Network Science (2025), **13**, e19, 1–33
doi:10.1017/nws.2025.10014

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Model-based clustering for network data via a latent shrinkage position cluster model

Xian Yao Gwee, Isobel Claire Gormley* and Michael Fop*

School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
Corresponding author: Michael Fop; Email: michael.fop@ucd.ie

*Isobel Claire Gormley and Michael Fop have contributed equally to this work.

Abstract

Low-dimensional representation and clustering of network data are tasks of great interest across various fields. Latent position models are routinely used for this purpose by assuming that each node has a location in a low-dimensional latent space and by enabling node clustering. However, these models fall short through their inability to simultaneously determine the latent space dimension and number of clusters. Here we introduce the latent shrinkage position cluster model (LSPCM), which addresses this limitation. The LSPCM posits an infinite-dimensional latent space and assumes a Bayesian nonparametric shrinkage prior on the latent positions' variance parameters resulting in higher dimensions having increasingly smaller variances, aiding the identification of dimensions with non-negligible variance. Further, the LSPCM assumes the latent positions follow a sparse finite Gaussian mixture model, allowing for automatic inference on the number of clusters related to non-empty mixture components. As a result, the LSPCM simultaneously infers the effective dimension of the latent space and the number of clusters, eliminating the need to fit and compare multiple models. The performance of the LSPCM is assessed via simulation studies and demonstrated through application to two real Twitter network datasets from sporting and political contexts. Open-source software is available to facilitate widespread use of the LSPCM.

Keywords: Latent position model; network analysis; mixture models; Bayesian nonparametric priors; multiplicative gamma process

1. Introduction

Network data are an important class of structured data where objects are represented as nodes and the relationships between these objects are represented as edges. Network data arise in various fields, including epidemiology (Jo et al., 2021), neuroscience (Yang et al., 2020), brain connectivity (Aliverti and Durante, 2019), and sociology (D'Angelo et al., 2019). A key interest in the analysis of network data is the task of clustering the nodes in the network, where the aim is to group together nodes that share similar characteristics or behaviors.

While a variety of network models exist, many originate from the influential Erdös–Rényi random graph model (Erdos and Rényi, 1959; Gilbert, 1959), including the widely utilized stochastic block model (Holland, et al., 1983; Snijders and Nowicki, 1997) and the latent position model (LPM, Hoff, et al., 2002). Here, we focus on the LPM which assumes each node in a network has a position in a latent p-dimensional space. Under the LPM, the probability of an edge between two nodes is determined by their proximity in the latent space. The latent position cluster model (LPCM, Handcock, et al., 2007) extended the LPM to allow for model-based clustering of nodes in

a network by assuming the latent positions are generated from a finite mixture of Gaussian distributions. Various extensions to the LPCM have been proposed, including approaches that account for random effects (Krivitsky et al., 2009), that incorporate covariates through a mixture of experts framework (Gormley and Murphy, 2010), that employ a variational approach to Bayesian inference (Salter-Townshend and Murphy, 2013), that model longitudinal networks (Sewell and Chen, 2017), that allow for edge clustering (Sewell, 2020; Pham and Sewell, 2024), and that model multiplex networks (D'Angelo et al., 2023); see Rastelli, et al. (2016); Sosa and Buitrago (2021); Kaur et al. (2024) for comprehensive reviews.

Although the LPCM is widely used, inferring the number of clusters and the dimensionality of the latent space remains a challenging task. In practice, the numbers of clusters and dimensions are typically treated as user-defined parameters, with the latter set as 2 in the majority of cases to allow for simple visualization and interpretation (D'Angelo et al., 2023; Liu and Chen, 2022). Often, a set of LPCMs with different numbers of clusters and dimensions are fitted to the network data and a model selection criterion is used to select the optimal model. Many model selection criteria have been used for this purpose e.g., Handcock et al. (2007) use a variant of the Bayesian information criterion (BIC) to select optimal numbers of clusters and dimensions but highlight its lack of robustness. Other model selection criteria such as the Watanabe–Akaike information criterion (WAIC) (Ng et al., 2021; Sosa and Betancourt, 2022), the Akaike information criterion (AIC) and the integrated completed likelihood (ICL) (Sewell, 2020) have also been used. While these criteria have found some success, they can give conflicting inference on the same data and their use requires fitting a large set of models, each with a different combination of number of clusters and number of dimensions, which becomes computationally prohibitive as the set of possible models grows.

Alternative, automated strategies for inferring the numbers of clusters and dimensions from the network data have emerged. In the context of stochastic block models, Yang et al. (2020) utilize a frequentist framework, while Passino and Heard (2020) employ a Bayesian framework for this purpose. In the LPCM setting, automatic inference on the number of clusters has been considered in D'Angelo et al. (2023) via an infinite mixture model, while Durante and Dunson (2014) and Gwee, et al. (2024a) employed nonparametric shrinkage priors to address the latent space dimensionality. However, automated, simultaneous inference of both the number of clusters and the effective latent space dimension in the LPCM setting has not yet been considered.

Here the latent shrinkage position cluster model (LSPCM) is introduced, which simultaneously infers the node clustering structure and the effective latent space dimension. To achieve clustering of the nodes, a sparse finite mixture model (Malsiner-Walli et al., 2014, 2017) is employed that overfits the number of components in the mixture model. The adoption of a sparse prior on the mixture weights encourages emptying of redundant components, thereby allowing inference on the number of clusters. Within each cluster, a Bayesian nonparametric truncated gamma process shrinkage prior (Bhattacharya and Dunson, 2011; Gwee, et al., 2024a) is placed on the variance of the nodes' positions in the latent space. While the latent space is assumed to have infinitely many dimensions, the shrinkage prior implies that higher dimensions have negligible variance and therefore are non-informative. The LSPCM eliminates the need for choosing a model selection criterion and only requires the fitting of a single model to simultaneously infer both the optimal number of clusters and the effective latent space dimension, reducing the required computation time. Additionally, the Bayesian framework naturally provides uncertainty quantification for the number of clusters and the number of effective dimensions.

The remainder of this article is structured as follows: Section 2 describes the proposed LSPCM while Section 3 outlines the inferential process along with practical implementation details. Section 4 describes simulation studies conducted to explore the performance of the LSPCM in a variety of settings where the numbers of nodes, dimensions, and clusters, as well as cluster volumes, vary. Section 5 applies the proposed LSPCM to two real Twitter network data sets: one

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

concerning football players in the English premier league and one concerning the Irish political context. Section 6 concludes and discusses potential extensions. R (R Core Team, 2024) code with which all results presented herein were produced is freely available from the lspm GitLab repository.

2. The latent shrinkage position cluster model

To cluster nodes in a network, we introduce the novel LSPCM which draws on and fuses together ideas from both the latent shrinkage position model (LSPM) of Gwee et al. (2024a), the LPCM of Handcock et al. (2007), and the sparse finite mixture models of Malsiner-Walli et al. (2014).

2.1 The latent shrinkage position model

Network data typically take the form of an $n \times n$ adjacency matrix, $\mathbf{Y}$, where n is the number of nodes and entry $y_{i,j}$ denotes the relationship or edge between nodes i and j. Self-loops are not permitted and thus the diagonal elements of $\mathbf{Y}$ are zero. Here we consider binary edges but a variety of edge types can be considered.

Under the LPM (Hoff, et al., 2002), edges are assumed independent, conditional on the latent positions of the nodes. The sampling distribution is then

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

where $\mathbf{Z}$ is the matrix of latent positions, with $\mathbf{z}_i$ denoting the latent position of node i; α is a global parameter that captures the overall connectivity level in the network. Denoting by $q_{i,j}$ the probability of an edge between nodes i and j, i.e. $\mathbb{P}(y_{i,j} = 1 \mid \alpha, \mathbf{z}_i, \mathbf{z}_j)$, a logistic regression model formulation is used where the log odds of an edge between nodes i and j depends on the Euclidean distance between their respective positions $\mathbf{z}_i$ and $\mathbf{z}_j$ in the latent space,

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

As in Gollini and Murphy (2016) and D'Angelo et al. (2019), in (1) the distance is taken to be the squared Euclidean distance, giving higher weight to the latent space in the model than that induced by a linear Euclidean distance.

Gwee et al. (2024a) propose the LSPM, a Bayesian nonparametric extension of the LPM that allows automatic inference on the number of the infinite latent space dimensions that are necessary to fully describe the network i.e., the number that are effective since they have non-negligible latent position variance. Under the LSPM, the latent positions are assumed to have a zero-centered Gaussian distribution with diagonal precision matrix Ω, whose entries ω_ℓ denote the precision of the latent positions in dimension ℓ, for $\ell = 1, \ldots, \infty$. Spherical Gaussian distributions are assumed, as is standard in latent position models, since the likelihood is invariant to rotations of the latent space (Handcock, et al., 2007). The LSPM employs a multiplicative truncated gamma process (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 δ_1, while a truncated gamma distribution is assumed for the remaining dimensions to ensure shrinkage. Specifically, for $i = 1, \ldots, n$

$$\mathbf{z}_i \sim \text{MVN}(\mathbf{0}, \Omega^{-1})$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

$$\Omega = \begin{bmatrix} \omega_1^{-1} & \dots & 0 \\ \vdots & \ddots & \\ 0 & & \omega_\infty^{-1} \end{bmatrix} \qquad \omega_\ell = \prod_{h=1}^{\ell} \delta_h \text{ for } \ell = 1, \dots, \infty$$

$$\delta_1 \sim \text{Gam}(a_1, b_1 = 1) \qquad \delta_h \sim \text{Gam}^{\text{T}}(a_2, b_2 = 1, t_2 = 1) \text{ for } h > 1.$$

Here a_1 and b_1 are the gamma prior's shape and rate parameters on the first dimension's shrinkage parameter, while a_2 is the shape parameter, b_2 is the rate parameter, and t_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 of the positions in the higher dimensions of the latent space. Further details regarding these hyperparameters and their choices are given in Appendix A. 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 meaningful information encoded in them as the distances between nodes will be close to zero. Thus, under the LSPM, the effective dimensions are those in which the variance is non-negligible (Gwee, et al., 2024a).

2.2 The latent shrinkage position cluster model

To infer the unknown number of clusters of nodes, we propose the novel LSPCM which results from fusing the LSPM (Gwee, et al., 2024a), the latent position cluster model (LPCM) of Handcock, et al. (2007) and the sparse finite mixture model framework of Malsiner-Walli et al. (2014, 2017). Like the LPCM, the proposed LSPCM assumes that latent positions arise from a finite mixture of G spherical multivariate normal distributions; however, unlike the LPCM, the LSPCM assumes an infinite-dimensional latent space as in the LSPM. Further, the LSPCM employs a sparsity-inducing Dirichlet prior on the mixture weights, facilitating automatic inference on the number of clusters.

Specifically, the LSPCM assumes, for $i = 1, \dots, n$,

$$\mathbf{z}_i \sim \sum_{g=1}^{G} \tau_g \text{MVN}(\boldsymbol{\mu}_g, \psi_g^{-1} \Omega^{-1}) \tag{2}$$

$$\boldsymbol{\mu}_g | \Omega \sim \text{MVN}(\mathbf{0}, \xi \Omega^{-1}) \qquad \psi_g \sim \text{Gam}(a_\psi, b_\psi)$$

$$\Omega = \begin{bmatrix} \omega_1^{-1} & \dots & 0 \\ \vdots & \ddots & \\ 0 & & \omega_\infty^{-1} \end{bmatrix} \qquad \omega_\ell = \prod_{h=1}^{\ell} \delta_h \text{ for } \ell = 1, \dots, \infty$$

$$\delta_1 \sim \text{Gam}(a_1, b_1 = 1) \quad \delta_h \sim \text{Gam}^{\text{T}}(a_2, b_2 = 1, t_2 = 1) \text{ for } h > 1.$$

Here τ_g denotes the probability that a node belongs to the g-th component, so that $\tau_g \geq 0$ $(g = 1, \dots, G)$ and $\sum_{g=1}^{G} \tau_g = 1$. For the mean latent position $\boldsymbol{\mu}_g$ of the g-th component, a multivariate Gaussian prior with zero mean and a scaled version of the same precision matrix as the latent positions with its MTGP prior is assumed; this ensures the latent positions will all shrink towards zero at higher dimensions. Consequently, the component means are increasingly shrunk towards zero, resulting in higher dimensions characterized by increasingly overlapping mixture components, and, as a result, becoming less informative for cluster separation. Further, as in Ryan et al. (2017), the prior covariance on $\boldsymbol{\mu}_g$ is inflated by a scaling factor ξ; values of

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

$\xi > 1$ result in cluster means that are more dispersed than cluster members. While a range of values of $\xi = \{1, 9, 100\}$ were considered, posterior predictive checks suggested that, similar to Ryan et al. (2017), $\xi = 9$ showed good model fit by encouraging both separate yet connected clusters. Additionally, a cluster-specific factor ψ_g scales the precision matrix of the latent positions providing the flexibility to model clusters of different volumes.

While the LPCM assumes a finite mixture of spherical multivariate normal distributions and casts the problem of inferring the number of clusters as one of model selection, the LSPCM considers an overfitted (or sparse) finite mixture model which provides automatic inference on the number of clusters. Here, following Malsiner-Walli et al. (2014); Fruhwirth-Schnatter, et al. (2019), a symmetric Dirichlet prior is placed on the mixture weights $\boldsymbol{\tau}$ with a gamma prior placed on the Dirichlet's influential sparsity-inducing hyperparameter ν, i.e.,

$$\boldsymbol{\tau} = (\tau_1, \ldots, \tau_G) \sim \text{Dir}(\nu, \ldots, \nu) \qquad \nu \sim \text{Gam}(a_\nu, Gb_\nu).$$

Under such an overfitted mixture, a distinction is made between the number of mixture components G, the number of non-empty mixture components G_+, and the number of clusters in the data G^*. A large number G of initial components is specified and under a gamma prior that favors small values of ν, unnecessary components are emptied out during the inferential process. The number of non-empty components G_+ then serves as an estimate of the number of clusters G^*. Thus, the LSPCM allows for automatic, simultaneous inference of the number of clusters and of the number of effective latent space dimensions, while requiring only a single model to be fit to the network data.

While the LSPM adopts a nonparametric approach, an overfitted sparse finite mixture approach for the clustering aspect of LSPCM is employed here as the sampling algorithm is simpler than that required to fit an infinite mixture model. Further, as detailed in Frühwirth-Schnatter and Malsiner-Walli (2018), under certain hyperparameter specifications, similar inference with respect to the number of clusters is obtained under the sparse finite and nonparametric approaches. Additionally, Frühwirth-Schnatter and Malsiner-Walli (2018) and Murphy, et al. (2020) highlight that overfitted mixtures serve well in situations where the data arise from a moderate number of clusters, even with growing sample size, whereas infinite mixtures are suited to cases where the number of clusters increases with sample size. As the networks considered here are anticipated to arise from a moderate number of clusters, an overfitted approach is preferred and taken.

To facilitate inference, we introduce $C = (c_1, \ldots, c_n)$ which contains the latent component membership vectors $c_i = (c_{i1}, \ldots, c_{iG})$, for $i = 1, \ldots, n$, where $c_{ig} = 1$ if node i belongs to component g and $c_{ig} = 0$ otherwise. Thus, $c_i \sim \text{MultiNom}(1, \boldsymbol{\tau})$. A summarizing graphical model representation of the LSPCM is shown in Figure 1.

3. Inference

Denoting by $\boldsymbol{\Theta}$ the component means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_G$, the precision scaling factors $\psi_1, \ldots, \psi_G$ and the precision matrix $\boldsymbol{\Omega}$, the joint posterior distribution of the LSPCM is then

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

where $\mathbb{P}(\alpha)$ is the non-informative $N(\mu_\alpha = 0, \sigma_\alpha^2 = 4)$ prior for the α parameter, and $\mathbb{P}(\boldsymbol{\tau})$ and $\mathbb{P}(\boldsymbol{\Theta})$ denote the prior distributions outlined in Section 2.2. For the MTGP priors, similar to Durante (2017) and Gwee, et al. (2024a), we set the hyperparameters to $a_1 = 2$ and $a_2 = 3$. A gamma hyperprior is used on the hyperparameter ν with shape $a_\nu = 5$ and rate $b_\nu = 5$ that is found to work well empirically in the simulation studies. For the gamma prior on ψ_g, as LPMs can only model networks of connected nodes, and as uncovering clustering structure is not possible if some clusters are so variable that they mask others, only values of ψ_g near 1 are of interest. A strongly informed gamma prior centered on 1 is therefore assumed for ψ_g with $a_\psi = 400$ and $b_\psi = 400$,

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

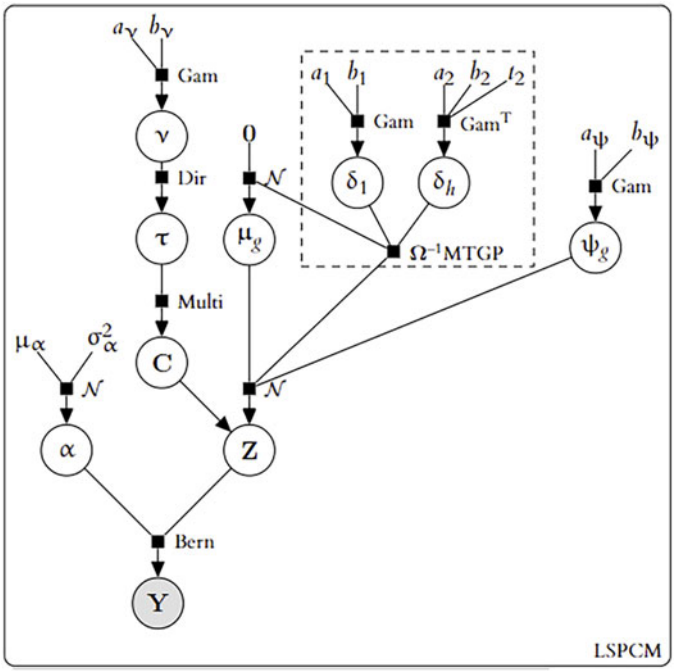

Figure 1. The graphical model representation of the LSPCM.

giving a prior concentrated around 1. A summary of the LSPCM hyperparameter settings used is detailed in Appendix A.

3.1 An adaptive Metropolis-within-Gibbs sampler

Markov chain Monte Carlo (MCMC) is employed to draw samples from the joint posterior distribution. After defining necessary notation in Appendix B, derivations of the full conditional distributions of the latent positions, cluster membership vectors and model parameters are given in Appendix C. As the MTGP prior is nonparametric and assumes an infinite number of latent dimensions, in practice setting an initial finite truncation level, p_0, on the number of dimensions fitted is required. Following Bhattacharya and Dunson (2011), an adaptive Metropolis-within-Gibbs sampler is employed where the truncation dimension $p \leq p_0$ is dynamically shrunk or augmented as the sampler proceeds, with inferential validity ensured as detailed in Bhattacharya and Dunson (2011). The number of effective dimensions, $\hat{p}$, can then be subsequently inferred (see Section 3.2.3).

Denoting the current iteration by s for $s = 1, \ldots, S$ (but for clarity s is omitted where it is unnecessary), a single pass of the sampler proceeds as follows:

1. Sample the component mean $\boldsymbol{\mu}_g$ for $g = 1, \ldots, G$ from $\text{MVN}_p\left(\frac{\sum_{i=1}^n c_{ig}\mathbf{z}_i}{\sum_{i=1}^n c_{ig}+\xi^{-1}}, \left[\boldsymbol{\Omega}\left(\sum_{i=1}^n c_{ig} + \xi^{-1}\right)\right]^{-1}\right)$.
2. Sample $\check{\nu}$ from a $\text{Gam}(\sigma_\nu, \frac{\sigma_\nu}{\nu^{(s)}})$ where σ_ν is a step size factor. Accept $\check{\nu}$ as $\nu^{(s+1)}$ with probability $\frac{p(\nu=\check{\nu}|-)}{p(\nu=\nu^{(s)}|-)}\frac{f(\check{\nu};a_\nu,b_\nu)}{f(\nu^{(s)};a_\nu,b_\nu)}$, where $p(\nu \mid -)$ is the full conditional of ν and $f(\,\cdot\,;a_\nu, b_\nu)$ is the gamma density. Otherwise, set $\nu^{(s+1)} = \nu^{(s)}$.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

3. Sample the mixing weights $\boldsymbol{\tau}$ from $\text{Dir}(\sum_{i=1}^{n} c_{i1} + \nu, \ldots, \sum_{i=1}^{n} c_{iG} + \nu)$.
4. Sample the cluster-specific precision scaling parameter ψ_g for $g = 1, \ldots, G$ from $\text{Gam}\left(a_\psi + \frac{p \sum_{i=1}^{n} c_{ig}}{2}, \quad b_\psi + \frac{1}{2} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T \Omega (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig}\right)$.
5. Sample the latent memberships c_i for each node $i = 1, \ldots, n$ from a Multinomial with G categories by drawing the g-th category with probability $\frac{\tau_g \phi_p(\mathbf{z}_i; \boldsymbol{\mu}_g, \psi_g^{-1} \boldsymbol{\Omega}^{-1})}{\sum_{r=1}^{G} \tau_r \phi_p(\mathbf{z}_i; \boldsymbol{\mu}_r, \psi_g^{-1} \boldsymbol{\Omega}^{-1})}$ where $\phi_p(\mathbf{z}_i; \boldsymbol{\mu}, \psi_g^{-1} \boldsymbol{\Omega}^{-1})$ is the p-dimensional multivariate normal density.
6. Sample $\check{\mathbf{Z}}$ from a $\text{MVN}_p(\mathbf{Z}^{(s)}, k\psi_g^{-1} \Omega^{-1(s)})$ proposal distribution where k is a step size factor. Accept $\check{\mathbf{Z}}$ as $\mathbf{Z}^{(s+1)}$ with probability $\frac{\mathbb{P}(\mathbf{Y} | \alpha^{(s)}, \check{\mathbf{Z}})}{\mathbb{P}(\mathbf{Y} | \alpha^{(s)}, \mathbf{Z}^{(s)})} \frac{\phi_p(\mathbf{z}_i; \check{\mathbf{Z}}, k\psi_g^{-1} \Omega^{-1(s)})}{\phi_p(\mathbf{z}_i; \mathbf{Z}^{(s)}, k\psi_g^{-1} \Omega^{-1(s)})}$, otherwise set $\mathbf{Z}^{(s+1)} = \mathbf{Z}^{(s)}$.
7. Sample $\check{\alpha}$ from an informed Gaussian proposal distribution $N(\bar{\mu}_\alpha, \bar{\sigma}_\alpha^2)$ (Gormley and Murphy, 2010) where

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

and

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

Then, accept $\check{\alpha}$ as $\alpha^{(s+1)}$ following the Metropolis-Hastings acceptance ratio.
8. Sample δ_1 from

$$\text{Gam}\left(\frac{(n+G)\, p}{2} + a_1, \ \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T \psi_g^{-1} \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} + \frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_1 \right).$$

9. Sample $\delta_h^{(s+1)}$ for $h = 2, \ldots, p$ from

$$\text{Gam}^{\text{T}}\left(\frac{(n+G)\,(p-h+1)}{2} + a_2, \ \frac{1}{2} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T \psi_g^{-1} \left(\prod_{m=1, m \neq h}^{\ell} \delta_m^{(s^*)} \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} + \frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=1, m \neq h}^{\ell} \delta_m^{(s^*)} \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_2, \quad 1 \right)$$

where $s^* = s + 1$ for $m < h$ and $s^* = s$ for $m > h$.
10. Calculate ω_ℓ by taking the cumulative product of δ_1 to δ_ℓ.
11. At iteration s, adapt the truncation dimension $p^{(s)}$ with probability $\exp(-\kappa_0 - \kappa_1 s)$ as detailed in Section 3.2.2.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

3.2 Practical implementation details

When implementing the adaptive Metropolis-within-Gibbs sampler, practical details such as initialization of latent variables and parameters, adapting the truncation dimension, and post-processing of the MCMC chain require attention.

3.2.1 Initialization

While the LSPCM assumes an infinite number of latent dimensions and a potentially large number of components in the overfitted mixture, this is computationally impractical due to finite computational resources. Setting initial values for both the truncation level of the latent dimensions, p_0, and for the number of components in the mixture, G, prevents the sampler from exploring computationally unfeasible models. Here, $p_0 = 5$ and $G = 20$ are used throughout as these settings were empirically observed to strike a balance between exploring complex models and computational feasibility for the type of networks analyzed.

The LSPCM's parameters, cluster allocations and latent positions also require initialization, achieved here as follows, where $s = 0$:

1. Calculate the geodesic distances (Kolaczyk and Csárdi, 2020) between the nodes in the network. Apply classical multidimensional scaling (Cox and Cox, 1994) to the geodesic distances and set $\mathbf{Z}^{(0)}$ to be the resulting $n \times p_0$ positions. Empirically, this approach was more computationally efficient than setting $\mathbf{Z}^{(0)}$ to the LPM solution, with little notable loss in inferential performance.
2. Fit a standard regression model, with regression coefficients α and β, to the vectorized adjacency matrix where $\text{log odds}(q_{i,j} = 1) = \alpha - \beta \|\mathbf{z}_i^{(0)} - \mathbf{z}_j^{(0)}\|_2^2$ to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$. Set $\alpha^{(0)} = \hat{\alpha}$.
3. As the LSPCM model (1) constrains $\beta = 1$, center and rescale the latent positions by setting $\mathbf{Z}^{(0)} = \sqrt{|\hat{\beta}|}\tilde{\mathbf{Z}}^{(0)}$, where $\tilde{\mathbf{Z}}^{(0)}$ are the mean-centered initial latent positions.
4. Apply model-based clustering to $\mathbf{Z}^{(0)}$ via `mclust` (Scrucca et al., 2016) with up to G clusters and the `EEI` covariance structure to obtain initial cluster allocations.
5. Obtain $\omega_\ell^{(0)}$ for $\ell = 1, \ldots, p_0$ by calculating the empirical precision of each column of $\mathbf{Z}^{(0)}$.
6. Set $\delta_1^{(0)} = \omega_1^{(0)}$ and calculate $\delta_h^{(0)} = \frac{\omega_h^{(0)}}{\omega_{h-1}^{(0)}}$ where $h = 2, \ldots, p_0$.

3.2.2 Adapting the truncation dimension

As in Bhattacharya and Dunson (2011) and Gwee et al. (2024a), the LSPCM adapts the truncation dimension p as the MCMC sampler runs. This adaptation approach is employed to ease computational load as there are often relatively few important factors. Further, under the LSPCM, later dimensions will have very small, but not exactly zero, variance, thus contributing to the distance between nodes in (1) and potentially introducing bias to parameter estimates. This behavior is considered in Gwee et al. (2024a). Similarly to Bhattacharya and Dunson (2011), the probability of adapting p at iteration s is taken as $\mathbb{P}(s) = \exp(-\kappa_0 - \kappa_1 s)$, which decreases exponentially as the MCMC chain evolves. Here, setting $\kappa_0 = 4$ and $\kappa_1 = 5 \times 10^{-4}$ demonstrated good performance in empirical experiments.

After the burn-in period, at an adaptation step, when $p > 1$, p is reduced based on the cumulative proportion of the latent position variance that the dimensions $\ell = 1, \ldots, p-1$ contain. If the dimensions up to the ℓ-th cumulatively contain a proportion greater than ϵ_1 of the total variance, then the dimensions from $\ell + 1$ to p add little information, and p is reduced to ℓ. In general,

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

$\epsilon_1 = 0.8$ was found to work well but higher values of ϵ_1 also showed good performance in networks with large n. If the 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 $\delta_p^{-1} > \epsilon_2$, p is increased by 1, with the additional associated parameters drawn from their respective priors. We consider $\epsilon_2 = 0.95$ which was found to work well in practice. The case where $p = 1$ at an adaptation step requires a different criterion: if the proportion of latent positions with absolute deviation from their empirical mean > 1.96 (the 95% critical value of a standard Normal) is greater $0.05 \times \epsilon_3$, then p is increased to 2; setting $\epsilon_3 = 5$ was found to work well in practice. Reducing p is not considered if $p = 1$.

3.2.3 Post-processing of the MCMC chain

The likelihood function of the LSPCM depends on the Euclidean distances between the latent positions; hence, it remains unaffected by rotations, reflections, or translations of these positions, giving rise to identifiability issues. To ensure valid posterior inference, similar to Gormley and Murphy (2010), a Procrustean transformation of the sampled latent positions $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(S)}$ is considered. The transformation aligns the sampled positions with a reference configuration $\tilde{\mathbf{Z}}$, which is selected based on the configuration that yields the highest log-likelihood during the burn-in phase of the MCMC chain. Although this choice is arbitrary, it has little effect as the reference configuration solely serves the purpose of addressing identifiability.

The MCMC sampler provides insight on the number of effective latent dimensions $\hat{p}$. Here, post hoc, the number of effective dimensions at each iteration s, $\hat{p}^{(s)}$, is determined as $\hat{p}^{(s)} = p^{(s)} - m^{(s)}$, where $m^{(s)}$ is the number of non-effective dimensions, i.e. dimensions with negligible variance. For each iteration, the cumulative proportion of variance of each dimension is computed. If the first $\ell^{(s)}$ dimensions cumulatively contain a proportion greater than $\epsilon_1 = 0.8$ of the total variance, then the dimensions from $\ell^{(s)} + 1$ to $p^{(s)}$ are considered as non-effective and $m^{(s)} = p^{(s)} - \ell^{(s)}$. Similar to Bhattacharya and Dunson (2011) $\hat{p}_m$, the mode of the thinned $\hat{p}^{(1)}, \ldots, \hat{p}^{(S)}$, is used to denote the modal effective dimension, with empirical intervals quantifying the associated uncertainty.

Since the MCMC samples of the component allocations will have varying numbers of non-empty components, and because of the label switching problem caused by invariance of the mixture with respect to permutation of the component labels, care must be taken when deriving posterior summaries of cluster labels and parameters. While Fruhwirth-Schnatter et al. (2019) give an overview of many potential approaches, here, as it demonstrated robust and accurate performance for the networks examined, we adopt the procedure of Fritsch and Ickstadt (2009) to estimate the cluster labels of the nodes and subsequently the number of clusters. The method, implemented in the R package `mcclust` (Fritsch, 2022), first estimates the posterior similarity matrix containing the proportion of times a pair of nodes are placed in the same cluster and then maximizes the posterior expected adjusted Rand index (PEAR) to obtain the optimal cluster labels. In addition, as in Malsiner-Walli et al. (2014), the posterior distribution of the number of filled components G_+ and the corresponding posterior mode G_m are examined to inspect the uncertainty in the number of clusters. In some situations (see Section 5.1 for an example), the posterior mode may differ from the number of clusters estimated by the PEAR method. Finally, for summarizing the posterior distributions of the cluster parameters, as in Gormley and Murphy (2010), cluster parameters are obtained after permuting cluster labels to minimize a loss function based on the cluster means.

4. Simulation studies

The performance of the LSPCM is assessed on simulated data scenarios by evaluating its ability to recover the effective latent dimension of the latent space and to correctly infer the latent positions, the number of clusters, and the cluster allocations. The latent positions are simulated according to

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

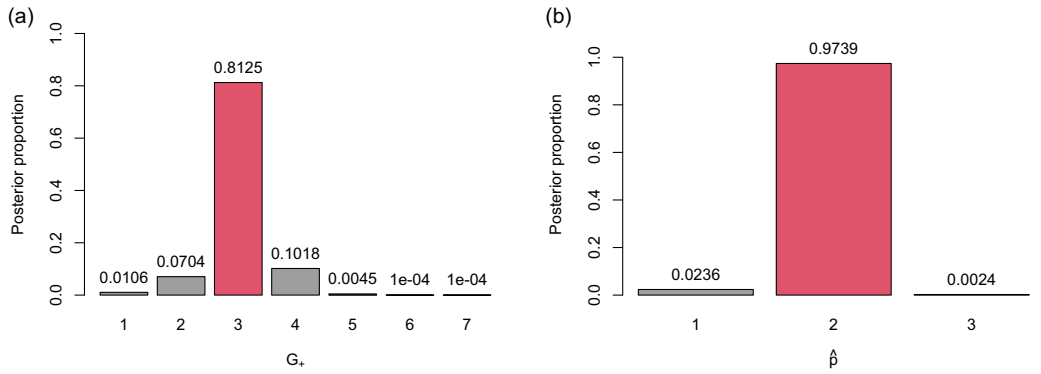

Figure 2. (a) Posterior distribution of the number of active clusters G_+ and (b) the distribution of the effective latent space dimension $\hat{p}$ across 30 small simulated networks. The true number of clusters and latent dimensions are highlighted in red.

(2) and the mixture weights $\boldsymbol{\tau}$ are generated from a symmetric Dir(10, . . . , 10) which gives rise to clusters with similar numbers of nodes. Given the latent positions, a network is then generated as in (1). Three scenarios are considered: Sections 4.1 and 4.2, respectively, assess the performance of LSPCM on networks with small and moderate numbers of nodes, dimensions, and clusters, while Section 4.3 assesses the performance when clusters have varying volumes i.e., when $\psi_g \neq \psi_g'$ for $g \neq g'$. A total of 30 networks are simulated in each scenario. In Sections 4.1 and 4.3, the MCMC chains are run for 500,000 iterations with a burn-in of 50,000 iterations, thinned every 1,000th iteration while in Section 4.2, 1,000,000 iterations are considered, with a burn-in of 100,000 iterations, thinned every 2,000th iteration due to the larger settings considered. To ensure acceptance rates in the 20% − 40% range, step sizes of $\sigma_\nu = 0.5$ and k values between 1 and 2.5 were employed in the proposal distributions.

4.1 Scenario 1: small networks

Networks with $n = 50$ are generated with the true number of latent dimensions $p^* = 2$ and true number of clusters $G^* = 3$, with shrinkage strength $\boldsymbol{\delta} = (1, 1.05)$, $\psi_1 = \psi_2 = \psi_3 = 1$, and cluster mean positions $\boldsymbol{\mu} = \{(0, 0), (-4, 0), (-4, 4)\}$ giving well-separated clusters. Across the 30 simulated networks, the smallest cluster contained 6 nodes while the largest cluster contained 31, and fixing $\alpha = 6$ gave rise to network densities between 27% and 38%.

Figure 2a shows that the posterior modal number of clusters $G_m = 3$ and Figure 2b shows that the modal effective number of dimensions is $\hat{p}_m = 2$. Overall, the adjusted Rand index (ARI, Hubert and Arabie, 1985), which measures agreement between the inferred cluster memberships and the true cluster labels, shows good correspondence, with mean value of 0.88 (standard deviation (sd) = 0.11). Procrustes correlations (PC, Peres-Neto and Jackson, 2001) between inferred and true latent positions also show that the latent positions are accurately estimated, with average value of 0.97 (sd = 0.03). Additional visualization of the posterior distributions on the variances, shrinkage strengths, and the parameter α are available in Appendix D.

Additionally, in this scenario sensitivity to the setting of ϵ_1 was explored by also considering $\epsilon_1 = 0.9$ when deciding to add a truncation dimension but with the lower value of 0.8 employed when inferring the effective number of dimensions; there was no change in the modal effective dimension inferred, with only a small change in the posterior probability of $\hat{p} = 3$ to 0.9225.

For comparison, the LPCM is fitted on the same simulated data using the R package `latentnet` (Krivitsky and Handcock, 2008). Upon fitting 25 LPCMs with different combinations of $p = \{1, \ldots, 5\}$ and $G = \{1, \ldots, 5\}$ to each of the 30 simulated networks, the BIC suggested

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

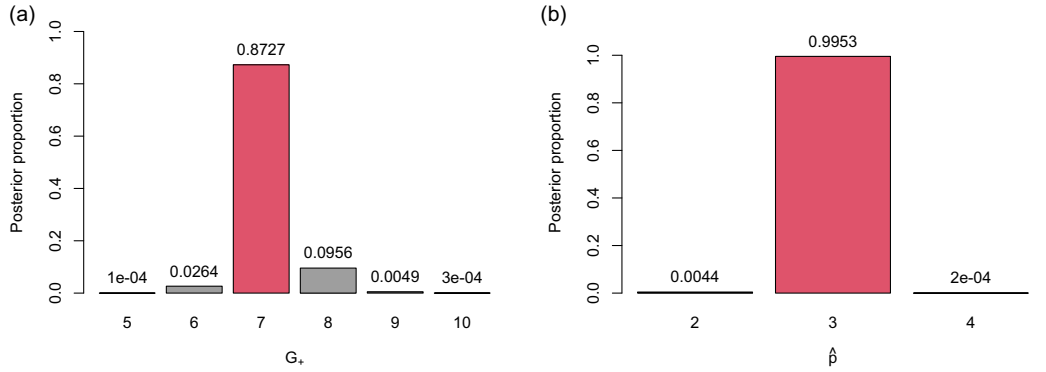

Figure 3. (a) Posterior distribution of the number of active clusters G_+ and (b) the distribution of the effective latent space dimension $\hat{p}$ across 30 moderately sized simulated networks. The true number of clusters and latent dimensions are highlighted in red.

2 dimensions and 3 clusters for all of them, with average ARI and PC values between the LPCM's posterior modal cluster labels and the true labels of 0.87 (sd = 0.12) and 0.96 (sd = 0.04), respectively. In terms of computational cost, fitting the LSPCM on a computer with an i9-13900H CPU and 32 GB RAM took on average 18.5 minutes. The LPCM with the correct p and G took 5.55 minutes on average to run for the same number of iterations. Across the various simulated networks, the total time taken to fit the 25 LPCMs with different combinations of p and G was comparable to or longer than the time taken to fit the LSPCM. However, notably, when fitting the LPCMs no quantification of the uncertainty in p and G is provided.

4.2 Scenario 2: moderately sized networks

Networks with $n = 200$ were generated with $p^* = 3$ latent dimensions and $\boldsymbol{\delta} = (1, 1.1, 1.05)$. The true number of clusters $G^* = 7$ with $\boldsymbol{\mu} = \{(-5, 0, 0), (-5, 5, 0), (0, -5, 5), (0, 0, -5), (2, 0, 2), (-2, 2, -2), (0, -2, 0)\}$ and $\psi_g = 1$ for $g = 1, \ldots, 7$. The 7 clusters have different degrees of separation across the dimensions. Among the 30 simulated networks, the smallest cluster had 4 nodes while the largest had 54. The parameter $\alpha = 20$, which resulted in networks with density varying between 23% and 31%.

Figure 3a shows that the posterior modal number of clusters is correctly inferred with $G_m = 7$ and Figure 3b suggests the modal effective number of dimensions is also correctly inferred with $\hat{p}_m = 3$.

The ARI values indicate very accurate and robust inference of cluster membership with mean value of 0.941 (sd = 0.024). The average PC value was 0.994 (sd = 0.002). The PC values are calculated using only dimensions up to and including the modal effective number of dimensions; in cases where $\hat{p}_m < p^*$, the higher dimensions are not included in the comparison resulting in lower PC values. Additional visualization of the posterior distributions of the variances, shrinkage strengths, and the parameter α for this setting are available in Appendix D. In scenarios with larger p^*, underestimation of the true number of dimensions often occurred as empirically with only $n = 200$ and $G^* = 7$ the relative variance in later dimensions was intuitively small.

Upon fitting the LPCM with 3 dimensions and 7 clusters the average ARI and PC values between the LPCM's posterior modal cluster labels and the true labels were 0.84 (0.09) and 0.98 (0.01), respectively (standard deviation in brackets). The average time for the completion of one MCMC chain for the LSPCM was 149 minutes. However, fitting the LPCM for multiple combinations of p and G would incur a considerably larger computational cost. In fact, solely for

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

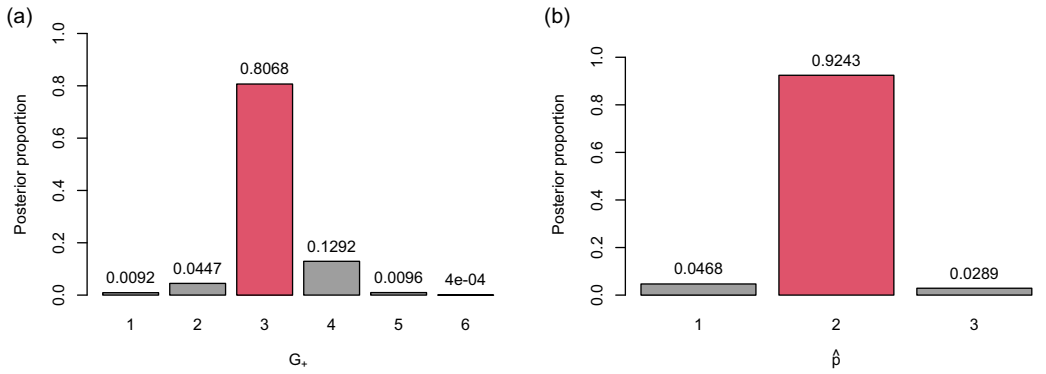

Figure 4. (a) Posterior distribution of the number of active clusters G_+ and (b) the distribution of the effective latent space dimension $\hat{p}$ across 30 small simulated networks with clusters of slightly different volumes. The true number of clusters and dimensions are highlighted in red.

$p = 3$ and $G = 7$ it took 142 minutes to run the MCMC via `latentnet` for the same number of iterations, and again no uncertainty quantification is provided.

4.3 Scenario 3: networks with clusters of different volumes

Networks are generated with the same settings as in Section 4.1 except with clusters of different volumes. Two settings are considered: clusters of slightly different volumes where ψ_g values are set as $(\psi_1, \psi_2, \psi_3) = (\frac{4}{5}, 1, \frac{5}{4})$ and clusters of highly different volumes with $(\psi_1, \psi_2, \psi_3) = (\frac{1}{5}, 1, 5)$. For comparison purposes, the networks simulated in Section 4.1 will be considered as networks with clusters of the same volume as $\psi_g = 1 \forall g$. For the setting with slightly different volumes, network densities varied between 26% and 38% while in the highly different volumes setting, networks had densities between 22% and 38%.

Figure 4 shows that, for clusters of slightly different volumes, the number of clusters and effective dimensions were correctly inferred, but with lower certainty than shown in Figure 2 for networks with clusters of the same volume. For the highly different volumes setting, Figure 5a shows the posterior modal number of clusters $G_m = 4$ which underestimates the true number of clusters, while 5b shows the effective number of dimensions is correctly inferred but again with less certainty than in the setting with clusters of the same volume. For networks with clusters of highly different volumes, intuitively the ARI values indicate poor clustering performance with mean value of 0.617 (sd = 0.208); for the slightly different volumes setting the average ARI was 0.882 (sd = 0.105). Similarly, the average PC values were 0.899 (sd = 0.091) and 0.973 (sd = 0.032) for the highly and slightly different volumes settings, respectively.

Figure 6a highlights the intuitive difficulty in clustering nodes in networks with clusters of highly different volumes. The nodes from the cluster with the largest volume (indicated by black squares in Figure 6a) are far apart in the latent space, and many lie closer to nodes from another cluster e.g., to nodes from the medium volume cluster indicated by red circles. The result, as illustrated in Figure 6b, is that the LPSCM intuitively fits more clusters with smaller volumes than the truth.

5. Application to Twitter network data

The LSPCM is fitted to two binary Twitter networks with different characteristics: a football players network with a small number of nodes, in which each player is known to belong to one of

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

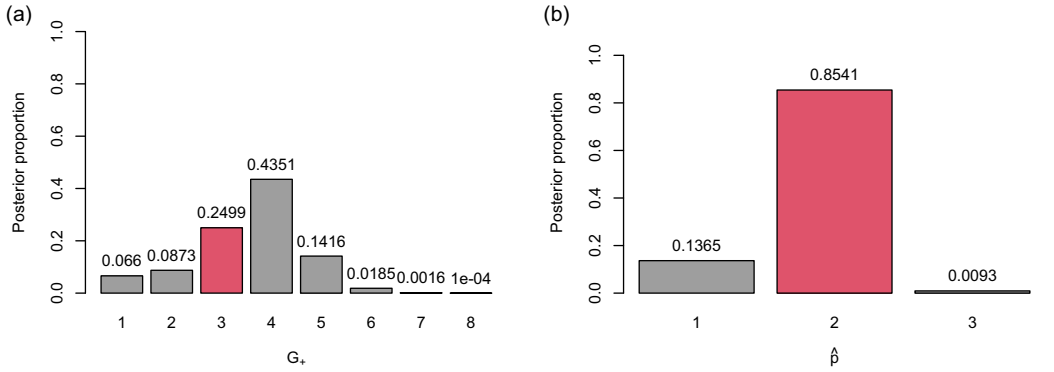

Figure 5. (a) Posterior distribution of the number of active clusters G_+ and (b) the distribution of the effective latent space dimension $\hat{p}$ across 30 small simulated networks with clusters of highly different volumes. The true number of clusters and dimensions are highlighted in red.

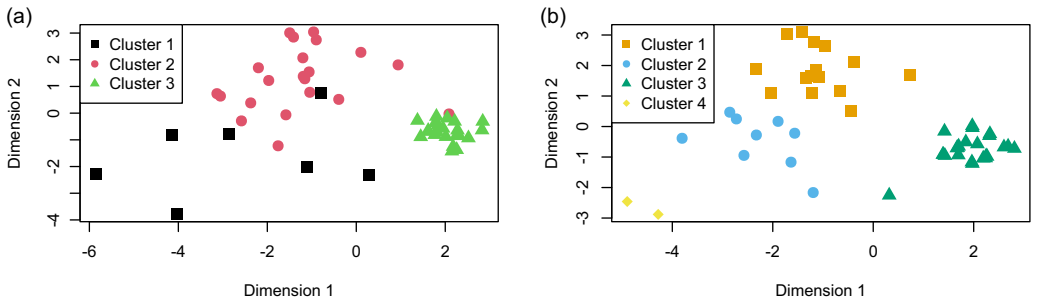

Figure 6. One of the 30 simulated networks when clusters have highly different volumes with (a) the true positions and cluster labels and (b) the LSPCM posterior mean positions conditioned on the modal effective number of dimensions ($\hat{p}_m = 2$) and the estimated cluster labels.

three football clubs, and a network among Irish politicians with a moderate number of nodes, where each politician is affiliated to one of seven Irish political parties. The data are publicly available at this link (Greene and Cunningham, 2013). In the analyses, hyperparameters, initial values, and step sizes are set as in Section 4.

5.1 Football players network

This binary network consists of directed edges indicating the presence of Twitter mentions from one English Premier League football player's Twitter account to another. The data were adapted from those provided in Greene and Cunningham (2013) by considering only the top 3 clubs that have the most player Twitter accounts: 55 players playing for 3 different Premier League clubs are considered with 15 players from Stoke City football club (Stoke), 23 players from Tottenham Hotspur (Spurs), and 17 players from West Bromwich Albion (West Brom). In total, there were 497 mentions between players, giving a network density of 16.73%. To fit the LSPCM to these data, 10 MCMC chains are run, each for 1,000,000 iterations with a burn-in period of 100,000 and thinned every 1,000th. The average time for the completion of one MCMC chain was 30.6 minutes. With 10 MCMC chains, the Gelman–Rubin convergence criterion (Gelman et al., 2014), computed conditional on the modal effective number of dimensions, was satisfied with $1.0 \leq \hat{R} < 1.1$ for the α , ν, and ψ parameters, and with $\hat{R} = 1.3$ and $\hat{R} = 1.2$ for δ_1 and δ_2 respectively. Trace plots are provided in Appendix E.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

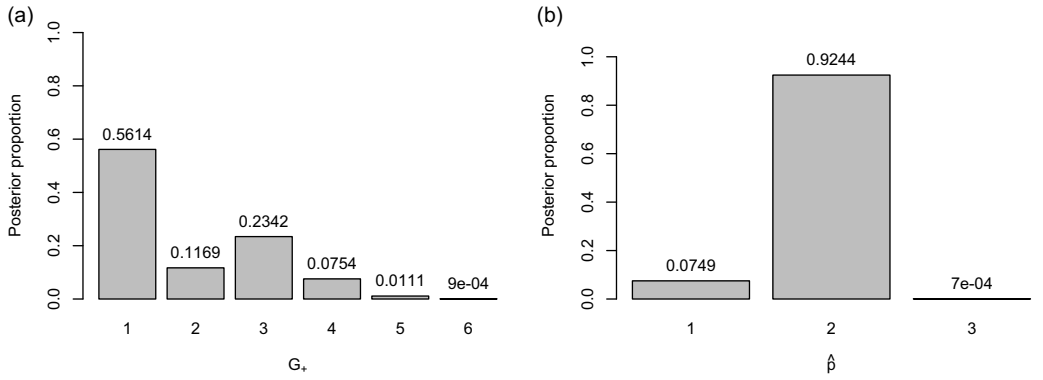

Figure 7. For the football players network, (a) the posterior distribution of the number of non-empty components G_+, and (b) the distribution of the number of effective latent space dimensions $\hat{p}$.

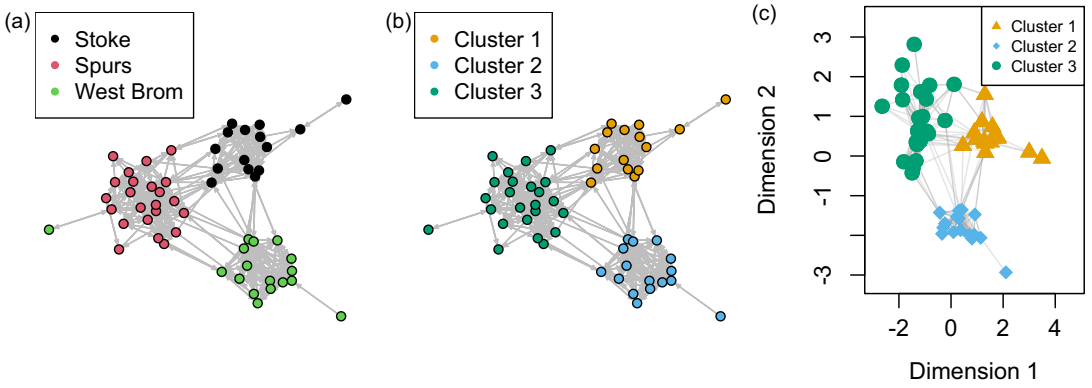

Figure 8. Football players network, (a) the Fruchterman-Reingold layout with players colored by club, (b) the Fruchterman-Reingold layout with players colored by inferred cluster label and (c) posterior mean latent positions on the $\hat{p}_m = 2$ effective latent dimensions colored by inferred cluster label.

Figure 7 illustrates that under the LSPCM the posterior modal number of clusters $G_m = 1$ (1, 4) and the modal effective number of dimensions is $\hat{p}_m = 2$ (1, 2), with 95% empirical intervals in brackets. There is uncertainty in G_+ as demonstrated by the wide posterior interval and notable support for $G_+ = 3$, but there is strong certainty on the number of effective dimensions $\hat{p}$.

Despite $G_m = 1$, maximizing PEAR across the 10 chains resulted in an estimate of 3 clusters, with an ARI of 0.94 between this final clustering and the players' clubs. Figures 8a and 8b show the Fruchterman-Reingold layouts (Kolaczyk and Csárdi, 2020) of the network with the players' clubs and inferred clusters detailed respectively. Only one player was clustered differently to the other players in his club: the West Brom player Romelu Lukaku who was originally from Chelsea football club and was on a season-long loan deal with West Brom. Lukaku mentioned and was mentioned by only one player, the Spurs player Jan Vertonghen and intuitively the LSPCM has clustered them together. Figure 8c shows the posterior mean latent positions colored by the cluster labels that maximizes the PEAR across the 10 chains. Cluster 1 (which contains all of the Stoke players only) and cluster 2 (all players from Spurs, and Romelu Lukaku) are separate from each other on the first dimension, while in the second dimension cluster 3 (which captures West Brom players except Romelu Lukaku) is separated from clusters 1 and 2, indicating both effective latent dimensions are necessary for representation of the nodes in the three clusters.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

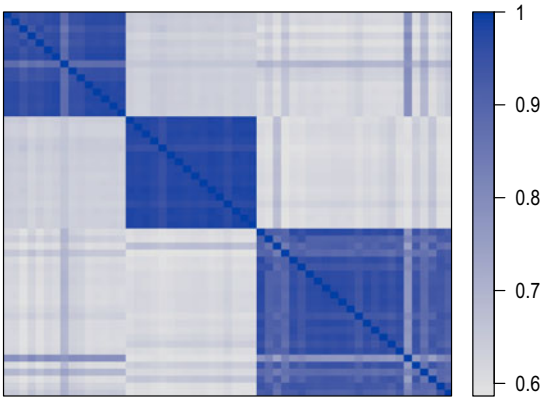

Figure 9. Heat map of the posterior similarity matrix of the cluster labels inferred from the football network, ordered by the cluster labels.

Through the posterior similarity matrix, Figure 9 illustrates the uncertainty in the football players' cluster labels where the color intensity indicates the probability of a player being clustered together with another player. Cluster 2 (predominantly players from West Brom) has the strongest certainty in its cluster labels, while there is some uncertainty in the membership of players in clusters 1 and 3.

Upon fitting 16 LPCMs from all combinations of $p = \{1, \ldots, 4\}$ and $G = \{1, \ldots, 4\}$, the BIC suggests 3 clusters and 2 dimensions as optimal. The ARI between the cluster labels of the optimal LPCM and LSPCM is 1, indicating that both approaches provide the same cluster allocations. The average Procrustes correlation between the posterior mean latent positions under the optimal LPCM and the LSPCM was 0.973 (standard deviation of 0.005) across the 10 LSPCM chains. Additional results regarding the posterior distributions of variance and shrinkage strength parameters can be found in the Appendix E.

5.2 Irish politicians network

The LSPCM is used to analyze a Twitter network between 348 Irish politicians from 2012. The network consists of binary directed edges indicating if one politician follows another (Greene and Cunningham, 2013). Each of the politicians is affiliated with one of the seven Irish political parties: 49 are affiliated with Fianna Fáil (FF), 143 with Fine Gael (FG), 7 with the Green Party (Green), 79 with the Labour Party (Lab), 31 with Sinn Féin (SF), 8 with the United Left Alliance (ULA) and 31 are Independent (Ind). There are 16,856 directed relationships between the politicians, giving a network density of 13.96. For inference, 10 MCMC chains are run, each for 2,000,000 iterations with a burn-in period of 100,000 and thinned every 4,500th sample. Here, the setting of $G = 20$ and $b_\nu = 5$ resulted in very many sparsely populated clusters which provided little clustering insight. Motivated by the context, $G = 10$ and $b_\nu = 1000$ were instead used resulting in fewer, more populated clusters the results of which are presented here; results under $G = 20$ and $b_\nu = 5$, and $G = 10$ and $b_\nu = 100$ are provided in Appendix E for the interested reader. The average time for the completion of one MCMC chain was 11.71 hours. With 10 MCMC chains, the Gelman–Rubin convergence criterion (Gelman et al., 2014), computed conditional on the modal effective number of dimensions, was satisfied with $1.0 \leq \hat{R} < 1.1$ for the α, ν, and ψ_g parameters, and $1.0 \leq \hat{R} < 1.6$ for $\boldsymbol{\delta}$ parameters. Trace plots are provided in Appendix E.

Figure 10 shows that, under the LSCPM, the posterior modal number of clusters is 6 (5, 7) and the modal number of effective dimensions is 4 (3, 4), with the 95% empirical intervals reported in brackets. Upon fitting multiple LPCMs across all combinations of $p = \{2, \ldots, 8\}$ and

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

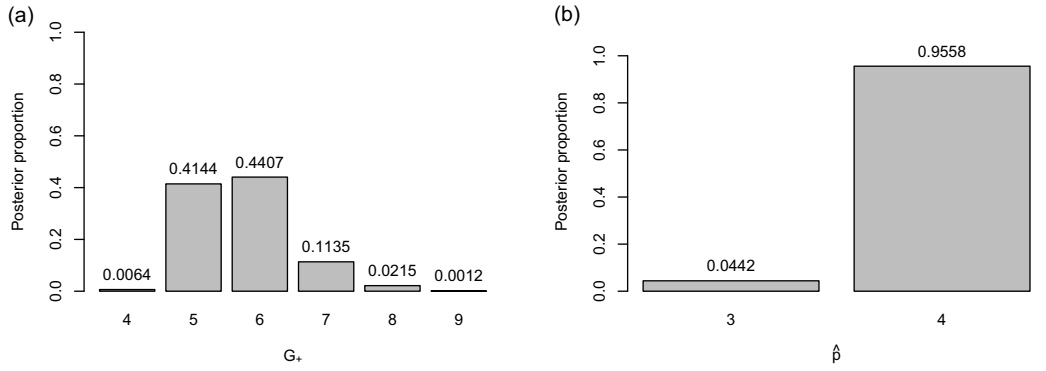

Figure 10. For the Irish politicians network: (a) the posterior distribution of the number of non-empty components G_+, and (b) the distribution of the effective latent space dimension $\hat{p}$.

$G = \{2, \ldots, 8\}$, the BIC suggests 6 clusters and 6 dimensions. The ARI between the cluster labels of the optimal LPCM and LSPCM is 0.55, indicating different clustering solutions under the 2 approaches. The mean Procrustes correlation between the posterior mean positions under the LPCM and LSPCM was 0.71 (standard deviation of 0.01) across the 10 LSPCM chains. Additional results regarding the posterior distributions of variance and shrinkage strength parameters can be found in Appendix E.

Despite $G_m = 6$, maximizing the PEAR across the 10 chains resulted in 5 clusters and the ARI between the inferred cluster labels and the politicians' political affiliations is 0.89. Table 1 presents the cross-tabulation of the political party membership and the LSPCM cluster labels. Additionally, Figure 11 shows the Fruchterman-Reingold layout of the network with nodes colored by political party affiliation and LSPCM inferred cluster label. The inferred clustering structure captures the composition and nature of the 2012 Irish political landscape: a coalition government of Fine Gael and the Labour Party were in power with Fianna Fáil the main opposition party. Intuitively, cluster 3 captures the large number of Fine Gael politicians with cluster 1 capturing the Labour politicians in power with them at the time. Cluster 5 captures all of the opposition's Fianna Fáil politicians. Cluster 2 solely contains the Sinn Féin candidates, with cluster 4 capturing many independent candidates and those from the United Left Alliance. Through the posterior similarity matrix, Figure 12 shows that politicians in cluster 2 (Sinn Féin) have the strongest certainty in their cluster labels while other politicians have some uncertainty.

Figure 13 shows the politicians' posterior mean latent positions on the $\hat{p}_m = 4$ dimensions with nodes colored by the maximized PEAR estimated cluster allocation across the 10 chains. On each dimension, one cluster tends to be located separately to the others, indicating that all the dimensions provide relevant information to distinguish the clusters. For example, on dimension 1 the politicians in cluster 3 are located separately from the other politicians. Similar patterns are apparent for politicians in cluster 5 on dimension 2, politicians in cluster 2 on dimension 3 and politicians in cluster 4 on dimension 4. Visually, the combination of all four dimensions allows for separation of the five clusters.

6. Discussion

The LSPCM enables simultaneous inference on the number of clusters of nodes in a network and the effective dimension of the latent space required to represent it. This is achieved, respectively, through employing a sparse prior on the weights of a mixture model and a Bayesian nonparametric shrinkage prior on the variance of the nodes' positions in the latent space. The LSPCM

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

Table 1. Cross-tabulation of political party membership and the LSPCM representative cluster labels

		Cluster				
		1	*2*	*3*	*4*	*5*
	Fine Gael	3		140		
	Fianna Fáil					49
Political	*Labour Party*	79				
Parties	*Independent*	9	1	1	20	
	Green Party	3			3	1
	Sinn Féin		31			
	United Left Alliance				8	

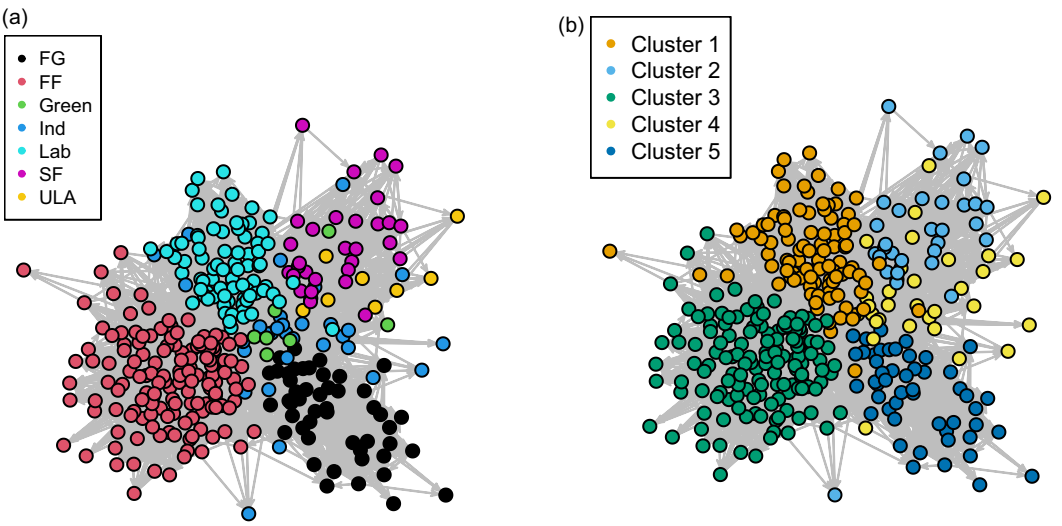

Figure 11. Fruchterman-Reingold layout of the Irish politicians network with nodes colored by (a) political party affiliation and (b) LSPCM inferred cluster labels.

eliminates the need for the computationally expensive procedure of fitting multiple latent position cluster models with different numbers of clusters and dimensions and then choosing the best model using a range of model selection criteria. The performance of the LSPCM was assessed through simulation studies and its application to sporting and political social networks uncovered interesting and intuitive clustering structures.

While the LSPCM demonstrated good performance, it is important to consider its sensitivity to specification of the parameters and threshold parameters employed. As the focus here is on the clustering solution, it is particularly important to consider sensitivity to the parameters of the gamma prior on the Dirichlet parameter, especially in small networks, but the context of the network at hand can provide some guidance. Additionally, sensitivity to the latent position dimension threshold settings should be considered carefully by practitioners as underestimating the dimension could result in invalid inference while overestimating the effective number of dimensions could result in bias. In practice, the values required to recover the effective dimension may change according to the size and density of the network. Using a prior may enable more

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

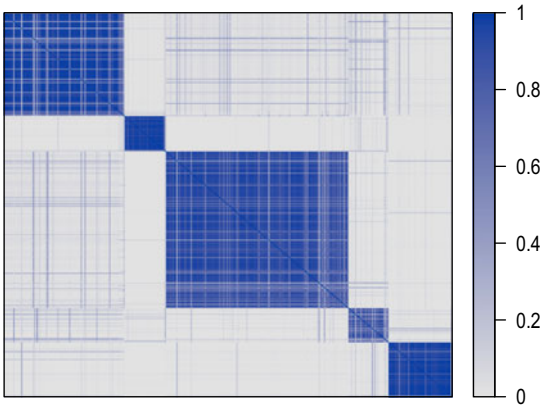

Figure 12. Heat map of the posterior similarity matrix of the Irish politicians ordered by the cluster labels.

robust inference on this parameter. Regarding identifiability, as in the case of LPCM, the interplay between the number of dimensions and the number of clusters is of interest. Here, due to the shrinkage of the variance of the latent dimensions induced by the prior, higher dimensions are less likely to exhibit clustering structure as the cluster means become closer in higher dimensions. Thus, any clustering structure should be captured in the earlier dimensions. Moreover, while the influence of the parameters of the Dirichlet prior and the MTGP shrinkage prior are intuitively related as they are specified within the same model, there is perhaps surprisingly little interaction between the dimension of the latent space and the number of clusters, as observed in Handcock et al. (2007). Relatedly, assessing model fit in a Bayesian clustering setting for network data is challenging; efforts to do so via posterior predictive checks (Gwee, et al., 2024a) and cross validation (Sosa and Rodríguez, 2021) have received attention, albeit in a non-clustering setting.

Fitting the LSPCM is computationally feasible on networks of the scale considered here; however, it would be computationally burdensome to apply it to larger networks. As the number of nodes increases, their computational cost increases quadratically (Salter-Townshend and Murphy, 2013; Rastelli, et al., 2024), making the LSPCM challenging to scale especially with the use of Metropolis-within-Gibbs sampling. Spencer, et al. (2022) propose faster inference for the LPM by combining Hamiltonian Monte Carlo and firefly Monte Carlo, which could be adapted for the LSPCM. Further, while adapting the LSPCM to facilitate modeling of networks with more complex edge types is a natural extension, such advances would come with additional computational cost. Addressing these issues could be possible by employing case–control approaches for the likelihood function (Raftery et al., 2012) and/or avoiding MCMC through the use of variational inference methods (Salter-Townshend and Murphy, 2013; Gwee, et al., 2024b). Many variational algorithms have shown empirical success, and recent theoretical contributions by Yang, et al. (2020) and Liu and Chen (2022) have provided better understanding into the statistical guarantees of these approaches.

Finally, while here a MTGP shrinkage prior was employed on the latent dimensions' variances and an overfitted mixture was used to infer the number of clusters, the LSPCM could be viewed as a member of a broader family of such models given the variety of alternative shrinkage priors and clustering approaches available. Grushanina (2023) provides a broad review of approaches to infinite factorisations. For example, the Indian buffet process (IBP) has been employed to penalize increasing dimensionality in latent factor models; while it allows for the contribution from a dimension to be exactly zero (Knowles and Ghahramani, 2011; Rocková and George, 2016), the sparsity the IBP enforces could be too restrictive here. The spike-and-slab approach of the cumulative shrinkage prior of Legramanti, et al. (2020) is an alternative, likely fruitful approach.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

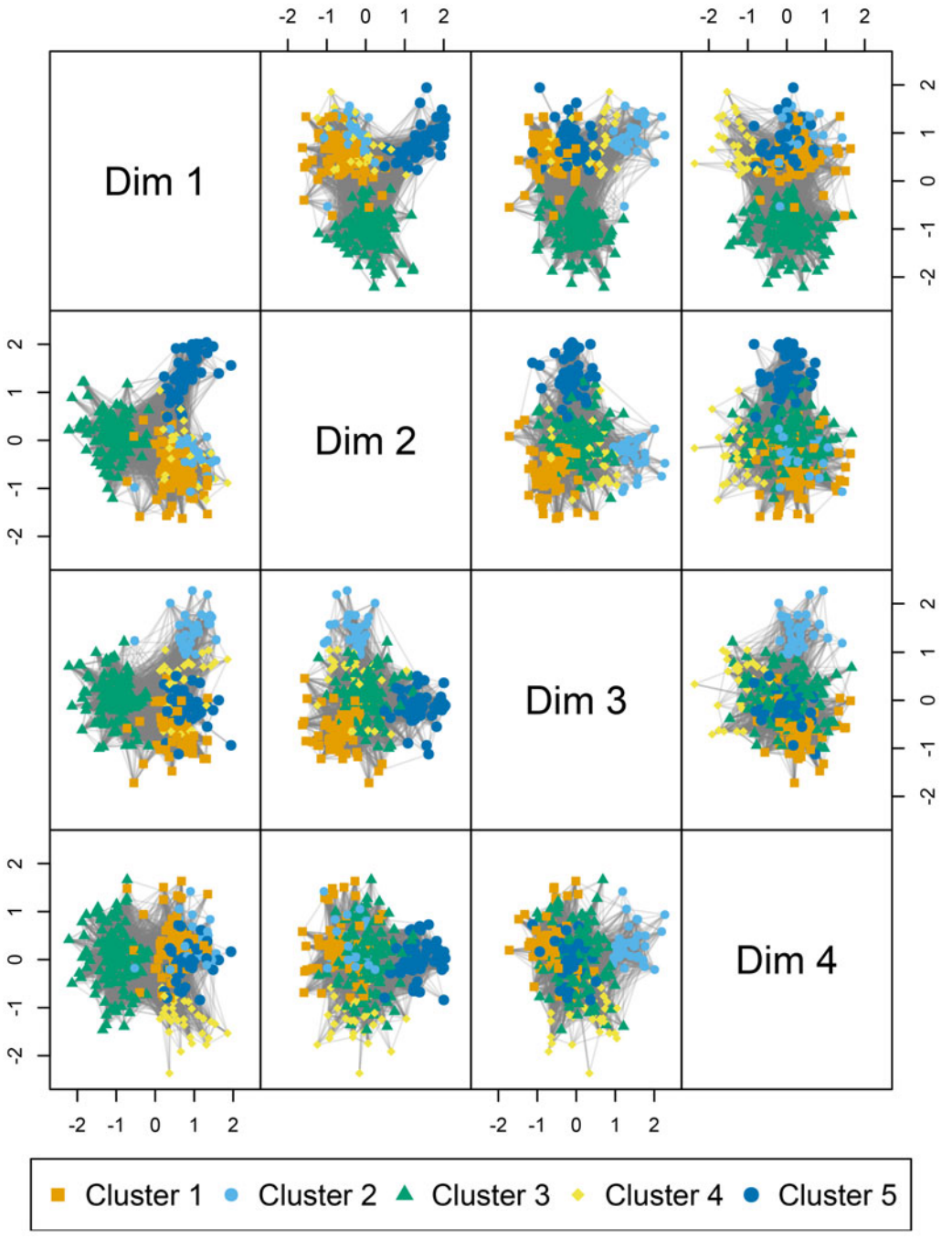

Figure 13. LSPCM inferred posterior mean latent positions of the Irish politicians on the $\hat{p}_m = 4$ dimensions with nodes colored by cluster membership.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

From the clustering point of view, an infinite mixture model could be used to infer the number of clusters. Frühwirth-Schnatter and Malsiner-Walli (2018) discuss linkages between infinite and overfitted mixtures, where they highlight overfitted and infinite mixtures yield comparable clustering performance when the hyperpriors are matched.

Acknowledgements. The authors are grateful for discussions with members of the Working Group in Model-based Clustering and for the reviewers' suggestions which greatly contributed to this work.

Competing interests. None.

Funding statement. This publication has emanated from research conducted with the financial support of Research Ireland under grant number 18/CRT/6049. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Data and code availability. R code implementing the LPSCM and with which all results presented in the manuscript were produced is freely available from the lspm GitLab repository. The two Twitter networks examined in the main manuscript are publicly available at this link.

References

Aliverti, E., & Durante, D. (2019). Spatial modeling of brain connectivity data via latent distance models with nodes clustering. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *12*(3), 185–196. doi: 10.1002/sam.11412.

Bhattacharya, A., & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, *98*(2), 291–306. doi: 10.1093/biomet/asr013.

Cox, T. F., & Cox, M. A. A. (1994). *Multidimensional scaling*. Chapman & Hall.

D'Angelo, S., Alfò, M., & Fop, M. (2023). Model-based clustering for multidimensional social networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(3), 481–507. doi: 10.1093/jrsssa/qnac011.

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. *The Annals of Applied Statistics*, *13*. doi: 10.1214/18-aoas1221.

Durante, D. (2017). A note on the multiplicative gamma process. *Statistics & Probability Letters*, *122*, 198–204. doi: 10.1016/j.spl.2016.11.014.

Durante, D., & Dunson, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, *101*(4), 883–898. doi: 10.1093/biomet/asu040.

Erdos, P., & Rényi, A. (1959). *On random graphs. i.* Vol. *6*, (pp. 290–297). Publicationes Mathematicae.

Fritsch, A. (2022). mcclust: Process an MCMC Sample of Clusterings, doi: 10.32614/CRAN.package.mcclust, R package version 1.0.1, https://CRAN.R-project.org/package=mcclust.

Fritsch, A., & Ickstadt, K. (2009). Improved criteria for clustering based on the posterior similarity matrix. *Bayesian Analysis*, *4*(2), 367–391. doi: 10.1214/09-ba414.

Fruhwirth-Schnatter, S., Celeux, G., & Robert, C. P. (2019). *Handbook of mixture analysis*. CRC Press.

Frühwirth-Schnatter, S., & Malsiner-Walli, G. (2018). From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification*, *13*(1), 33–64. doi: 10.1007/s11634-018-0329-y.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/Crc.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144. doi: 10.1214/aoms/1177706098. https://projecteuclid.org/euclid.aoms/1177706098

Gollini, I., & Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, *25*(1), 246–265. doi: 10.1080/10618600.2014.978006.

Gormley, I. C., & Murphy, T. B. (2010). A mixture of experts latent position cluster model for social network data. *Statistical Methodology*, *7*(3), 385–405. doi: 10.1016/j.stamet.2010.01.002.

Greene, D., & Cunningham, P. (2013). Producing a unified graph representation from multiple social network views. In *Proceedings of the 5th annual acm web science conference, WebSci '13*, pp. 118–121. Paris, France: Association for Computing Machinery. 10.1145/2464464.2464471. isbn: 9781450318891.

Grushanina, M. (2023). A review of Bayesian methods for infinite factorisations. arXiv preprint arXiv:2309.12990.

Gwee, X. Y., Gormley, I. C., & Fop, M. (2024a). A latent shrinkage position model for binary and count network data. *Bayesian Analysis*, *20*(2), 405–433. doi: 10.1214/23-BA1403.

Gwee, X. Y., Gormley, I. C., & Fop, M. (2024b). Variational inference for the latent shrinkage position model. *Stat*, *13*(2), e685. doi: 10.1002/sta4.685.

Handcock, M. S., Raftery, A. E., & 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. doi: 10.1111/j.1467-985x.2007.00471.x.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. doi: 10.1198/016214502388618906.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks*, *5*(2), 109–137. doi: 10.1016/0378-8733(83)90021-7.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. doi: 10.1007/BF01908075.

Jo, W., Chang, D., You, M., & Ghim, G.-H. (2021). A social network analysis of the spread of COVID-19 in South Korea and policy implications. *Scientific Reports*, *11*, 8581. doi: 10.1038/s41598-021-87837-0.

Kaur, H., Rastelli, R., Friel, N., & Raftery, A. (2024). Latent position network models. In: The Sage Handbook of Social Network Analysis. John McLevey, John Scott and Peter J. Carrington Editors, 2nd edn, Sage Publications Ltd.

Knowles, D., & Ghahramani, Z. (2011). Nonparametric bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics*, *5*(2B), 1534–1552. Accessed April 8, 2021. doi: 10.1214/10-aoas435.

Kolaczyk, E. D., & Csárdi, G. (2020). *Statistical analysis of network data with r* (2nd ed.). Springer.

Krivitsky, P. N., & Handcock, M. S. (2008). Fitting latent cluster models for networks with latentnet. *Journal of Statistical Software*, *24*, 1–23. doi: 10.18637/jss.v024.i05.

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. *Social Networks*, *31*(3), 204–213. doi: 10.1016/j.socnet.2009.04.001.

Legramanti, S., Durante, D., & Dunson, D. B. (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika*, *107*(3), 745–752. doi: 10.1093/biomet/asaa008.

Liu, Y., & Chen, Y. (2022). Variational inference for latent space models for dynamic networks. *Statistica Sinica*, *32*, 2147–2170. doi: 10.5705/ss.202020.0506.

Malsiner-Walli, G., Frühwirth-Schnatter, S., & Grün, B. (2014). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing*, *26*(1-2), 303–324. doi: 10.1007/s11222-014-9500-2.

Malsiner-Walli, G., Frühwirth-Schnatter, S., & Grün, B. (2017). Identifying mixtures of mixtures using Bayesian estimation. *Journal of Computational and Graphical Statistics*, *26*(2), 285–295. doi: 10.1080/10618600.2016.1200472.

Murphy, K., Viroli, C., & Gormley, I. C. (2020). Infinite mixtures of infinite factor analysers. *Bayesian Analysis*, *15*, 937–963. doi: 10.1214/19-ba1179.

Ng, T. L. J., Murphy, T. B., Westling, T., McCormick, T. H., & Fosdick, B. (2021). Modeling the social media relationships of irish politicians using a generalized latent space stochastic blockmodel. *The Annals of Applied Statistics*, *15*, 1923–1944. doi: 10.1214/21-aoas1483.

Passino, F. S., & Heard, N. A. (2020). Bayesian estimation of the latent dimension and communities in stochastic blockmodels. *Statistics and Computing*, *30*(5), 1291–1307. doi: 10.1007/s11222-020-09946-6.

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*(2), 169–178. doi: 10.1007/s004420100720.

Pham, H. T. D., & Sewell, D. K. (2024). Automated detection of edge clusters via an overfitted mixture prior. [in en] *Network Science*, *12*(1), 88–106. doi: 10.1017/nws.2023.22 issn: 2050-1242, 2050-1250.

R Core Team (2024). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. *Available at* https://www.R-project.org/.

Raftery, A. E., Niu, X., Hoff, P. D., & 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. doi: 10.1080/10618600.2012.679240.

Rastelli, R., Friel, N., & Raftery, A. E. (2016). Properties of latent variable network models. *Network Science*, *4*(4), 407–432. doi: 10.1017/nws.2016.23.

Rastelli, R., Maire, F., & Friel, N. (2024). Computationally efficient inference for latent position network models, *Electronic Journal of Statistics*, *18*(1), 2531–2570. doi: 10.1214/24-EJS2256.

Rocková, V., & George, E. I. (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association*, *111*(516), 1608–1622. Accessed January 14, 2022. doi: 10.1080/01621459.2015.1100620.

Ryan, C., & JasonWyse, N. F. (2017). Bayesian model selection for the latent position cluster model for social networks. *Network Science*, *5*(1), 70–91. doi: 10.1017/nws.2017.6.

Salter-Townshend, M., & Murphy, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, *57*(1), 661–671. doi: 10.1016/j.csda.2012.08.004.

Scrucca, L., Fop, M., Brendan Murphy, T., & Adrian, E. R. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*, 289–317. doi: 10.32614/rj-2016-021 Accessed May 4, 2020 https://journal.r-project.org/archive/2016/RJ-2016-021/RJ-2016-021.pdf.

Sewell, D. K. (2020). Model-based edge clustering. *Journal of Computational and Graphical Statistics*, *30*(2), 390–405. doi: 10.1080/10618600.2020.1811104.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

Sewell, D. K., & Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, *12*(2), 351–377. doi: 10.1214/16-ba1000.
Snijders, T. A. B., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, *14*(1), 75–100. doi: 10.1007/s003579900004.
Sosa, J., & Betancourt, B. (2022). A latent space model for multilayer network data. *Computational Statistics & Data Analysis*, *169*, 107432. doi: 10.1016/j.csda.2022.107432.
Sosa, J., & Buitrago, L. (2021). A review of latent space models for social networks. *Revista Colombiana de Estadística*, *44*(1), 171–200. doi: 10.15446/rce.v44n1.89369.
Sosa, J., & Rodríguez, A. (2021). A latent space model for cognitive social structures data. *Social Networks*, *65*, 85–97. doi: 10.1016/j.socnet.2020.12.002.
Spencer, N. A., Junker, B. W., & Sweet, T. M. (2022). Faster MCMC for Gaussian latent position network models. *Network Science*, *10*(1), 20–45. doi: 10.1017/nws.2022.1. Accessed May 23, 2022.
Yang, C., Priebe, C. E., Park, Y., & Marchette, D. J. (2020). Simultaneous dimensionality and complexity model selection for spectral graph clustering. *Journal of Computational and Graphical Statistics*, *30*(2), 422–441. doi: 10.1080/10618600.2020.1824870.
Yang, Y., Pati, D., & Bhattacharya, A. (2020). α-variational inference with statistical guarantees. *The Annals of Statistics*, *48*, 886–905. doi: 10.1214/19-aos1827. Accessed April 27, 2022.

A. Hyperparameter specifications

Table 2. Hyperparameter specifications for the LSPCM model

Parameter	Hyperparameters	Values	Comments
p_0		Generally 5	
G		Generally 20	
α	$(\mu_\alpha, \sigma_\alpha^2)$	(0, 4)	
ν	(a_ν, Gb_ν)	Generally Gamma $(a_\nu = 5, Gb_\nu = G5)$	Increasing b_ν increases the likelihood of a smaller number of clusters.
ψ_g	(a_ψ, b_ψ)	(400,400)	Require expected value near 1 with low variance to avoid very variable clusters.
δ_1	(a_1, b_1)	(2,1)	Based on the work from Durante (2017); Gwee, et al. (2024a).
δ_h	(a_2, b_2, t_2)	(3,1,1)	Based on the work from Durante (2017); Gwee, et al. (2024a).
	ξ	9	Based on Ryan et al. (2017).
	κ_0	Generally $\kappa_0 = 4$	Increasing κ_0 decreases adaptation.
	κ_1	Generally $\kappa_1 = 5 \times 10^{-4}$	Increasing κ_1 decreases adaptation.
	ϵ_1	Generally $\epsilon_1 = 0.8$	Increasing ϵ_1 makes it harder to decrease the number of dimensions.
	ϵ_2	Generally $\epsilon_2 = 0.9$	Increasing ϵ_2 (up to $\epsilon_2 = 1$) makes it harder to increase the number of dimensions.
	ϵ_3	Generally $\epsilon_3 = 5$	Increasing ϵ_3 makes it harder to increase the number of dimensions.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

B. Notation and terminology

n: Number of nodes.
p: The truncation dimension in the fitted LSPM.
$\hat{p}$: The number of effective dimensions.
p^*: The true effective dimension of the latent space.
p_0: The initial truncation level of the number of dimensions.
$\hat{p}_m$: The modal number of effective dimensions.
G: The number of mixture components.
G_+: The number of non-empty mixture components.
G^*: The true number of clusters.
G_m: The posterior modal number of non-empty components.
$\mathbf{Y}$: $n \times n$ network adjacency matrix.
$y_{i,j}$: Value of an edge between node i and node j.
$\mathbf{Z}$: $n \times p$ matrix of latent positions.
$z_{i\ell}$: The latent position of node i in dimension ℓ.
$q_{i,j}$: Probability of forming an edge between node i and node j.
α: Global parameter that captures the overall connectivity level in the network.
τ_g: The mixing weight of component g.
ψ_g: The cluster g specific variance scaling parameter.
ν: The hyperparameter for τ_g.
c_{ig}: A binary indicator variable of membership of node i in component g.
$\boldsymbol{\mu}$: $G \times p$ matrix of the mean latent positions.
$\boldsymbol{\mu}_g$: The mean latent position parameter for component g.
Ω: $p \times p$ precision matrix of the latent positions.
ω_ℓ: Precision/global shrinkage parameter for dimension ℓ.
δ_h: Shrinkage strength from dimension h.
ξ: Scaling factor in the prior covariance of the component means.
$\boldsymbol{\Theta}$: A collective term for $\boldsymbol{\tau}$, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$.
a_1: Shape parameter of gamma distribution for δ_1 in dimension 1.
b_1: Rate parameter of the gamma distribution for δ_1 in dimension 1.
a_2: Shape parameter of gamma distribution for δ_h in dimension h.
b_2: Rate parameter of the gamma distribution for δ_h in dimension h.
t_2: Left truncation point of the gamma distribution for δ_h.
a_ν: Shape parameter of gamma distribution for ν.
b_ν: Rate parameter of the gamma distribution for ν.
a_ψ: Shape parameter of gamma distribution for ψ.
b_ψ: Rate parameter of the gamma distribution for ψ.
κ_0: The parameter that affects the horizontal translation of the dimension adaptation probability exponential function.
κ_1: The parameter that affects the horizontal scaling of the dimension adaptation probability exponential function.
ϵ_1: The threshold for decreasing dimensions in the dimension adaptation step.
ϵ_2: The threshold for adding a dimension in the dimension adaptation step.
ϵ_3: The threshold for adding a dimension in the dimension adaptation step when the truncation dimension is 1.

C. Derivation of full conditional distributions

Links between nodes i and j are assumed to form probabilistically from a Bernoulli distribution:

$$y_{i,j} \sim \text{Bin}(q_{i,j})$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

For a binary network, the probability $q_{i,j}$ is expressed in terms of a logistic model i.e.

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

Denoting $\eta_{i,j} = \alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2$, then $q_{i,j} = \frac{\exp(\eta_{i,j})}{1+\exp(\eta_{i,j})}$.
The likelihood function of the LSPCM is

$$\begin{aligned} L(\mathbf{Y}|\mathbf{Z},\alpha) &= \prod_{i\neq j} P(y_{i,j}|\mathbf{z}_i,\mathbf{z}_j,\alpha) \\ &= \prod_{i\neq j} \left[\frac{\exp(\eta_{i,j})}{1+\exp(\eta_{i,j})}\right]^{y_{i,j}} \left[1-\frac{\exp(\eta_{i,j})}{1+\exp(\eta_{i,j})}\right]^{1-y_{i,j}} \\ &= \prod_{i\neq j} \frac{\exp(\eta_{i,j}y_{i,j})}{1+\exp(\eta_{i,j})}, \end{aligned}$$

The joint posterior distribution of the LSPM is

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

Then

$$\begin{aligned} \mathbb{P}(\alpha,\mathbf{Z},\mathbf{C},\boldsymbol{\tau},\boldsymbol{\Theta}\mid\mathbf{Y}) \propto & \left\{\prod_{i\neq j}\left[\frac{\exp(\eta_{i,j}y_{i,j})}{1+\exp(\eta_{i,j})}\right]\right\} \\ & \times \left\{\frac{1}{\sqrt{2\pi\sigma_\alpha^2}}\exp\left[-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\right]\right\} \\ & \times \prod_{i=1}^{n}\prod_{g=1}^{G}\left\{\left(\frac{1}{2\pi}\right)^{\frac{p}{2}}[\det(\psi_g\Omega)]^{\frac{1}{2}}\right. \\ & \qquad \left.\exp\left[-\frac{1}{2}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)\right]\right\}^{c_{ig}} \\ & \times \prod_{i=1}^{n}\left[\frac{n!}{c_{i1}!\cdots c_{iG}!}\tau_1^{c_{i1}}\cdots\tau_G^{c_{iG}}\right] \\ & \times \left[\frac{1}{\beta(\nu)}\prod_{g=1}^{G}\tau_g^{\nu-1}\right] \\ & \times \left\{\prod_{g=1}^{G}\left(\frac{1}{2\pi}\right)^{\frac{p}{2}}[\det(\xi^{-1}\Omega)]^{-\frac{1}{2}}\right. \\ & \qquad \left.\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_g-\mathbf{0})^T(\xi^{-1}\Omega)(\boldsymbol{\mu}_g-\mathbf{0})\right]\right\} \\ & \times \left\{\frac{b_\psi^{a_\psi}}{\Gamma(a_\psi)}(\psi_g)^{a_\psi-1}\exp\left[-b_\psi(\psi_g)\right]\right\} \\ & \times \left\{\frac{b_\nu^{a_\nu}}{\Gamma(a_\nu)}(\nu)^{a_\nu-1}\exp\left[-b_\nu(\nu)\right]\right\} \end{aligned}$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

$$\times \left\{ \frac{b_1^{a_1}}{\Gamma(a_1)} (\delta_1)^{a_1 - 1} \exp\left[-b_1(\delta_1) \right] \right\}$$
$$\times \left\{ \prod_{h=2}^{p} \frac{b_2^{a_2}}{\Gamma(a_2)} (\delta_h)^{a_2 - 1} \exp\left[-b_2(\delta_h) \right] \right\}$$

Indicating with $-$ the conditioning on all the remaining variables, the full conditional distribution for α is:

$$\mathbb{P}(\alpha \mid -) \propto \left\{ \prod_{i \neq j} \left[\frac{\exp(\eta_{i,j} y_{i,j})}{1 + \exp(\eta_{i,j})} \right] \right\} \times \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left[-\frac{1}{2} \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} \right]$$
$$\propto \left\{ \prod_{i \neq j} \left[\frac{\exp(\eta_{i,j} y_{i,j})}{1 + \exp(\eta_{i,j})} \right] \right\} \times \exp\left[-\frac{1}{2} \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2} \right]$$
$$\log \mathbb{P}(\alpha \mid -) \propto \sum_{i \neq j} \left\{ \eta_{i,j} y_{i,j} - \log\left[1 + \exp(\eta_{i,j}) \right] \right\} - \frac{1}{2} \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2}$$
$$\propto \sum_{i \neq j} \left\{ \eta_{i,j} y_{i,j} - \log\left[1 + \exp(\eta_{i,j}) \right] \right\} - (\alpha - \mu_\alpha)^2$$
$$\propto \sum_{i \neq j} \left\{ (\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) y_{i,j} - \log\left[1 + \exp(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) \right] \right\} - (\alpha - \mu_\alpha)^2$$
$$\log \mathbb{P}(\alpha \mid -) \propto \sum_{i \neq j} \left\{ \alpha y_{i,j} - \log\left[1 + \exp(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) \right] \right\} - (\alpha - \mu_\alpha)^2$$

As this is not a recognizable distribution, the Metropolis-Hasting algorithm is employed.

The full conditional distribution for $\mathbf{Z}$ is:

$$\mathbb{P}(\mathbf{Z} \mid -) \propto \left\{ \prod_{i \neq j} \left[\frac{\exp(\eta_{i,j} y_{i,j})}{1 + \exp(\eta_{i,j})} \right] \right\}$$
$$\times \prod_{i=1}^{n} \prod_{g=1}^{G} \left\{ \left(\frac{1}{2\pi} \right)^{\frac{p}{2}} \left[\det(\psi_g \Omega) \right]^{\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g \Omega) (\mathbf{z}_i - \boldsymbol{\mu}_g) \right] \right\}^{c_{ig}}$$
$$\log \mathbb{P}(\mathbf{Z} \mid -) \propto \sum_{i \neq j} \left\{ \eta_{i,j} y_{i,j} - \log\left[1 + \exp(\eta_{i,j}) \right] \right\} - \sum_{i=1}^{n} \sum_{g=1}^{G} \left[\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g \Omega) (\mathbf{z}_i - \boldsymbol{\mu}_g) \right]^{c_{ig}}$$
$$\propto \sum_{i \neq j} \left\{ \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 y_{i,j} - \log\left[1 + \exp(\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2^2) \right] \right\}$$
$$- \sum_{i=1}^{n} \sum_{g=1}^{G} \left[\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g \Omega) (\mathbf{z}_i - \boldsymbol{\mu}_g) \right]^{c_{ig}}$$

As this is not a recognizable distribution, the Metropolis-Hasting algorithm is employed.

The full conditional distribution for $c_{ig} = 1$ is:

$$\mathbb{P}(c_{ig} = 1 \mid -) = \frac{\mathbb{P}(- \mid c_{ig} = 1) \mathbb{P}(c_{ig} = 1)}{\mathbb{P}(-)} = \frac{\mathbb{P}(\mathbf{z}_i \mid c_{ig} = 1) \mathbb{P}(c_{ig} = 1)}{\sum_{r=1}^{G} \mathbb{P}(\mathbf{z}_{ir} \mid c_{ir} = 1) \mathbb{P}(c_{ir} = 1)}$$
$$= \frac{\mathrm{MVN}_p(\mathbf{z}_i; \boldsymbol{\mu}_g, \psi_g^{-1} \boldsymbol{\Omega}^{-1}) \tau_g}{\sum_{r=1}^{G} \mathrm{MVN}_p(\mathbf{z}_{ir}; \boldsymbol{\mu}_r, \psi_g^{-1} \boldsymbol{\Omega}^{-1}) \tau_r} = \frac{\tau_g \mathrm{MVN}_p(\mathbf{z}_i; \boldsymbol{\mu}_g, \psi_g^{-1} \boldsymbol{\Omega}^{-1})}{\sum_{r=1}^{G} \tau_r \mathrm{MVN}_p(\mathbf{z}_{ir}; \boldsymbol{\mu}_r, \psi_g^{-1} \boldsymbol{\Omega}^{-1})}$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

The full conditional distribution for τ_g is:

$$\begin{aligned}\mathbb{P}(\tau_g \mid -) &\propto \left[\frac{1}{\beta(\nu)}\tau_g^{\nu-1}\right]\left[\prod_{i=1}^{n}\tau_g^{c_{ig}}\right] \\ &\propto \tau_g^{\nu-1}\tau_g^{\sum_{i=1}^{n} c_{ig}} \\ &\propto \tau_g^{\sum_{i=1}^{n} c_{ig}+\nu-1} \\ &\sim \text{Dir}(\sum_{i=1}^{n} c_{ig}+\nu)\end{aligned}$$

The full conditional distribution for ν is:

$$\begin{aligned}\mathbb{P}(\nu \mid -) &\propto \left[\frac{1}{\beta(\nu)}\prod_{g=1}^{G}\tau_g^{\nu-1}\right]\left\{\frac{b_\nu^{a_\nu}}{\Gamma(a_\nu)}(\nu)^{a_\nu-1}\exp\left[-b_\nu(\nu)\right]\right\} \\ &\propto \prod_{g=1}^{G}\tau_g^{\nu-1}\nu^{a_\nu-1}\exp\left[-b_\nu(\nu)\right]\end{aligned}$$

As this is not a recognizable distribution, the Metropolis-Hasting algorithm is employed.

The full conditional distribution for ψ_g is:

$$\begin{aligned}\mathbb{P}(\psi_g \mid -) &\propto \prod_{i=1}^{n}\left\{\left(\frac{1}{2\pi}\right)^{\frac{p}{2}}\left[\det(\psi_g\Omega)\right]^{\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)\right]\right\}^{c_{ig}} \\ &\quad\times\left\{\frac{b_\psi^{a_\psi}}{\Gamma(a_\psi)}(\psi_g)^{a_\psi-1}\exp\left[-b_\psi(\psi_g)\right]\right\} \\ &\propto \psi_g^{\frac{p\sum_{i=1}^{n} c_{ig}}{2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)c_{ig}\right] \\ &\quad\times\left\{\frac{b_\psi^{a_\psi}}{\Gamma(a_\psi)}(\psi_g)^{a_\psi-1}\exp\left[-b_\psi(\psi_g)\right]\right\} \\ &\propto \psi_g^{\frac{p\sum_{i=1}^{n} c_{ig}}{2}}(\psi_g)^{a_\psi-1}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)c_{ig}\right]\exp\left[-b_\psi(\psi_g)\right] \\ &\propto \psi_g^{a_\psi+\frac{p\sum_{i=1}^{n} c_{ig}}{2}-1}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)c_{ig}-b_\psi(\psi_g)\right] \\ &\propto \psi_g^{a_\psi+\frac{p\sum_{i=1}^{n} c_{ig}}{2}-1}\exp\left\{-\psi_g\left[b_\psi+\frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)c_{ig}\right]\right\} \\ &\sim \text{Gam}\left(a_\psi+\frac{p\sum_{i=1}^{n} c_{ig}}{2},\qquad b_\psi+\frac{1}{2}\sum_{i=1}^{n}(\mathbf{z}_i-\boldsymbol{\mu}_g)^T(\Omega)(\mathbf{z}_i-\boldsymbol{\mu}_g)c_{ig}\right)\end{aligned}$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

The full conditional distribution for $\boldsymbol{\mu}_g$ is:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\mu}_g \mid -) &\propto \prod_{i=1}^{n} \left\{ \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} [\det(\psi_g \Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{z}_i - \boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i - \boldsymbol{\mu}_g)\right] \right\}^{c_{ig}} \\
&\quad \times \left\{ \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} [\det(\xi^{-1}\Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu}_g - \mathbf{0})^T(\xi^{-1}\boldsymbol{\Omega})(\boldsymbol{\mu}_g - \mathbf{0})\right] \right\} \\
&\propto \exp\left\{ \sum_{i=1}^{n} \left[-\frac{1}{2}(\mathbf{z}_i - \boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i - \boldsymbol{\mu}_g)(c_{ig})\right] - \frac{\boldsymbol{\mu}_g^T(\xi^{-1}\boldsymbol{\Omega})\boldsymbol{\mu}_g}{2} \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left[\sum_{i=1}^{n} \left[(\mathbf{z}_i^T\psi_g\boldsymbol{\Omega}\mathbf{z}_i c_{ig} - 2\mathbf{z}_i^T\psi_g\boldsymbol{\Omega}\boldsymbol{\mu}_g c_{ig} + \boldsymbol{\mu}_g^T\psi_g\boldsymbol{\Omega}\boldsymbol{\mu}_g c_{ig}) \right] + \boldsymbol{\mu}_g^T\xi^{-1}\boldsymbol{\Omega}\boldsymbol{\mu}_g \right] \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left[-2\sum_{i=1}^{n}\psi_g c_{ig}\mathbf{z}_i^T\boldsymbol{\Omega}\boldsymbol{\mu}_g + \sum_{i=1}^{n}\psi_g c_{ig}\boldsymbol{\mu}_g^T\boldsymbol{\Omega}\boldsymbol{\mu}_g + \boldsymbol{\mu}_g^T\xi^{-1}\boldsymbol{\Omega}\boldsymbol{\mu}_g \right] \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left[\boldsymbol{\mu}_g^T\boldsymbol{\Omega}\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\boldsymbol{\mu}_g - 2\boldsymbol{\mu}_g^T\left(\psi_g\boldsymbol{\Omega}\sum_{i=1}^{n}(c_{ig}\mathbf{z}_i)\right) \right] \right\} \\
&\propto \exp\left\{ -\frac{1}{2}\left[\boldsymbol{\mu}_g^T - \left\{\psi_g\boldsymbol{\Omega}\sum_{i=1}^{n}(c_{ig}\mathbf{z}_i)\right\}^T \left\{\boldsymbol{\Omega}\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\right\}^{-1} \right] \right. \\
&\quad \left. \times \left[\boldsymbol{\Omega}\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\right] \left[\boldsymbol{\mu}_g - \left\{\boldsymbol{\Omega}^T\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\right\}^{-1} \left\{\psi_g\boldsymbol{\Omega}\sum_{i=1}^{n}(c_{ig}\mathbf{z}_i)\right\} \right] \right\}
\end{aligned}$$

since $\boldsymbol{\Omega}$ is a diagonal matrix, $\boldsymbol{\Omega} = \boldsymbol{\Omega}^T$ and $\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1} = \mathbf{I}$, thus,

$$\begin{aligned}
&\propto \exp\left\{ -\frac{1}{2}\left[\boldsymbol{\mu}_g - \frac{\sum_{i=1}^{n} c_{ig}\mathbf{z}_i}{\sum_{i=1}^{n} c_{ig} + \xi^{-1}} \right]^T \left[\boldsymbol{\Omega}\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\right] \right. \\
&\quad \left. \times \left[\boldsymbol{\mu}_g - \frac{\sum_{i=1}^{n} c_{ig}\mathbf{z}_i}{\sum_{i=1}^{n} c_{ig} + \xi^{-1}} \right] \right\} \\
&\sim \mathrm{MVN}_p\left(\frac{\sum_{i=1}^{n} c_{ig}\mathbf{z}_i}{\sum_{i=1}^{n} c_{ig} + \xi^{-1}}, \quad \left[\boldsymbol{\Omega}\left(\psi_g\sum_{i=1}^{n}c_{ig} + \xi^{-1}\right)\right]^{-1} \right)
\end{aligned}$$

The full conditional distribution for δ_1 is:

$$\begin{aligned}
\mathbb{P}(\delta_1 \mid -) &\propto \prod_{i=1}^{n}\prod_{g=1}^{G} \left\{ \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} [\det(\psi_g \Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{z}_i - \boldsymbol{\mu}_g)^T(\psi_g\Omega)(\mathbf{z}_i - \boldsymbol{\mu}_g)\right] \right\}^{c_{ig}} \\
&\quad \times \prod_{g=1}^{G} \left\{ \left(\frac{1}{2\pi}\right)^{\frac{p}{2}} [\det(\xi^{-1}\Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\mu}_g - \mathbf{0})^T(\xi^{-1}\boldsymbol{\Omega})(\boldsymbol{\mu}_g - \mathbf{0})\right] \right\} \\
&\quad \times \left[\frac{b_1^{a_1}}{\Gamma(a_1)}(\delta_1)^{a_1 - 1}\exp(-b_1\delta_1) \right] \\
&\propto \left\{ \boldsymbol{\omega}^{\frac{p\sum_{i=1}^{n}\sum_{i=g}^{G} c_{ig}}{2}} \mathbf{I}_p \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\sum_{i=g}^{G}(\mathbf{z}_i - \boldsymbol{\mu}_g)^T(\psi_g\boldsymbol{\omega})\mathbf{I}_p(\mathbf{z}_i - \boldsymbol{\mu}_g)c_{ig} \right] \right\}
\end{aligned}$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

$$\times \left\{ \boldsymbol{\omega}^{\frac{Gp}{2}} \mathbf{I}_p \exp\left[-\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}\boldsymbol{\omega}) \mathbf{I}_p \boldsymbol{\mu}_g \right] \right\} \times \left[\delta_1^{a_1-1} \exp(-b_1\delta_1) \right]$$

$$\propto \left\{ \delta_1^{\frac{np}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{i=g}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=1}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right] \right\}$$

$$\times \left\{ \delta_1^{\frac{Gp}{2}} \exp\left[-\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=1}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g \right] \right\} \times \left[\delta_1^{a_1-1} \exp(-b_1\delta_1) \right]$$

$$\propto \delta_1^{\frac{np}{2}} \delta_1^{\frac{Gp}{2}} \delta_1^{a_1-1} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{i=g}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\delta_1 \prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right.$$

$$\left. -\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\delta_1 \prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g - b_1\delta_1 \right]$$

$$\propto \delta_1^{\frac{(n+G)p}{2}+a_1-1} \exp\left\{ -\left[\frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right.\right.$$

$$\left.\left. +\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_1 \right] \delta_1 \right\}$$

$$\sim \text{Gam}\left(\frac{(n+G)\,p}{2} + a_1, \quad \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right.$$

$$\left. +\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=2}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_1 \right)$$

The full conditional distribution for δ_h, where $h \geq 2$ is:

$$\mathbb{P}(\delta_h \mid -) \propto \prod_{i=1}^{n} \prod_{g=1}^{G} \left\{ \left(\frac{1}{2\pi} \right)^{\frac{p}{2}} [\det(\psi_g \Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g \Omega) (\mathbf{z}_i - \boldsymbol{\mu}_g) \right] \right\}^{c_{ig}}$$

$$\times \prod_{g=1}^{G} \left\{ \left(\frac{1}{2\pi} \right)^{\frac{p}{2}} [\det(\xi^{-1}\Omega)]^{\frac{1}{2}} \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_g - \mathbf{0})^T (\xi^{-1}\boldsymbol{\Omega}) (\boldsymbol{\mu}_g - \mathbf{0}) \right] \right\}$$

$$\times \left[\frac{b_2^{a_2}}{\Gamma(a_2)} (\delta_h)^{a_2-1} \exp(-b_2\delta_h) \right]$$

$$\propto \left\{ \boldsymbol{\omega}^{\frac{(p-h+1)\sum_{i=1}^{n}\sum_{g=1}^{G} c_{ig}}{2}} \mathbf{I}_p \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g \boldsymbol{\omega}) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right] \right\}$$

$$\times \left\{ \boldsymbol{\omega}^{\frac{Gp}{2}} \mathbf{I}_p \exp\left[-\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}\boldsymbol{\omega}) \mathbf{I}_p \boldsymbol{\mu}_g \right] \right\}$$

$$\times \left[\delta_h^{a_2-1} \exp(-b_2\delta_h) \right]$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

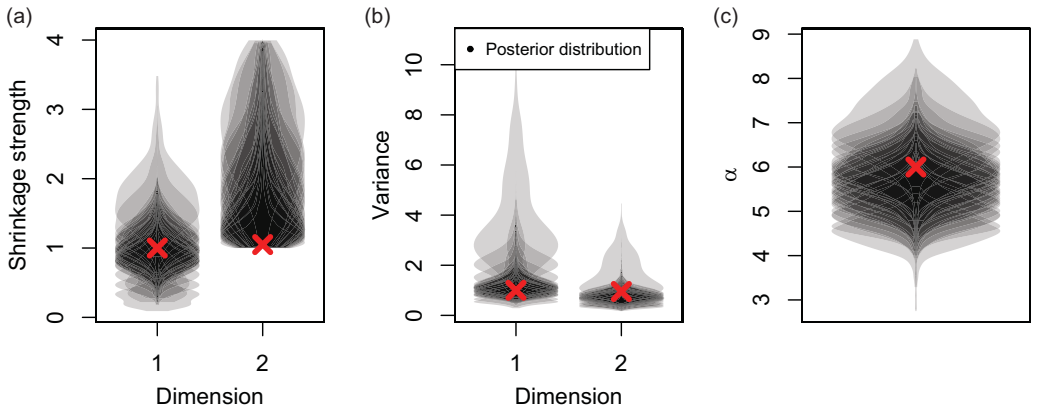

Figure 14. For the simulated network with $p^* = 2$ and $G^* = 3$, the posterior distribution of (a) the shrinkage strength parameter, (b) the latent position variance parameter, and (c) the α parameter.

$$
\begin{aligned}
&\propto \left\{ \delta_h^{\frac{(p-h+1)n}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=1}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right] \right\} \\
&\quad \times \left\{ \delta_h^{\frac{G(p-h+1)}{2}} \exp\left[-\frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=1}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g \right] \right\} \\
&\quad \times \left[\delta_h^{a_2 - 1} \exp(-b_2 \delta_h) \right] \\
&\propto \delta_h^{\frac{n(p-h+1)}{2}} \delta_h^{\frac{G(p-h+1)}{2}} \delta_h^{a_2 - 1} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\delta_h \prod_{m=1, m \neq h}^{\ell} \delta_m \right) \right. \\
&\quad \left. \times \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} - \frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\delta_h \prod_{m=1, m \neq h}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g - b_2 \delta_h \right] \\
&\propto \delta_h^{\frac{(n+G)(p-h+1)}{2} + a_2 - 1} \exp\left\{ -\left[\frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=1, m \neq h}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=1, m \neq h}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_2 \right] \delta_h \right\}
\end{aligned}
$$

since δ_h is bounded between $[1, \infty)$,

$$
\sim \text{Gam}^{\text{T}} \left(\frac{(n+G)(p-h+1)}{2} + a_2, \quad \frac{1}{2} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu}_g)^T (\psi_g) \left(\prod_{m=1, m \neq h}^{\ell} \delta_m \right) \mathbf{I}_p (\mathbf{z}_i - \boldsymbol{\mu}_g) c_{ig} \right.
$$
$$
\left. + \frac{1}{2} \sum_{g=1}^{G} \boldsymbol{\mu}_g^T (\xi^{-1}) \left(\prod_{m=1, m \neq h}^{\ell} \delta_m \right) \mathbf{I}_p \boldsymbol{\mu}_g + b_2, \quad 1 \right)
$$

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

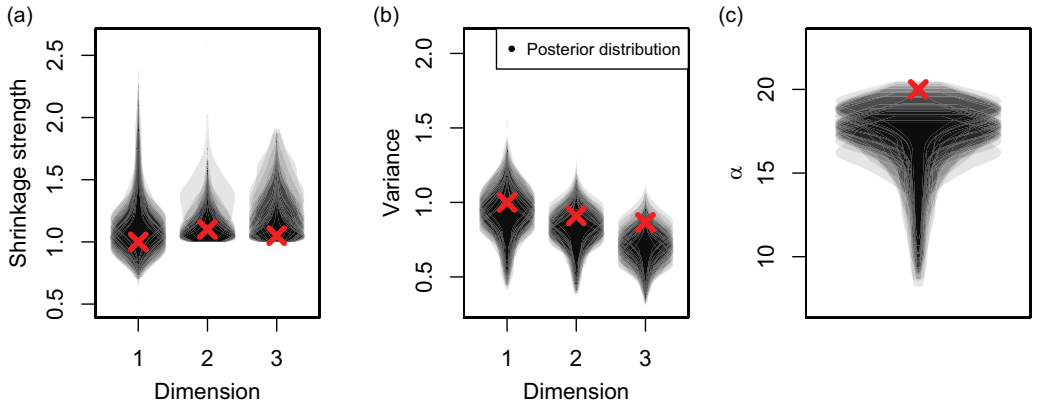

Figure 15. For the simulated network with $p^* = 3$ and $G^* = 7$, the posterior distribution of (a) the shrinkage strength parameter, (b) the latent position variance parameter, and (c) the α parameter.

D. Simulation studies' additional posterior distribution plots

The performance of LSPCM is also assessed through the posterior distributions of latent positions' variance, shrinkage strength, and the parameter α. Included here are supplementary plots to assist in assessing the performance of the LSPCM.

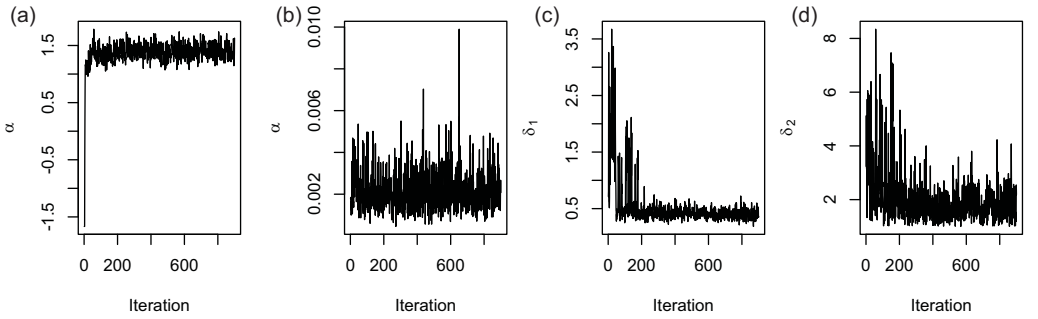

Figure 16. For the football Twitter network: an example of a trace plot for the parameters (a) α, (b) ν, (c) δ_1, and (d) δ_2.

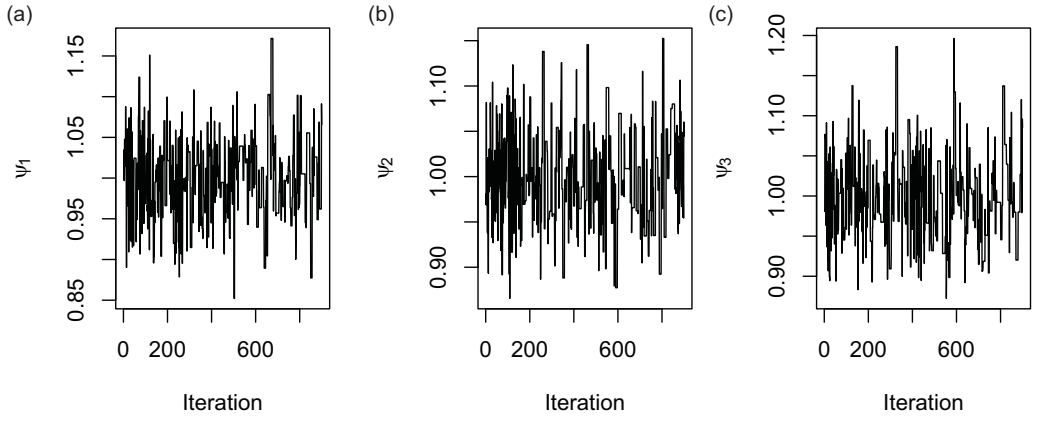

Figure 17. For the football Twitter network: an example of a trace plot for the parameters (a) ψ_1, (b) ψ_2 and (c) ψ_3.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

E. Additional results from analyses of Twitter networks

Supplementary results from applying the LSPCM to the Twitter networks are provided here, including the trace plots (Figures 16, 17, 19 and 20), and the posterior distributions of the variance, shrinkage strength, and the α parameters (Figures 18 and 21). For the Irish politicians' network, Figures 22 and 23 show the LSPCM clustering solution from a single MCMC chain at different values of G and b_ν, which then informed the prior hyperparameter choices used in Section 5.2. Figure 22a and Table 3 show that under the hyperparameter settings detailed, the LSPCM estimated clustering solutions had high variance on the number of non-empty components and resulted in clusters with very small numbers, with some having as few as 3 nodes in a cluster. The context and knowledge of the political composition of the Irish political landscape prompted a lower value of G and and a higher value of b_ν to be employed to better reflect the anticipated

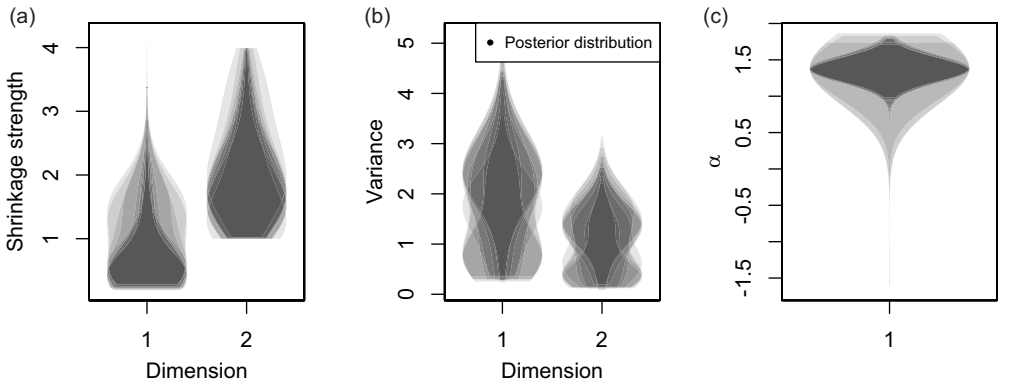

Figure 18. For the football Twitter network, (a) the posterior distributions of the shrinkage strength parameters across dimensions, (b) the posterior distributions of the variance parameters across dimensions, and (c) the posterior distribution of α.

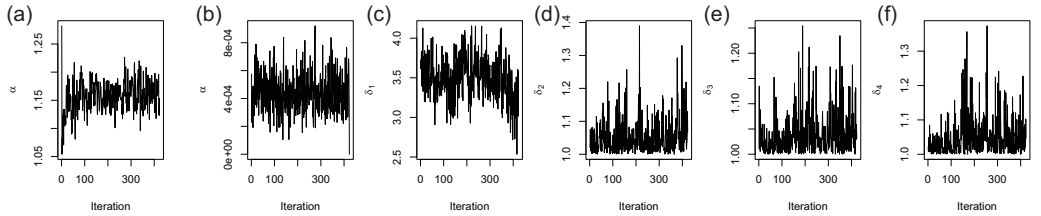

Figure 19. For the Irish politicians Twitter network: an example of a trace plot for the parameters (a) α, (b) ν, (c) δ_1, (d) δ_2, (e) δ_3, and (f) δ_4.

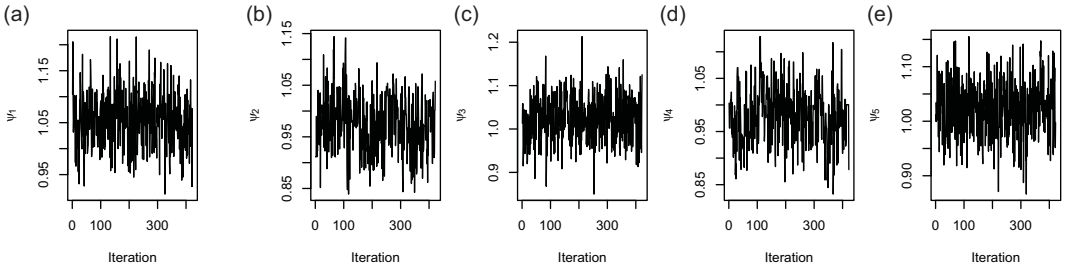

Figure 20. For the Irish politicians' Twitter network: an example of a trace plot for the parameters (a) ψ_1, (b) ψ_2, (c) ψ_3, (d) ψ_4 and (e) ψ_5.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

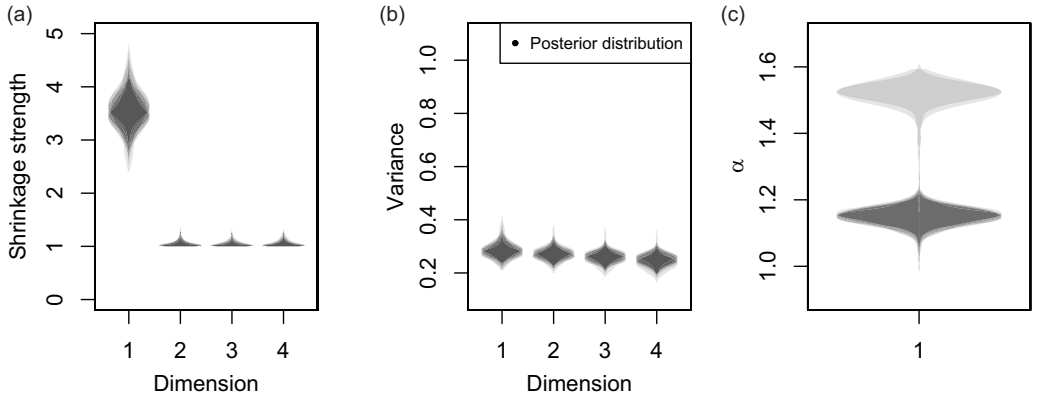

Figure 21. Irish politicians Twitter network, (a) the posterior distribution of the shrinkage strength parameters across dimensions, (b) the posterior distributions of the variance parameters across dimensions, and (c) the posterior distribution of α.

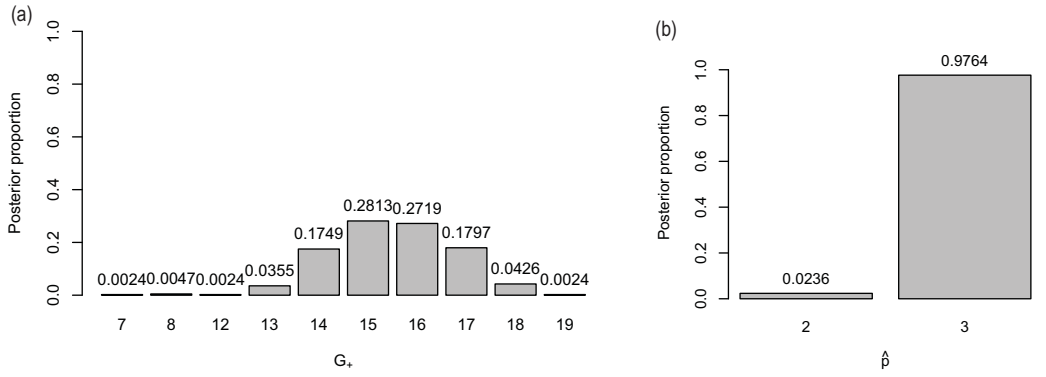

Figure 22. For the Irish politicians network with $a_\nu = 5$, $b_\nu = 5G$, where $G = 20$: (a) the posterior distribution of the number of non-empty components G_+, and (b) the distribution of the effective latent space dimension $\hat{p}$.

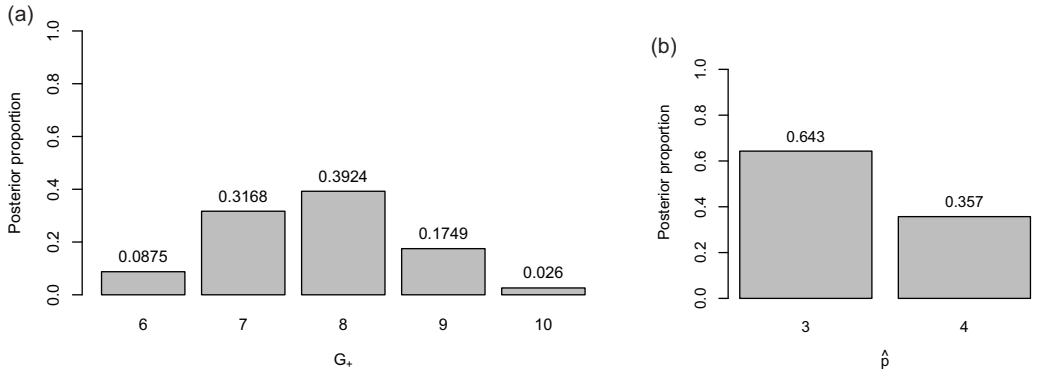

Figure 23. For the Irish politicians network with $a_\nu = 5$, $b_\nu = 10G$, where $G = 10$: (a) the posterior distribution of the number of non-empty components G_+, and (b) the distribution of the effective latent space dimension $\hat{p}$.

https://doi.org/10.1017/nws.2025.10014 Published online by Cambridge University Press

Table 3. For the Irish politicians network with $a_\nu = 5, b_\nu = 5G$, where $G = 20$: cross-tabulation of political party membership and the LSPCM representative cluster labels

		Cluster														
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
	Fine Gael			89		14		2		10		12	11	5		
	Fianna Fáil				3		43	1							2	
Political	*Labour Party*	71						2			6					
Parties	*Independent*	1		2	16			3	3		2				1	3
	Green Party				5				2							
	Sinn Féin		30								1					
	United Left Alliance				1				7							

Table 4. For the Irish politicians network with $a_\nu = 5, b_\nu = 10G$, where $G = 10$: Cross-tabulation of political party membership and the LSPCM representative cluster labels

		Cluster							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	Fine Gael			119	20			2	2
	Fianna Fáil						49		
Political	*Labour Party*	77			2				
Parties	*Independent*	6		2	2	16		5	
	Green Party	1				6			
	Sinn Féin		31						
	United Left Alliance				8				

range of the likely numbers of clusters and to encourage stronger emptying of components to concentrate the nodes in fewer clusters for clearer insight to the clustering structure.

Cite this article: Gwee X. Y., Gormley I. C. and Fop M. (2025). Model-based clustering for network data via a latent shrinkage position cluster model. *Network Science* **13**, 1–33. https://doi.org/10.1017/nws.2025.10014