the mean that exceeds the 95% critical value of a standard Normal distribution. When this proportion is ϵ_3 times greater than the expected 0.05 proportion, p is increased to 2. Setting $\epsilon_3 = 5$ demonstrated good performance in simulation studies, with higher values of ϵ_3 increasing the tendency to remain at 1 dimension.

The hyperparameters κ_0 and a_2 can influence mixing in the adaptive MCMC sampler. When p is increased or decreased, it can take time for α to mix well as it compensates for the change in dimension. Small κ_0 increases the adaptation frequency, which can make achieving sufficient mixing of α challenging. Similarly, small a_2 encourages the addition of a dimension with variance similar to that of the p th dimension, which results in large distances between latent positions and α needs time to adjust. Conversely, an a_2 that is too large adds a dimension that has very little variance and encodes little information. The aforementioned $3 \leq \kappa_0 \leq 5$ and $a_2 = 3$ were found to be a good combination in practice.

The adaptive sampler allows inference on the posterior distribution of the number of active dimensions p . The posterior mode p_m is used as the estimate of the effective p , with credible intervals quantifying the associated uncertainty.

3.3 Posterior predictive checking

Posterior predictive checking is a useful way of assessing the fit of a model to the data. Any systematic differences between the networks simulated from the posterior predictive distribution and the observed network may indicate potential failings of the model (Gelman et al., 2013). Sections 4 and 5 include posterior predictive checks to assess the fit of the LSPM to simulated and observed networks respectively. Samples drawn from the posterior predictive distribution are used to simulate replicate networks under the fitted model. These simulated networks are then compared to the observed network by checking similarity metrics, network properties, and distances between networks. The type of metrics used depends on the network type.

Binary valued networks

For binary networks, the similarity metrics considered here are the accuracy and the F_1 -score between the observed network and the replicate networks drawn from the posterior predictive distribution. Accuracy is a measure of the correctly identified edges i.e., $\text{accuracy} = (\text{TP} + \text{TN})/[n(n - 1)]$, where TP is the number of true positives and TN is the number of true negatives. The F_1 -score is the harmonic mean of precision and recall i.e., $F_1\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ where $\text{precision} = \text{TP}/(\text{TP} + \text{FP})$ and $\text{recall} = \text{TP}/(\text{TP} + \text{FN})$, with FP the number of false positives, and FN the number of false negatives.

Network properties such as density and transitivity are also considered. A network’s density is the ratio of the number of observed dyadic connections over the number of possible dyadic connections, while transitivity is three times the number of triangles divided by the total number of connected triples. These metrics assess if the LSPM captures network properties of the observed data.

The Hamming distance is also considered, measuring the normalised difference between the observed network and a replicate network. The Hamming distance is given by $\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} |y_{i,j} - y_{i,j}^{(r)}|$, where $y_{i,j}^{(r)}$ is the link between nodes i and j in the r th replicate network from the posterior predictive distribution.

Count valued networks

As binary network measures are not directly applicable to count valued networks, the (log) frequencies of counts in replicate networks from the posterior predictive distribution are compared to those of the observed network. The mean absolute difference between replicate and observed counts is also considered.

In the simulation studies outlined in Section 4, the fit of the Poisson LSPM is also evaluated in terms of Euclidean distances for estimated and actual latent positions. Specifically, the ratios between the $n(n-1)/2$ distances derived from the posterior mean latent positions and the distances derived from the true latent positions are computed. A distribution of these ratios tightly centered around 1 will indicate good model fit.

4 Simulation studies

The performance of the LSPM is assessed on simulated data scenarios. The simulated data are generated by Bernoulli trials for each node pair based on the probabilities derived from the distances between the nodes’ latent positions. For count valued edges, a Poisson distribution is employed instead of the Bernoulli.

The latent positions are simulated according to (2) with the shrinkage strengths being manually set to explore their effect across different settings. Hyperparameters are set as $\mu_\alpha = 0$, $\sigma_\alpha = 3$, $a_1 = 2$, $a_2 = 3$, $b_1 = b_2 = 1$. A total of 30 networks are simulated in each case. Different step sizes between 0.1 to 3 are used in the proposal distributions to ensure acceptance rates are within the 20–40% range. The MCMC chains are run for 1,000,000 iterations for Section 4.1 and 4.3 with a burn-in period of 100,000 iterations, thinning every 1,500th. In Section 4.2 and Section 4.4, 500,000 iterations are considered, the burn-in period is 50,000 with thinning every 1,500th iteration for the logistic LSPM and every 1,000th iteration for the Poisson LSPM.

The simulation studies are structured as follows: Section 4.1 examines the impact of different initial truncation levels, p_0 . Section 4.2 assesses LSPM capability under different network sizes, n . Section 4.3 studies the performance of the logistic LSPM under different network densities; Section 4.4 explores the performance of the Poisson LSPM under different levels of overdispersion. Where relevant, violin plots are used to visualise posterior distributions for each of the 30 simulated networks. Red crosses or

red lines within the violin plots represent the true values used to simulate the network. Throughout, the LSPM is compared with the LPM and posterior predictive checks are used to assess model fit. Additional results and posterior predictive checks for the simulation studies are available in the Supplementary Material, in Appendix D.

4.1 Study 1: initial truncation level

This section explores the effect of the initial number of latent dimensions p_0 on the inference of the number of active dimensions. Networks with $n = 100$ are generated with the true number of effective latent dimensions $p^* = 4$ and shrinkage strengths of $\delta_1 = 0.5$, $\delta_2 = 1.1$, $\delta_3 = 1.05$, and $\delta_4 = 1.15$ which mean that the second dimension has importance similar to the first. Here, $\alpha = 6$ meaning moderate network density (i.e. $\simeq 20\%$) in the case of binary networks. Four initial truncation levels are considered: p_0 initialised as described in step 3 of Section 3 (termed ‘auto’) and $p_0 = \{2, 4, 10\}$ representing situations where the initial truncation has been underestimated, correctly specified, and overestimated, respectively. Across the simulated networks, $3 \leq p_0 \leq 6$ were found under the ‘auto’ procedure, with $p_0 = \{4, 5\}$ in the majority of cases.

Figure 2 shows the posterior of the number of active dimensions p inferred from the 30 simulated networks. Under the ‘auto’ initialisation procedure, and when $p_0 > p^*$, the posterior concentrates around the true number of dimensions. However, when $p_0 < p^*$, the posterior tends to concentrate on dimensions lower than p^* . Further results are summarised in Table 1, which shows that for all initial values p_0 , the 95% credible intervals include the true dimension p^* . Across binary and count simulated networks,

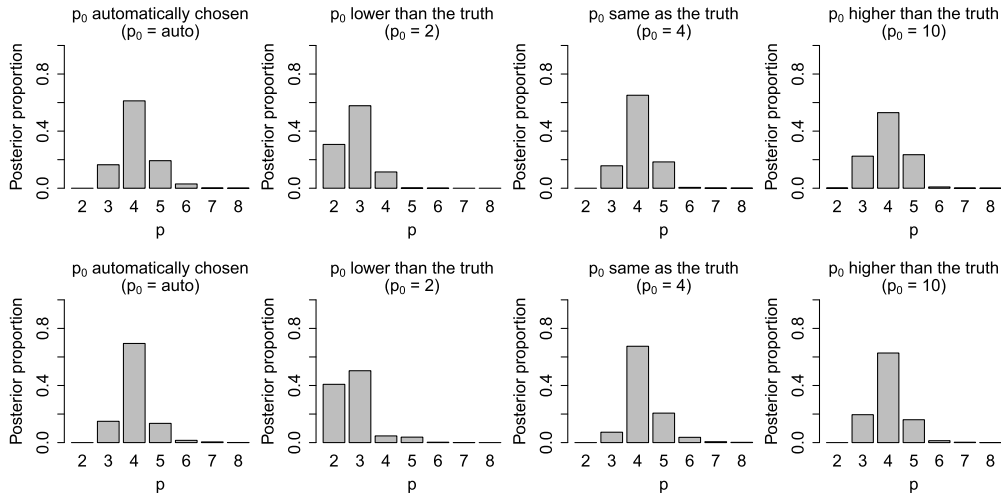


Figure 2: For different initial truncation levels, p_0 , the posterior distribution of the number of active dimensions p across 30 simulated networks under the logistic LSPM for binary networks (top) and under the Poisson LSPM for count networks (bottom).

p_0	Binary networks			Count networks		
	p_m	Proportion where $p_m = p^*$	Procrustes correlation	p_m	Proportion where $p_m = p^*$	Procrustes correlation
auto	4 (3, 6)	0.63	0.96 (0.89, 1.00)	4 (3, 5)	0.73	0.98 (0.84, 1.00)
2	3 (2, 4)	0.13	0.84 (0.69, 0.94)	3 (2, 5)	0.03	0.81 (0.69, 0.92)
4	4 (3, 5)	0.67	0.97 (0.89, 0.99)	4 (3, 6)	0.70	0.98 (0.80, 1.00)
10	4 (3, 5)	0.53	0.97 (0.89, 0.99)	4 (3, 5)	0.67	0.98 (0.89, 1.00)

Table 1: For different initial truncation levels, p_0 , the posterior mode p_m of the number of active dimensions, the proportion of the 30 simulated networks for which $p_m = p^* = 4$, and the Procrustes correlation of the 30 simulated networks’ latent positions with the posterior mean positions. The 95% credible intervals are given in the brackets.

the proportion of times the posterior modal dimension $p_m = p^* = 4$ is greater than 0.63 under the ‘auto’ initialisation, and is greater than 0.53 when $p_0 = 10$, with a poorer proportion of 0.13 or less when initialising $p_0 < p^*$. Table 1 also reports the Procrustes correlations between the true latent positions and the LSPM posterior mean positions, conditioned on p_m ; there is good agreement across network types and initialisation strategies. These results suggest that the ‘auto’ approach or starting with a reasonably large p_0 is advisable for accurate inference on the number of dimensions.

Figure 3 illustrates the posterior distributions of the shrinkage strength parameters, with the number of dimensions truncated at 8 for visualization purposes. When $p_0 = \{4, 10\}$ or is ‘auto’ initialised, the shrinkage parameter δ_5 tends to have a higher and more diffuse posterior than δ_4 , correctly indicating strong shrinkage after the 4th dimension. However, when $p_0 = 2 < p^*$, the shrinkage parameters δ_h for $h > p_0$ tend to be overestimated, causing excessive shrinkage below the true number of dimensions. When $p_0 < p^*$, the adaptive sampler can struggle to increase the dimension, with the shrinkage prior hyperparameter being very influential. These results support the suggestion that the ‘auto’ approach or starting with a reasonably large p_0 is advisable for accurate inference on the number of dimensions.

Posterior distributions for the intercept parameter α , conditional on different numbers of active dimensions $p = \{2, 3, 4, 5\}$, are illustrated in Figure 4. When $p = p^* = 4$, for any p_0 , α is accurately inferred, with more diffuse posteriors in the case of binary networks than for count networks. However, when $p \neq p^*$, the α parameter compensates for the additional or missing dimensions: when $p < p^*$ underestimation of α occurs, due to the missing contribution from the Euclidean distances in the omitted dimensions. Conversely, when $p > p^*$, α is overestimated to account for the additional distance contributed by the extra dimensions, but with a smaller bias than when $p < p^*$.

Posterior predictive checks under the logistic LSPM for binary simulated networks initialised using the ‘auto’ approach and $p_0 = \{2, 4, 10\}$ are shown in Figure 5. A number of LPM models were fitted using the `latentnet` R package (Krivitsky and Handcock, 2020) with $p = \{1, \dots, 8\}$; the BIC suggested $p = 4$ dimensions as optimal. The logistic LSPM fitted well and similarly across $p_0 \geq p^*$ while model fit was poorer across all metrics considered when $p_0 < p^*$. The LPM with $p = 4$ also fitted well, but required multiple models to be fit and the use of a model selection criterion. Figure 6 shows the posterior predictive checks for the Poisson LSPM fitted to simulated count networks. Similar behaviour to the binary case is observed with poor model fit

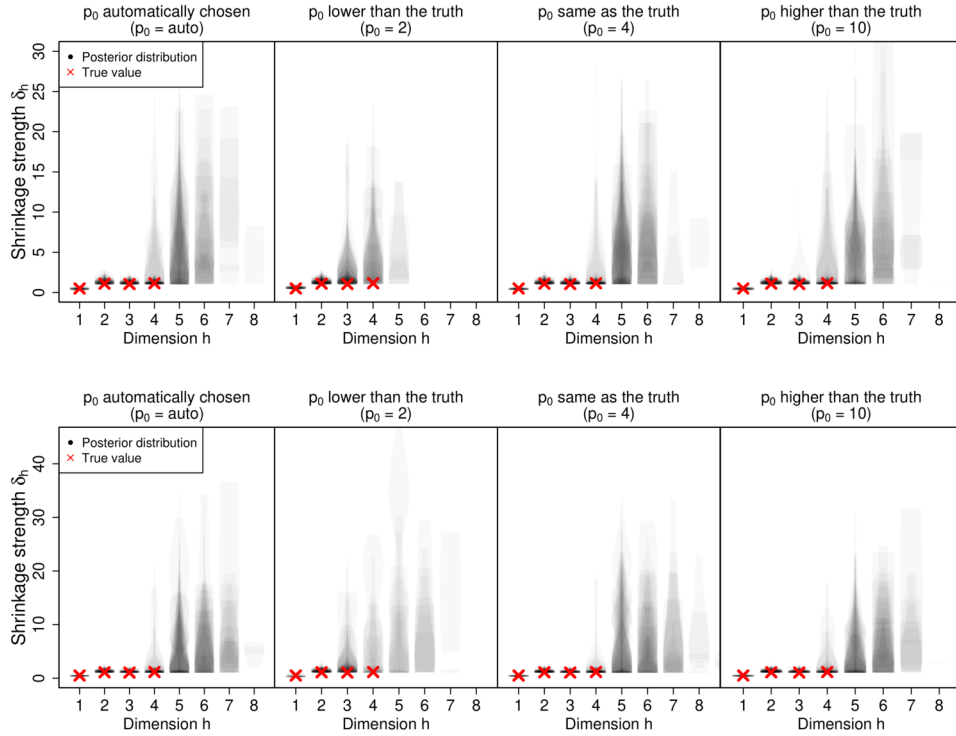


Figure 3: Posterior distributions of shrinkage strength across dimensions for different initial truncation levels p_0 , across 30 simulated networks for the logistic LSPM for binary data (top) and the Poisson LSPM for count data (bottom).

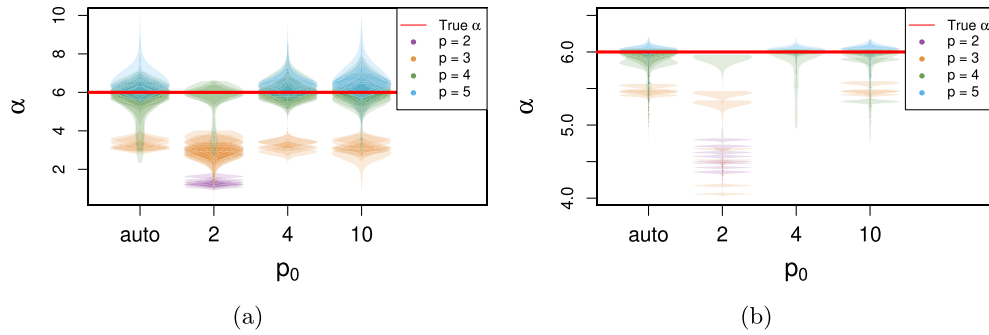


Figure 4: Posterior distributions of α for different initial truncation levels p_0 across 30 simulated networks under (a) the logistic LSPM for binary networks and (b) the Poisson LSPM for count networks.

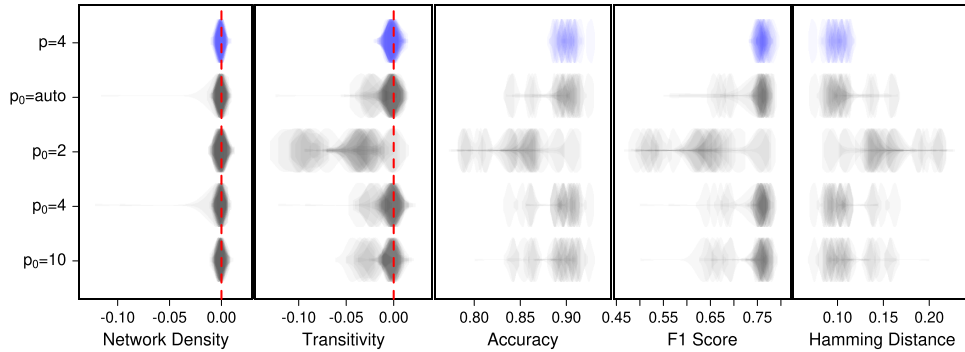


Figure 5: Posterior predictive checks, across 30 simulated binary networks, for the logistic LSPM (gray) under different initial truncation levels p_0 conditioned on p_m and the logistic LPM (blue) with $p = 4$. The violin plots indicate the metric distributions from 30 replicate networks. The network density and transitivity plots illustrate the differences between the posterior predictive and observed networks.

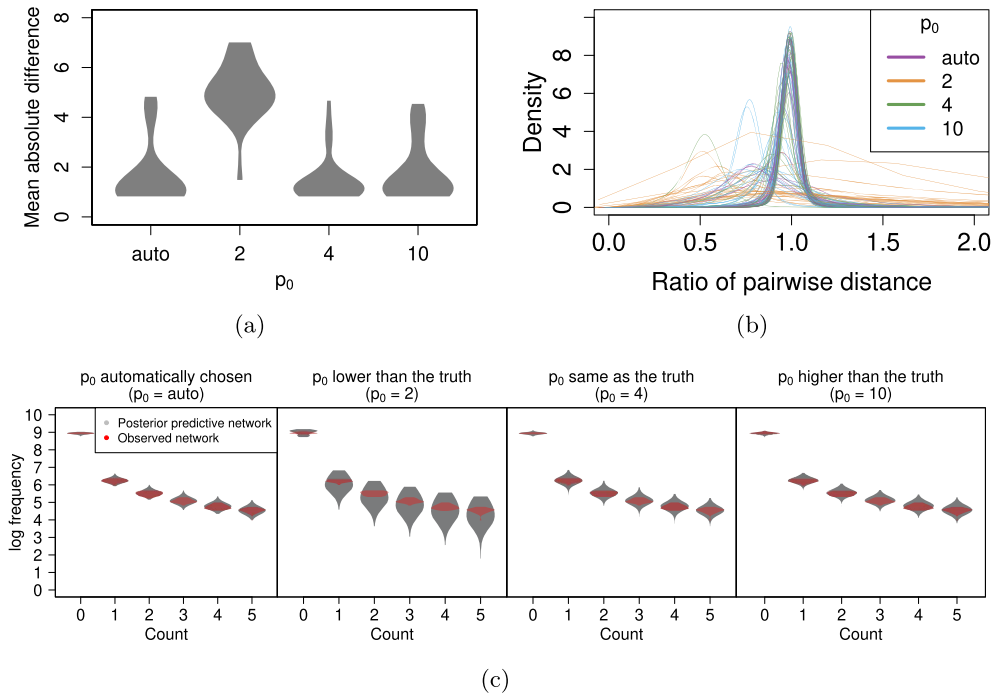


Figure 6: Assessment of model fit for the Poisson LSPM via (a) the mean absolute difference between replicate and observed network counts, (b) ratios of the pairwise distances between inferred and true node locations, and (c) the log frequency of replicate and observed network counts.

when $p_0 < p^*$ and improved fit when $p_0 \geq p^*$: mean absolute count differences are small and ratios of pairwise distances between true and inferred latent locations are centred around 1. There is good agreement between the log frequency of counts in posterior predictive networks and the observed counts, with $p_0 = 2$ having larger uncertainty.

In terms of computational cost, fitting the logistic LSPM on a computer with an i7-10510U CPU and 16 GB RAM took on average 40 minutes for a simulated binary network with $n = 100$. For comparison, fitting a single $p = 4$ LPM with the `latentnet` (Krivitsky and Handcock, 2020) R package, with default settings but the same burn in period and thinning as LSPM, took on average 37 minutes. However, fitting a number of LPM models each with a different p and choosing between them using the BIC is required in the LPM setting, thus incurring additional computational cost.

4.2 Study 2: network size

Here the focus is on assessing the effect of different network sizes. Networks with $n = \{20, 50, 100, 200\}$ are generated with $p^* = 2$, and shrinkage strengths of $\delta_1 = 0.5$ and $\delta_2 = 1.1$ giving $\omega_1^{-1} = 2$ and $\omega_2^{-1} = 1.82$ respectively. Here, $\alpha = 3$ leading to a moderate network density (i.e. $\simeq 20\%$) in the binary network case.

Inference on p_m , the posterior mode of p , and the Procrustes correlations between inferred and observed latent locations as n varies are detailed in Table 2. The posterior mode p_m is accurate across different n , while the 95% credible intervals' upper bounds tend to increase as n increases. This affects the proportion of times the posterior mode is equal to the optimal $p^* = 2$. Networks with larger n are characterized by larger uncertainty on the dimension of the latent space, as the sampler tends to explore higher dimensional solutions. In the majority of cases where p_m is not 2, the posterior mode indicates 3 dimensions. The Procrustes correlation between true and posterior mean latent positions, conditioned on p_m , are increasingly accurate and precise as n increases, with higher values for count than binary networks. As n increases, posterior predictive checks under the LSPM indicate improved model fit (see Supplementary Material, Appendix D).

n	Binary networks			Count networks		
	p_m	Proportion where $p_m = p^*$	Procrustes correlation	p_m	Proportion where $p_m = p^*$	Procrustes correlation
20	2 (2, 3)	1.00	0.964 (0.882, 0.985)	2 (2, 3)	0.93	0.992 (0.807, 0.998)
50	2 (2, 4)	0.80	0.992 (0.981, 0.995)	2 (2, 3)	0.97	0.998 (0.995, 0.999)
100	2 (2, 4)	0.87	0.996 (0.992, 0.997)	2 (2, 4)	0.93	0.999 (0.974, 1.000)
200	2 (2, 5)	0.87	0.998 (0.997, 0.998)	2 (2, 4)	0.77	1.000 (0.999, 1.000)

Table 2: For different n , the posterior mode p_m of the number of active dimensions, the proportion of the 30 simulated networks for which $p_m = p^* = 2$, and the Procrustes correlation of the 30 simulated networks' positions with the LSPM posterior mean positions. The 95% credible intervals are given in brackets.

True α	Empirical network density (%)	p_m	Proportion where $p_m = p^*$	Procrustes correlation
0	2–5	2 (2, 4)	0.23	0.38 (0.22, 0.64)
1	4–8	2 (2, 3)	0.48	0.50 (0.39, 0.75)
5	20–35	3 (2, 4)	0.77	0.98 (0.87, 0.99)
10	49–65	3 (2, 4)	0.90	0.99 (0.93, 0.99)
20	79–94	3 (2, 4)	0.57	0.96 (0.91, 0.98)
30	90–99	2 (2, 4)	0.33	0.70 (0.49, 0.90)

Table 3: For different network densities, the posterior mode p_m of the number active dimensions, the proportion of the 30 simulated networks for which $p_m = p^* = 3$, and the Procrustes correlation of the 30 simulated networks’ positions against the LSPM posterior mean positions. The 95% credible intervals are given in the brackets.

4.3 Study 3: density in binary networks

In the context of binary networks, the impact of network density on the performance of the LSPM is explored. Networks with $n = 50$, $p^* = 3$, $\delta_1 = 0.5$, $\delta_2 = 1.1$, and $\delta_3 = 1.05$ are generated. A range of α values between 0 and 30 are used to simulate networks with densities ranging from 2% to 99%. Table 3 reports the posterior mode of p and Procrustes correlations between inferred and true latent positions under different α values. Under the LSPM, the accuracy of p_m increases as the network density increases from 2% to 65%, but accuracy deteriorates for denser networks with density > 0.79 .

Figure 7(a) illustrates the effect of different network densities on the estimation of the α parameter. Values of $\alpha \leq 10$ are estimated accurately whereas higher values of α tend to be underestimated. Further, as it is evident from Figure 7(b) where the Procrustes correlations decrease for large values of α , the true latent positions are difficult to recover since their influence on the link probabilities is small in comparison to α when α is large. Similarly, the Procrustes correlations decrease for lower values of α as there are many possible locations that will produce a mostly unconnected network. In summary, the logistic LSPM performs best for networks of moderate density. Additional posterior distributions of dimensions, variances and shrinkage strength parameters, and posterior predictive checks, are available in the Supplementary Material, in Appendix D.

4.4 Study 4: overdispersion in count networks

Count network data are often characterised by overdispersion, whereby the variance of the counts is greater than their mean. Here, the impact on the performance of the Poisson LSPM of different levels of overdispersion in count network data is examined by varying the α and δ parameters. The first set of 30 networks are simulated with $\alpha = 0.5$ and $\delta_1 = \delta_2 = 1.5$ giving mean counts between 0.45 and 0.60 and variance between 0.65 and 0.85 (low overdispersion). Another 30 more overdispersed count networks are generated with $\alpha = 1.5$, $\delta_1 = 0.5$, and $\delta_2 = 1.5$ giving mean counts between 0.5 and 0.7 and variance between 1.4 and 2.0 (moderate overdispersion). The final set of 30 highly overdispersed count networks are generated with $\alpha = 5$, $\delta_1 = 0.1$, and

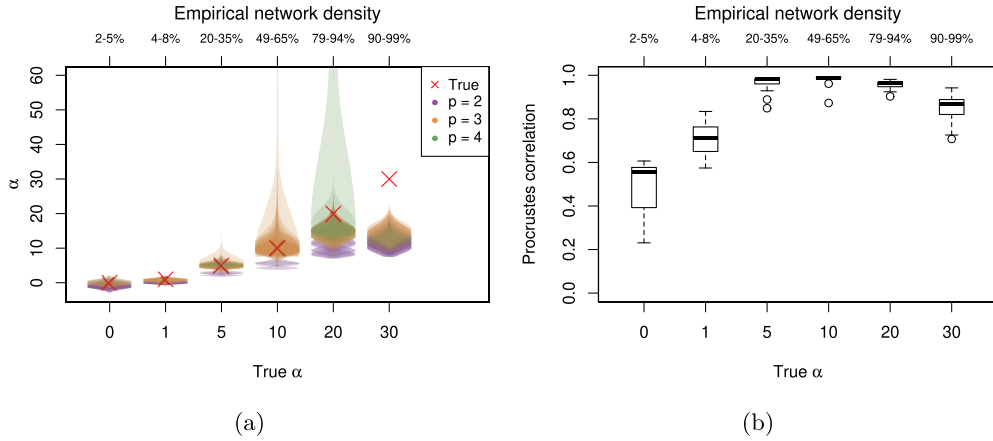


Figure 7: For different network densities the (a) posterior distributions of α and (b) Procrustes correlations between true and LSPM posterior mean latent positions.

Level of overdispersion	p_m	Proportion where $p_m = p^*$	Procrustes correlation
Low	2 (2, 3)	0.93	0.991 (0.950, 0.993)
Moderate	2 (2, 4)	0.80	0.997 (0.885, 0.998)
High	3 (2, 4)	0.43	0.855 (0.732, 0.914)

Table 4: For different levels of overdispersion, the posterior mode p_m of the number of active dimensions, the proportion of 30 simulated networks for which $p_m = p^* = 2$, and the Procrustes correlation of the 30 simulated networks' positions against the LSPM posterior mean positions. The 95% credible intervals are given in the brackets.

$\delta_2 = 1.5$ giving mean counts between 3 and 6 and variance between 220 and 420 (high overdispersion).

Table 4 reports on the posterior mode of p and on the Procrustes correlation between true and inferred positions for different overdispersion levels. As overdispersion increases, the LSPM tends to overestimate the number of effective dimensions. Figure 8 also illustrates that as overdispersion increases, inferential accuracy and precision for the variance parameters decreases. Additional posterior distributions of dimensions, variances and shrinkage strength parameters, and posterior predictive checks, are available in the Supplementary Material, in Appendix D.

5 Illustrative applications

The logistic and Poisson LSPMs are fitted to a range of real network datasets with different characteristics: the well-known Zachary karate club binary network which has a small number of nodes, a cat brain connectivity binary network with moderate number

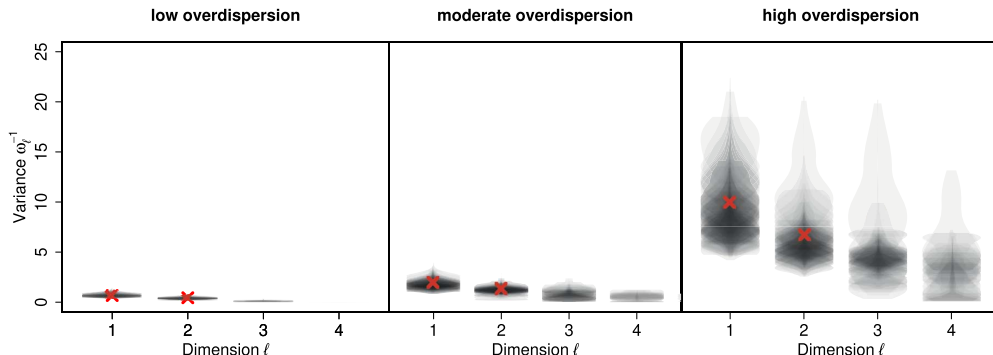


Figure 8: Posterior distributions of variance under the Poisson LSPM fitted to simulated networks with different levels of overdispersion.

of nodes and density, a relatively large but sparse worm nervous system binary network, and an overdispersed phone calls count network. In all cases hyperparameters, initial values, and step sizes are set as in Section 4.

5.1 Zachary karate club binary network data

The well-known Zachary’s karate club network (Zachary, 1977) is a social network with $n = 34$ members and describes the relationships in a university karate club. The karate teacher, Mr. Hi, and the club president, John, are the two central figures and the club was divided into two new clubs after an argument between them. The network is binary, undirected and has a density of 13.9%. For inference, 10 MCMC chains are run, each for 100,000 iterations with a burn-in period of 1,000 and every 400th sample thinned.

Figure 9(a) indicates that the posterior modal dimension is 2, with low associated uncertainty. Upon fitting multiple LPMs with $p = \{1, \dots, 5\}$, the BIC suggests 1 dimension is optimal. Posterior predictive checks for the LPM with $p = 1$ and the logistic LSPM with $p_m = 2$ (Figure 9(b)) indicate better fit for network density, transitivity and F_1 score under the LSPM while the LPM has better accuracy and Hamming distance. The Procrustes correlation between the posterior mean positions from the 1 dimensional LPM and the first dimension of the LSPM has a median value of 0.92 (standard deviation of 0.06) across 10 logistic LSPM and LPM chains. Additional results regarding the posterior distributions of variance and shrinkage strength parameters can be found in the Supplementary Material, Appendix E.

5.2 Cat brain connectivity binary network data

The logistic LSPM is used to analyse a cat brain connectivity network which includes $n = 65$ non-overlapping regions in the cortex treated as nodes and 1139 interregional macroscopic axonal projections treated as edges (Scannell et al., 1995; de Reus and van den Heuvel, 2013). The cortex regions were classified into four categories namely:

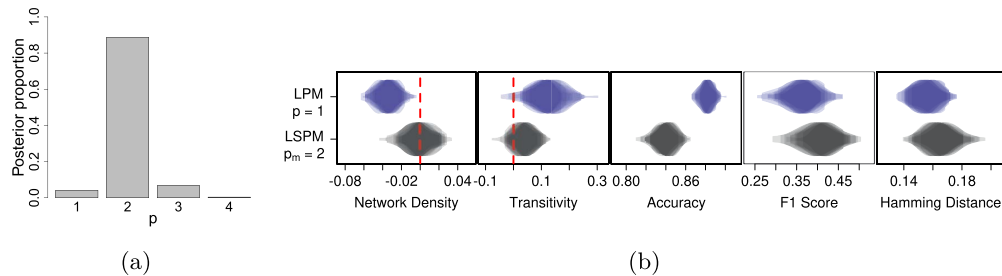


Figure 9: Zachary karate club network: (a) the posterior distribution of p and (b) posterior predictive checks for LPM with $p = 1$ and the logistic LSPM for $p_m = 2$.

visual regions (18 areas), auditory regions (10 areas), somatomotor regions (18 areas), and frontolimbic regions (19 areas). These classifications were based on neurophysiological information about the functional role of each brain area. Here, the network is viewed as a binary directed network, with $y_{i,j} = 1$ indicating there is a connection between regions i and j , while $y_{i,j} = 0$ indicating there is not. The network density is 27.37%.

To assess convergence, the logistic LSPM is fitted 10 times using different initial latent positions by introducing noise to the initial configuration obtained as outlined in Section 3. The MCMC chains are run for 500,000 iterations with a burn-in period of 50,000 and thinning every 2,000th sample. Model fitting took on average 10 minutes on a computer with an i7-10510U CPU and 16 GB RAM. For comparison, fitting an LPM with $p = 5$ with the `latentnet` (Krivitsky and Handcock, 2020) R package with default settings but 500,000 iterations took 7 minutes. However, the LPM setting requires fitting multiple models, thus incurring additional computational cost. The Gelman-Rubin convergence criterion (Gelman et al., 2013) was satisfied with $1.0 \leq \hat{R} < 1.1$ for α and the shrinkage parameters. Visual inspection of trace plots also suggests convergence; an example is provided in the Supplementary Material, Appendix E.

Figure 10(a) suggests that $p_m = 3$ effective dimensions are required for these data, with low associated uncertainty. Upon fitting multiple LPMs with $p = \{1, \dots, 5\}$, the BIC suggests 2 dimensions are optimal, with 3 dimensions being next best. Posterior predictive check metrics from 30 posterior predictive replicate networks for each of the 10 logistic LSPM and LPM MCMC chains are shown in Figure 10(b). While the logistic LSPM tends to underestimate network transitivity, it performs better than the LPM across the majority of metrics, with the estimated density close to the network data density, higher accuracy and F_1 score, and lower Hamming distance.

Figure 11 shows posterior mean latent positions under the logistic LSPM with $p_m = 3$. The inferred latent positions under both the LPM and the LSPM are very similar with a mean Procrustes correlation of 0.97 (standard deviation 0.01) across the 10 chains. Nodes from the same cortical region also lie close to each other in the latent space under both the LPM and the logistic LSPM.

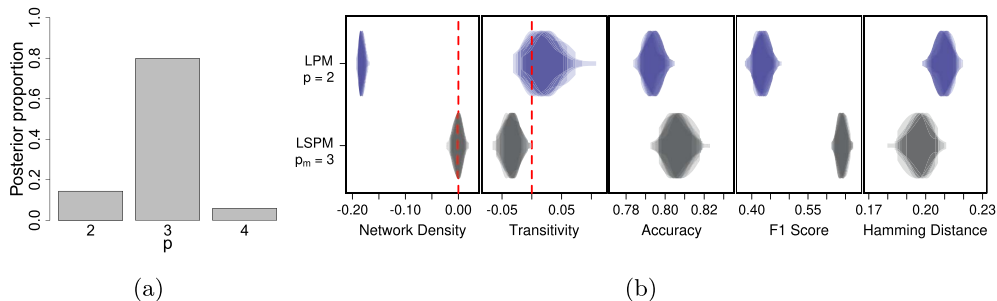


Figure 10: Cat brain connectivity data: (a) the posterior distribution of p and (b) posterior predictive checks for the LPM with $p = 2$ and the logistic LSPM with $p_m = 3$.

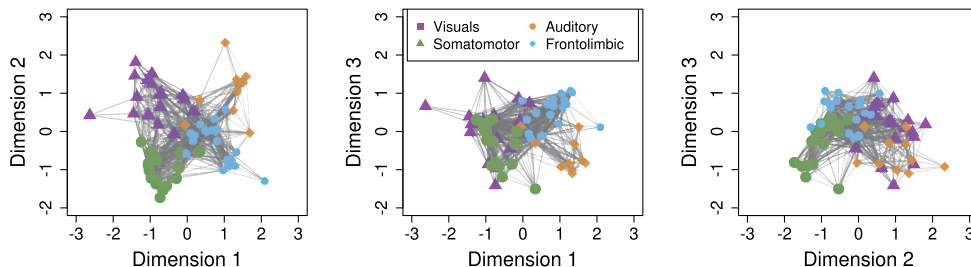


Figure 11: Posterior mean latent positions under the logistic LSPM with $p_m = 3$ for the cat brain connectivity network.

5.3 Worm nervous system binary network data

This binary directed network contains $n = 272$ nodes of neurons from the nervous system of the *Caenorhabditis elegans* adult male worm, with each of the 4451 edges representing the presence of either a chemical or electrical interaction between nodes. The data were reconstructed from serial electron micrograph sections by Jarrell et al. (2012). Three unconnected nodes are removed prior to analysis. The network is sparse with a density of 6.09%. To fit the logistic LSPM to these data, each of 10 MCMC chains were run for 1,000,000 iterations with a burn-in period of 100,000 iterations and every 3,000th thinned.

From a range of LPMs with $p = \{1, \dots, 7\}$, the BIC suggests $p = 4$ as optimal, closely followed by $p = 3$, as shown in Figure 12(a). Under the logistic LSPM, Figure 12(b) indicates that the posterior modal dimension is $p_m = 3$, with 4 dimensions the next most probable, *a posteriori*; in addition Figure 12(c) illustrates that the shrinkage strength parameter δ_4 has large uncertainty. The LSPM approach indicates that while 3 dimensions are required to adequately describe the network, there is high, quantifiable posterior uncertainty as to whether or not an additional dimension is needed.

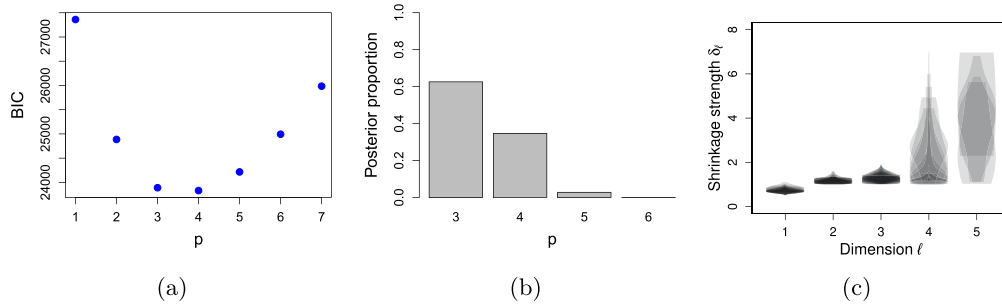


Figure 12: For the worm nervous system network: (a) the BIC for the LPM for $p = \{1, \dots, 7\}$ and under the logistic LSPM (b) the posterior distribution of p and (c) the posterior distribution of shrinkage strength parameters across the 10 MCMC chains.

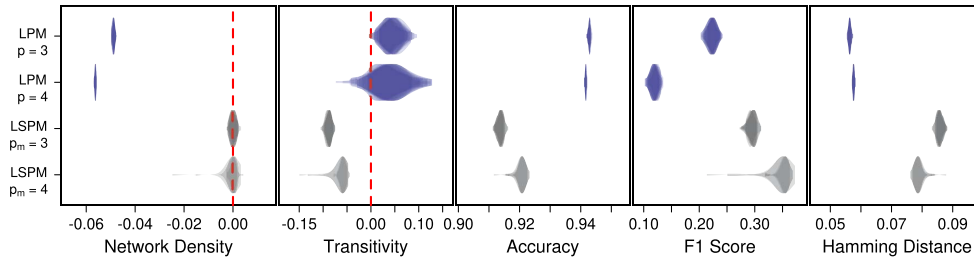


Figure 13: Posterior predictive checks on the worm nervous system network.

In terms of posterior predictive checks, Figure 13 shows that the LSPM with both 3 and 4 active dimensions fit better than the LPM with 3 and 4 dimensions in terms of network density and F_1 score, but worse in terms of transitivity, accuracy, and Hamming distance. The LSPM with $p_m = 3$ fits only slightly worse than its 4-dimensional counterpart, and so may be preferred as a lower dimensional representation. The Procrustes correlation between locations under the LSPM with $p_m = \{3, 4\}$ and the LPM with $p = \{3, 4\}$ have similar median values of more than 0.88 (standard deviation 0.05).

5.4 Phone calls count network data

Mobile phone call logs are available from the “Friends and Family” data collected by the MIT Human Dynamics Lab (Aharony et al., 2011). Here, the set of call logs from the November 2010 period is considered; the associated network contains the number of (directed) calls within a young community of $n = 120$ people. Modelling the number of calls rather than reducing the data to a binary network representing whether or not a call was made or received allows for deeper insight on the exchange of phone calls within the network. The Poisson LSPM is therefore fitted to this count network data. Fourteen disconnected nodes are removed prior to analysis. The data are overdispersed, with a mean count of 0.51 and a variance of 33.29.

A total of 10 MCMC chains were run, each from different starting configurations and with 500,000 iterations, a burn-in of 100,000 iterations and every 2,000th thinned. Figure 14(a) shows that under a Poisson LSPM the posterior explored models with 3 to 5 active dimensions, with $p_m = 4$. Fitting the Poisson LPM with $p = 1, \dots, 5$ suggested 1 dimension as optimal via the BIC. Comparing the posterior mean positions under the $p = 1$ LPM and the first dimension of the $p_m = 4$ Poisson LSPM gave a Procrustes correlation of 0.7, indicating some discrepancy. In terms of posterior predictive checks, Figure 14(b) shows that the Poisson LSPM posterior predictive networks tend to be more overdispersed than the observed network; considering a model that accounts for such overdispersion is likely to give improved model fit.

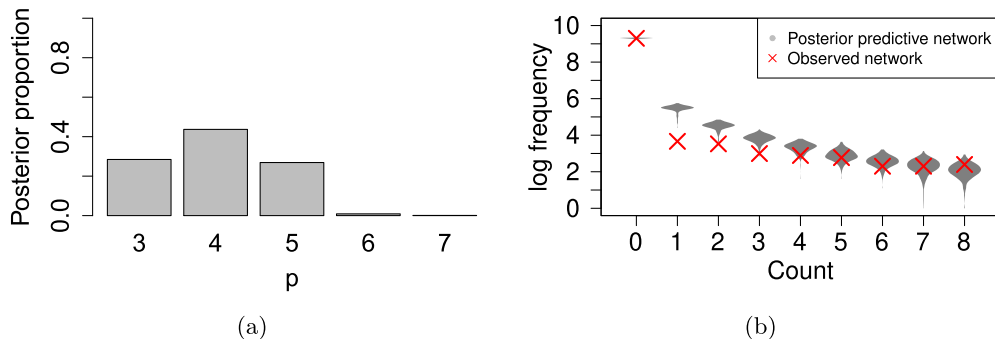


Figure 14: For the phone calls count network data: (a) the posterior distribution of p and (b) the log frequency of the observed and posterior predictive call counts for the first 8 counts in the network.

6 Discussion

The proposed latent shrinkage position model is a nonparametric Bayesian approach to model network data that focuses on the issue of inferring the dimension of the latent space. The LSPM extends the latent position model by employing a multiplicative truncated gamma process prior to allow an infinite number of dimensions in the latent space. In practice, the LSPM intrinsically infers the number of effective dimensions required to describe the network. The LSPM therefore circumvents the more computationally intensive approach where model selection criteria are used to choose the dimension after fitting a range of models with different dimensions p . Importantly, the LSPM retains the ease of interpretability inherent to the original LPM. The logistic and Poisson LSPM are developed for the analysis of binary and count network data respectively; extensions to alternative models for different link types are similarly feasible.

While fitting the LSPM is computationally feasible on the networks considered here, fitting the LSPM to large networks is relatively slow. There are many improvements that could be made to shorten run time, for example, by embedding the case-control concepts of Raftery et al. (2012) and by implementing the variational Bayesian inference of Salter-Townshend and Murphy (2013).

Since their introduction, both the concepts of the LPM for network data and the Bayesian nonparametric multiplicative gamma shrinkage prior have received much attention in a range of fields; as such the LSPM framework developed here should be of interest to a wide array of researchers and practitioners. The provision of the associated [lspm](#) R package will assist in usage of the Bayesian nonparametric LSPM models.

Supplementary Material

Supplementary to “A Latent Shrinkage Position Model for Binary and Count Network Data” (DOI: [10.1214/23-BA1403SUPP](https://doi.org/10.1214/23-BA1403SUPP); .pdf). The supplementary material contains the derivations of the properties of the LSPM and of the full conditional distributions required for the MCMC algorithm. Additional plots relating to the simulation studies and data applications are also included

References

- Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. (2011). “Social fMRI: Investigating and shaping social mechanisms in the real world.” *Pervasive and Mobile Computing*, 7(6): 643–659. [23](#)
- Armagan, A., Dunson, D. B., and Lee, J. (2013). “Generalized double Pareto shrinkage.” *Statistica Sinica*, 23(1): 119–143. [MR3076161](#). [5](#)
- Bhattacharya, A. and Dunson, D. B. (2011). “Sparse Bayesian infinite factor models.” *Biometrika*, 98(2): 291–306. [MR2806429](#). doi: <https://doi.org/10.1093/biomet/asr013>. [2](#), [3](#), [5](#), [10](#)
- Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). “Dirichlet–Laplace priors for optimal shrinkage.” *Journal of the American Statistical Association*, 110(512): 1479–1490. [MR3449048](#). doi: <https://doi.org/10.1080/01621459.2014.960967>. [5](#)
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). “The horseshoe estimator for sparse signals.” *Biometrika*, 97(2): 465–480. [MR2650751](#). doi: <https://doi.org/10.1093/biomet/asq017>. [5](#)
- Chen, S.-J., Liao, D.-L., Chen, C.-H., Wang, T.-Y., and Chen, K.-C. (2019). “Construction and analysis of protein-protein interaction network of heroin use disorder.” *Scientific Reports*, 9(1): 4980. [2](#)
- Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*, 2nd edition. Monographs on Statistics and Applied Probability. Boca Raton: Chapman & Hall/CRC. [9](#)
- de Reus, M. A. and van den Heuvel, M. P. (2013). “Rich club organization and intermodule communication in the cat connectome.” *Journal of Neuroscience*, 33(32): 12929–12939. [20](#)

- Durante, D. (2017). “A note on the multiplicative gamma process.” *Statistics & Probability Letters*, 122: 198–204. MR3584158. doi: <https://doi.org/10.1016/j.spl.2016.11.014>. 2, 5, 6
- Durante, D. and Dunson, D. B. (2014). “Nonparametric Bayes dynamic modelling of relational data.” *Biometrika*, 101(4): 883–898. MR3286923. doi: <https://doi.org/10.1093/biomet/asu040>. 3
- Durante, D., Dunson, D. B., and Vogelstein, J. T. (2017). “Nonparametric Bayes modeling of populations of networks.” *Journal of the American Statistical Association*, 112(520): 1516–1530. MR3750873. doi: <https://doi.org/10.1080/01621459.2016.1219260>. 3
- D’Angelo, S., Murphy, T. B., and Alfö, M. (2019). “Latent space modelling of multi-dimensional networks with application to the exchange of votes in Eurovision song contest.” *The Annals of Applied Statistics*, 13(2). MR3963557. doi: <https://doi.org/10.1214/18-A0AS1221>. 1, 4, 5
- Erdős, P. and Rényi, A. (1959). “On random graphs. I.” *Publicationes Mathematicae*, 6: 290–297. MR0120167. doi: <https://doi.org/10.5486/pmd.1959.6.3-4.12>. 2
- Fosdick, B. K., McCormick, T. H., Murphy, T. B., Ng, T. L. J., and Westling, T. (2019). “Multiresolution network models.” *Journal of Computational and Graphical Statistics*, 28(1): 185–196. MR3939381. doi: <https://doi.org/10.1080/10618600.2018.1505633>. 1
- Friel, N., Rastelli, R., Wyse, J., and Raftery, A. E. (2016). “Interlocking directorates in Irish companies using a latent space model for bipartite networks.” *Proceedings of the National Academy of Sciences*, 113(24): 6629–6634. 2
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*, 3rd edition. CRC Press. MR3235677. 11, 21
- Gilbert, E. N. (1959). “Random graphs.” *The Annals of Mathematical Statistics*, 30(4): 1141–1144. MR0108839. doi: <https://doi.org/10.1214/aoms/1177706098>. 2
- Gollini, I. and Murphy, T. B. (2016). “Joint modeling of multiple network views.” *Journal of Computational and Graphical Statistics*, 25(1): 246–265. MR3474046. doi: <https://doi.org/10.1080/10618600.2014.978006>. 4, 5
- Gormley, I. C. and Murphy, T. B. (2010). “A mixture of experts latent position cluster model for social network data.” *Statistical Methodology*, 7(3): 385–405. MR2643609. doi: <https://doi.org/10.1016/j.stamet.2010.01.002>. 2, 4, 8, 9
- Gwee, X. Y., Gormley, I. C., and Fop, M. (2023). “Supplementary Material for “A latent shrinkage position model for binary and count network data”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/23-BA1403SUPP>. 6
- Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). “Model-based clustering for social networks.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2): 301–354. MR2364300. doi: <https://doi.org/10.1111/j.1467-985X.2007.00471.x>. 2, 4

- Hoff, P. (2003). “Random effects models for network data.” In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, 303–312. The National Academies Press. 4
- Hoff, P. D. (2005). “Bilinear mixed-effects models for dyadic data.” *Journal of the American Statistical Association*, 100(469): 286–295. MR2156838. doi: <https://doi.org/10.1198/016214504000001015>. 4
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). “Latent space approaches to social network analysis.” *Journal of the American Statistical Association*, 97(460): 1090–1098. MR1951262. doi: <https://doi.org/10.1198/016214502388618906>. 2, 3, 4
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). “Stochastic blockmodels: First steps.” *Social Networks*, 5(2): 109–137. MR0718088. doi: [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7). 2
- Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., and Emmons, S. W. (2012). “The connectome of a decision-making neural network.” *Science*, 337(6093): 437–444. 22
- Jo, W., Chang, D., You, M., and Ghim, G.-H. (2021). “A social network analysis of the spread of COVID-19 in South Korea and policy implications.” *Scientific Reports*, 11(1): 8581. 2
- Kim, B., Lee, K. H., Xue, L., and Niu, X. (2018). “A review of dynamic network models with latent variables.” *Statistics Surveys*, 12: 105–135. MR3850294. doi: <https://doi.org/10.1214/18-SS121>. 2
- Kolaczyk E.D. and Csárdi G. (2020). *Statistical analysis of network data with R*, 2nd edition. Springer. 8
- Krivitsky, P. N. and Handcock, M. S. (2020). *latentnet: Latent Position and Cluster Models for Statistical Networks*. The Statnet Project (<https://statnet.org>). R package version 2.10.5. 14, 17, 21
- Legramanti, S., Durante, D., and Dunson, D. B. (2020). “Bayesian cumulative shrinkage for infinite factorizations.” *Biometrika*, 107(3): 745–752. MR4138988. doi: <https://doi.org/10.1093/biomet/asaa008>. 2, 5
- Liu, Y. and Chen, Y. (2021). “Variational inference for latent space models for dynamic networks.” *Statistica Sinica*. MR4478193. 1, 2
- Ma, Z., Ma, Z., and Yuan, H. (2020). “Universal latent space model fitting for large networks with edge covariates.” *Journal of Machine Learning Research*, 21(4): 1–67. MR4071187. doi: <https://doi.org/10.1109/tmls.2020.3010690>. 2
- McCullagh, P. and Nelder, J. A. (1998). *Generalized Linear Models*, 2nd edition. Monographs on Statistics and Applied Probability. Boca Raton: Chapman & Hall/CRC. MR3223057. doi: <https://doi.org/10.1007/978-1-4899-3242-6>. 4
- Murphy, K., Viroli, C., and Gormley, I. C. (2020). “Infinite mixtures of infinite factor

- analysers.” *Bayesian Analysis*, 15(3). MR4132655. doi: <https://doi.org/10.1214/19-BA1179>. 2, 10
- Ng, T. L. J., Murphy, T. B., Westling, T., McCormick, T. H., and Fosdick, B. (2021). “Modeling the social media relationships of Irish politicians using a generalized latent space stochastic blockmodel.” *The Annals of Applied Statistics*, 15(4). MR4355082. doi: <https://doi.org/10.1214/21-aos1483>. 2, 4
- Passino, F. S. and Heard, N. A. (2020). “Bayesian estimation of the latent dimension and communities in stochastic blockmodels.” *Statistics and Computing*, 30(5): 1291–1307. MR4137252. doi: <https://doi.org/10.1007/s11222-020-09946-6>. 2
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 3
- Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). “Fast inference for the latent space network model using a case-control approximate likelihood.” *Journal of Computational and Graphical Statistics*, 21(4): 901–919. MR3005803. doi: <https://doi.org/10.1080/10618600.2012.679240>. 24
- Rastelli, R. (2018). “The Sparse Latent Position Model for nonnegative weighted networks.” 2, 4
- Rastelli, R., Friel, N., and Raftery, A. E. (2016). “Properties of latent variable network models.” *Network Science*, 4(4): 407–432. 2
- Ryan, C., Wyse, J., and Friel, N. (2017). “Bayesian model selection for the latent position cluster model for social networks.” *Network Science*, 5(1): 70–91. 1
- Sajedianfard, N., Hadian, E., Samadi, A. H., Dehghan Shabani, Z., Sarkar, S., and Robinson, P. A. (2021). “Quantitative analysis of trade networks: Data and robustness.” *Applied Network Science*, 6(1): 46. 2
- Salter-Townshend, M. and McCormick, T. H. (2017). “Latent space models for multi-view network data.” *The Annals of Applied Statistics*, 11(3): 1217–1244. MR3709558. doi: <https://doi.org/10.1214/16-A0AS955>. 5
- Salter-Townshend, M. and Murphy, T. B. (2013). “Variational Bayesian inference for the Latent Position Cluster Model for network data.” *Computational Statistics & Data Analysis*, 57(1): 661–671. MR2981116. doi: <https://doi.org/10.1016/j.csda.2012.08.004>. 24
- Salter-Townshend, M., White, A., Gollini, I., and Murphy, T. B. (2012). “Review of statistical network analysis: models, algorithms, and software.” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(4): 243–264. MR2958152. doi: <https://doi.org/10.1002/sam.11146>. 4
- Sagr, M. and Alamro, A. (2019). “The role of social network analysis as a learning analytics tool in online problem based learning.” *BMC Medical Education*, 19(1): 160. 2
- Scannell, J., Blakemore, C., and Young, M. (1995). “Analysis of connectivity in the cat cerebral cortex.” *The Journal of Neuroscience*, 15(2): 1463–1483. 20

- Sewell, D. K. (2021). “Model-based edge clustering.” *Journal of Computational and Graphical Statistics*, 30(2): 390–405. MR4270512. doi: <https://doi.org/10.1080/10618600.2020.1811104>. 2
- Sewell, D. K. and Chen, Y. (2015). “Latent space models for dynamic networks.” *Journal of the American Statistical Association*, 110(512): 1646–1657. MR3449061. doi: <https://doi.org/10.1080/01621459.2014.988214>. 4, 5
- Sewell, D. K. and Chen, Y. (2016). “Latent space models for dynamic networks with weighted edges.” *Social Networks*, 44: 105–116. 2, 4
- Smith, A. L., Asta, D. M., and Calder, C. A. (2019). “The geometry of continuous latent space models for network data.” *Statistical Science*, 34(3). MR4017522. doi: <https://doi.org/10.1214/19-STS702>. 2
- Sosa, J. and Betancourt, B. (2022). “A latent space model for multilayer network data.” *Computational Statistics & Data Analysis*, 169: 107432. MR4369145. doi: <https://doi.org/10.1016/j.csda.2022.107432>. 2
- Tafakori, L., Pourkhanali, A., and Rastelli, R. (2021). “Measuring systemic risk and contagion in the European financial network.” *Empirical Economics*. 2
- Wu, J. (2012). *Advances in K-Means Clustering: A Data Mining Thinking*. Springer Theses. Heidelberg Berlin: Springer. MR2953540. doi: <https://doi.org/10.1007/978-3-642-29807-3>. 9
- Yang, C., Priebe, C. E., Park, Y., and Marchette, D. J. (2020). “Simultaneous dimensionality and complexity model selection for spectral graph clustering.” *Journal of Computational and Graphical Statistics*, 30(2): 422–441. MR4270514. doi: <https://doi.org/10.1080/10618600.2020.1824870>. 2
- Zachary, W. W. (1977). “An information flow model for conflict and fission in small groups.” *Journal of Anthropological Research*, 33(4): 452–473. 20
- Zhang, X., Xue, S., and Zhu, J. (2020). “A flexible latent space model for multilayer networks.” In Daumé, H. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 11288–11297. PMLR. 2

Acknowledgments

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The authors are grateful to the anonymous reviewers for their suggestions which greatly contributed to improving this work.