

# Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities

## Supplementary material

Adrien Todeschini, Xenia Miscouridou and François Caron

INRIA & Institut de Mathématiques de Bordeaux, France  
e-mail: [Adrien.Todeschini@inria.fr](mailto:Adrien.Todeschini@inria.fr)

Department of Statistics, University of Oxford, UK  
e-mail: [xenia.miscouridou@spc.ox.ac.uk](mailto:xenia.miscouridou@spc.ox.ac.uk); [caron@stats.ox.ac.uk](mailto:caron@stats.ox.ac.uk)

### A. Background on completely random measures

#### A.1. Completely random measures

Completely random measures (CRM) were introduced by [Kingman \(1967, 1993\)](#) and are now standard tools for constructing flexible Bayesian nonparametric (BNP) models; see for example the surveys of [Lijoi and Prünster \(2010\)](#) or [Daley and Vere-Jones \(2008, Section 10.1\)](#).

A CRM  $W$  on  $\mathbb{R}_+$  is a random measure such that, for any collection of disjoint measurable subsets  $A_1, \dots, A_n$  of  $\mathbb{R}_+$ ,  $W(A_1), \dots, W(A_n)$  are independent. A CRM can be decomposed into a sum of three independent parts: a non-random measure, a countable collection of atoms with random weights (or masses) at fixed locations, and a countable collection of atoms with random masses and random locations. Here, we will only consider CRMs with random masses and random locations, which take the form

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \quad (\text{S1})$$

where the  $w_i \in \mathbb{R}_+$  are the masses and  $\theta_i \in \mathbb{R}_+$  are the locations. The law of  $W$  can actually be characterized by a Poisson point process  $N = \{(w_i, \theta_i)_{i=1,2,\dots}\}$  on  $\mathbb{R}_+^2$  with mean measure  $\nu(dw, d\theta)$ . We focus here on the case where the CRM is homogeneous with independent increments. This implies that the locations  $\theta_i$  are independent of the weights  $w_i$  and the mean measure decomposes as  $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$  where  $\lambda$  is the Lebesgue measure and  $\rho$  is a measure on  $\mathbb{R}_+$  that satisfies the condition

$$\int_0^{\infty} (1 - e^{-w})\rho(dw) < \infty. \quad (\text{S2})$$

We write  $W \sim \text{CRM}(\rho, \lambda)$ . Note that  $W([0, T]) < \infty$  a.s. for any  $T > 0$  while  $W(\mathbb{R}_+) = \infty$  a.s. if  $\rho$  is not degenerate at 0. If

$$\int_0^{\infty} \rho(dw) = \infty, \quad (\text{S3})$$

then there will be a.s. an infinite number of jumps in any interval  $[0, T]$  and we refer to the CRM as infinite-activity. Otherwise, it is called finite-activity. Let  $\bar{\rho}$  be the tail Lévy intensity defined by

$$\bar{\rho}(x) = \int_x^{\infty} \rho(dw) \quad (\text{S4})$$

for  $x > 0$ . This function corresponds to the expected number of points  $(w_i, \theta_i)$  such that  $w_i > x$  and  $\theta_i \in [0, 1]$ , and its asymptotic properties play an important role in the characterization of the graph properties.

## A.2. Vectors of CRMs

Multivariate extensions of CRMs have been proposed by various authors (Epifani and Lijoi, 2010; Leisen and Lijoi, 2011; Leisen et al., 2013; Griffin et al., 2013; Lijoi et al., 2014). These models are closely related to Lévy copulas (Tankov, 2003; Cont and Tankov, 2003; Kallsen and Tankov, 2006) and multivariate subordinators on cones (Barndorff-Nielsen et al., 2001; Skorohod, 1991). A vector of CRMs  $(W_1, \dots, W_p)$  on  $\mathbb{R}_+$  is a collection of random measures  $W_k$ ,  $k = 1, \dots, p$ , such that, for any collection of disjoint measurable subsets  $A_1, \dots, A_n$  of  $\mathbb{R}_+$ , the vectors  $(W_1(A_1), \dots, W_p(A_1))$ ,  $(W_1(A_2), \dots, W_p(A_2))$ ,  $\dots$ ,  $(W_1(A_n), \dots, W_p(A_n))$  are mutually independent. We only consider here vectors of CRMs with both random weights and random locations. In this case, the measures  $W_k$ ,  $k = 1, \dots, p$ , are a.s. discrete and take the form

$$W_k = \sum_{i=1}^{\infty} w_{ik} \delta_{\theta_i}. \quad (\text{S5})$$

The law of the vector of CRMs can be characterized by a Poisson point process on  $\mathbb{R}_+^{p+1}$  with mean measure  $\nu(dw_1, \dots, dw_p, d\theta)$ . We focus again on vectors of homogeneous CRMs with independent increments where the mean measure of the Poisson process can be written as

$$\nu(dw_1, \dots, dw_p, d\theta) = \rho(dw_1, \dots, dw_p) \lambda(d\theta), \quad (\text{S6})$$

where  $\rho$  is a measure on  $\mathbb{R}_+^p$ , concentrated on  $\mathbb{R}_+^p \setminus \{\mathbf{0}\}$ , which satisfies

$$\int_{\mathbb{R}_+^p} \min \left( 1, \sum_{k=1}^p w_k \right) \rho(dw_1, \dots, dw_p) < \infty. \quad (\text{S7})$$

We use the same notation as for univariate (scalar) CRMs and write simply  $(W_1, \dots, W_p) \sim \text{CRM}(\rho, \lambda)$ . A key quantity is the multivariate Laplace exponent. For  $(t_1, \dots, t_p) \in (0, \infty)^p$ , this is defined by

$$\psi(t_1, \dots, t_p) := -\log \mathbb{E} \left[ e^{-\sum_{k=1}^p t_k W_k([0,1])} \right] \quad (\text{S8})$$

$$= \int_{\mathbb{R}_+^p} \left( 1 - e^{-\sum_{k=1}^p t_k w_k} \right) \rho(dw_1, \dots, dw_p). \quad (\text{S9})$$

Note that this quantity involves a  $p$ -dimensional integral which may not have an analytic form, therefore not easily computable, and may be expensive to evaluate numerically. As for CRMs, if

$$\int_{\mathbb{R}_+^p} \rho(dw_1, \dots, dw_p) = \infty \quad (\text{S10})$$

then there will be an infinite number of  $\theta_i \in [0, T]$  for which  $\sum_k w_{ik} > 0$  and the vector of CRMs is called infinite-activity. Otherwise, it is called finite-activity. Note that some (but not all) CRMs may still be marginally finite-activity.

## B. Proof of Propositions 4 and 5

The proof of (Caron and Fox, 2017, Appendix C) can be directly adapted to the model proposed in this paper which is the multivariate generalization of their framework in a multivariate setting. We provide a sketch of the proof. As  $Z$  is a jointly exchangeable point process verifying the expression in Eq. (21) of the main paper and the moment condition (24), it then follows from the law of large numbers that

$$N_\alpha^{(e)} = \Theta(\alpha^2) \text{ a.s. as } \alpha \rightarrow \infty.$$

**Finite-activity case.** If the vector of CRMs is finite-activity, the locations of the weights arise from a homogeneous Poisson process with finite rate, and  $N_\alpha = \Theta(\alpha)$  a.s. It follows that

$$N_\alpha^{(e)} = \Theta(N_\alpha^2) \text{ a.s. as } \alpha \rightarrow \infty.$$

**Infinite-activity case.** Following [Caron and Fox \(2017\)](#), one can lower bound the counting process of the nodes  $N_\alpha$ , by a counting process  $\tilde{N}_\alpha$  which is conditionally Poisson, and the same proof applies. For infinite-activity CCRMs, we use the fact that  $\psi(W_1([0, \alpha]), \dots, W_p([0, \alpha])) \rightarrow \infty$  a.s., it follows that  $N_\alpha = \Omega(\alpha)$  a.s., and therefore

$$N_\alpha^{(e)} = o(N_\alpha^2) \text{ a.s. as } \alpha \rightarrow \infty.$$

Finally, for compound CRMs with regularly varying  $\rho_0$  with exponent  $\sigma$  and slowly varying function  $\ell$  such that  $\lim_{t \rightarrow \infty} \ell(t) > 0$ , Proposition [S2](#) in Appendix [I](#) implies that  $N_\alpha = \omega(\alpha^{1+\sigma})$  a.s. and

$$N_\alpha^{(e)} = O(N_\alpha^{2/(1+\sigma)}) \text{ a.s. as } \alpha \rightarrow \infty.$$

## C. Details on some derivations

### C.1. Proof of Equation (20)

Note that for  $p \in \mathbb{R}, a, b > 0$  we have the identity

$$\int_0^\infty x^{p-1} e^{-ax-b/x} dx = \frac{2K_p(2\sqrt{ab})}{(a/b)^{p/2}}.$$

Additionally,  $K_\nu(z) = K_{-\nu}(z)$ . We have

$$\begin{aligned} \rho(dw_1, \dots, dw_p) &= \int_0^\infty w_0^{-p} F\left(\frac{dw_1}{w_0}, \dots, \frac{dw_p}{w_0}\right) \rho_0(dw_0) \\ &= \frac{1}{\Gamma(1-\sigma)} \int_0^\infty w_0^{-p-\sigma-1} e^{-\tau w_0} \prod_{k=1}^p \left(\frac{w_k}{w_0}\right)^{a_k-1} e^{-b_k w_k/w_0} \frac{b_k^{a_k}}{\Gamma(a_k)} dw_0 dw_{1:p} \\ &= \frac{1}{\Gamma(1-\sigma)} \left[ \prod_{k=1}^p w_k^{a_k-1} \frac{b_k^{a_k}}{\Gamma(a_k)} \right] \int_0^\infty w_0^{-\sum_k a_k - \sigma - 1} e^{-\tau w_0 - (\sum_k b_k w_k)/w_0} dw_0 dw_{1:p} \\ &= \frac{1}{\Gamma(1-\sigma)} \left[ \prod_{k=1}^p w_k^{a_k-1} \frac{b_k^{a_k}}{\Gamma(a_k)} \right] \frac{2K_{-\sigma - \sum_k a_k}(2\sqrt{\tau \sum_k b_k w_k})}{(\frac{\tau}{\sum_k b_k w_k})^{-(\sigma + \sum_k a_k)/2}} dw_{1:p} \\ &= \frac{2}{\Gamma(1-\sigma)} \left[ \prod_{k=1}^p w_k^{a_k-1} \frac{b_k^{a_k}}{\Gamma(a_k)} \right] \left(\frac{\tau}{\sum_k b_k w_k}\right)^{(\sigma + \sum_k a_k)/2} K_{\sigma + \sum_k a_k} \left(2\sqrt{\tau \sum_k b_k w_k}\right) dw_{1:p}. \end{aligned}$$

## D. Proof of Theorem 3

Let  $D_\alpha^* = \sum_{k=1}^p D_{k\alpha}([0, \alpha])$  be the number of edges in the directed graph of size  $\alpha$ ,  $W_{k,\alpha}^* = W_k([0, \alpha])$  and  $W_\alpha^* = (W_{1,\alpha}^*, \dots, W_{p,\alpha}^*)^T$ . Using Campbell's theorem,

$$\begin{aligned} \mathbb{E}[D_\alpha^*] &= \mathbb{E}[\mathbb{E}[D_\alpha^* | W_\alpha^*]] = \mathbb{E}[(W_\alpha^*)^T W_\alpha^*] \\ &= \mathbb{E}[W_\alpha^*]^T \mathbb{E}[W_\alpha^*] + \text{tr}(\text{cov}(W_\alpha^*)) \\ &= \alpha^2 \mu^T \mu + \alpha \text{tr}(\Sigma) \end{aligned} \tag{S11}$$

where we define

$$\mu = \int_{\mathbb{R}_+^p} w \rho(dw_1, \dots, dw_p), \quad \Sigma = \int_{\mathbb{R}_+^p} w w^T \rho(dw_1, \dots, dw_p).$$

Let  $w_i = (w_{i1}, \dots, w_{ip})$ . Then, using the extended Slivnyak-Mecke theorem ([Møller and Waagepetersen](#),

2003, Theorem 3.3) we obtain

$$\begin{aligned}
\mathbb{E}[N_\alpha^{(e)}] &= \mathbb{E} \left[ \mathbb{E}[N_\alpha^{(e)} | W_1, \dots, W_p] \right] \\
&= \mathbb{E} \left[ \sum_i \mathbf{1}_{\theta_i \leq \alpha} \left[ \left( 1 - e^{-w_i^T w_i} \right) + \frac{1}{2} \sum_{j \neq i} \mathbf{1}_{\theta_j \leq \alpha} \left( 1 - e^{-2w_i^T w_j} \right) \right] \right] \\
&= \alpha \int_{\mathbb{R}_+^p} \left( 1 - e^{-w^T w} \right) \rho(dw_1, \dots, dw_p) + \frac{\alpha^2}{2} \int_{\mathbb{R}_+^p} \psi(2w_1, \dots, 2w_p) \rho(dw_1, \dots, dw_p).
\end{aligned} \tag{S12}$$

Using the extended Slivnyak-Mecke theorem and then Campbell's theorem,

$$\begin{aligned}
\mathbb{E}[N_\alpha] &= \mathbb{E} [\mathbb{E} [N_\alpha | W_1, \dots, W_p]] \\
&= \mathbb{E} \left[ \sum_i \left( 1 - e^{-2w_i^T (\sum_{j \neq i} w_j \mathbf{1}_{\theta_j \leq \alpha}) - w_i^T w_i} \right) \mathbf{1}_{\theta_i \leq \alpha} \right] \\
&= \alpha \int_{\mathbb{R}_+^p} \mathbb{E} \left( 1 - e^{-2w^T (\sum_j w_j \mathbf{1}_{\theta_j \leq \alpha}) - w^T w} \right) \rho(dw_1, \dots, dw_p) \\
&= \alpha \int_{\mathbb{R}_+^p} \left( 1 - e^{-w^T w - \alpha \psi(2w_1, \dots, 2w_p)} \right) \rho(dw_1, \dots, dw_p).
\end{aligned} \tag{S13}$$

By monotone convergence, we have, as  $\alpha$  tends to infinity,

$$\mathbb{E}[N_\alpha] \sim \alpha \int_{\mathbb{R}_+^p} \rho(dw_1, \dots, dw_p)$$

if the CRM is finite-activity and  $\mathbb{E}[N_\alpha] = \omega(\alpha)$  otherwise.

### E. Approximation of the log-posterior density

The posterior probability density function, up to a normalizing constant, takes the form

$$\begin{aligned}
&p \left( (w_{1k}, \dots, w_{N_\alpha k})_{k=1, \dots, p}, \phi, \alpha \mid (z_{ij})_{1 \leq i, j \leq N_\alpha} \right) \\
&\propto \left[ \prod_{i=1}^{N_\alpha} \prod_{j=1}^{N_\alpha} \left( \frac{1 - e^{-\sum_k w_{ik} w_{jk}}}{e^{-\sum_k w_{ik} w_{jk}}} \right)^{z_{ij}} \right] e^{-\sum_{k=1}^p (w_{*k} + \sum_{i=1}^{N_\alpha} w_{ik})^2} \\
&\times \left[ \prod_{i=1}^{N_\alpha} \rho(w_{i1}, \dots, w_{ip}; \phi) \right] \alpha^{N_\alpha} p(\phi, \alpha) g_{*\alpha}(w_{*1}, \dots, w_{*p}; \phi)
\end{aligned} \tag{S14}$$

where  $m_{ik} = \sum_{j=1}^{N_\alpha} n_{ijk} + n_{jik}$  and  $g_{*\alpha}(w_{*1}, \dots, w_{*p}; \phi)$  is the probability density function of the random vector  $(W_1([0, \alpha]), \dots, W_p([0, \alpha]))$ .

This is intractable due to the lack of an analytic expression for  $g_{*\alpha}$ . We can however approximate the log-posterior. Noting that  $w_{*k} = o(\sum_{i=1}^{N_\alpha} w_{ik})$  as  $\alpha \rightarrow \infty$ , we have

$$(w_{*k} + \sum_{i=1}^{N_\alpha} w_{ik})^2 \simeq \left( \sum_{i=1}^{N_\alpha} w_{ik} \right)^2 + 2w_{*k} \sum_{i=1}^{N_\alpha} w_{ik}.$$

Using this approximation, one can now integrate out  $w_{*k}$  and we then obtain the approximation

$$\begin{aligned}
&p((w_{1k}, \dots, w_{N_\alpha k})_{k=1, \dots, p}, \phi, \alpha \mid (z_{ij})_{1 \leq i, j \leq N_\alpha}) \\
&\simeq \left[ \prod_{i=1}^{N_\alpha} \prod_{j=1}^{N_\alpha} \left( \frac{1 - e^{-\sum_k w_{ik} w_{jk}}}{e^{-\sum_k w_{ik} w_{jk}}} \right)^{z_{ij}} \right] e^{-\sum_{k=1}^p (\sum_{i=1}^{N_\alpha} w_{ik})^2} \\
&\times \left[ \prod_{i=1}^{N_\alpha} \rho(w_{i1}, \dots, w_{ip}; \phi) \right] \alpha^{N_\alpha} \exp \left[ -\alpha \psi \left( 2 \sum_{i=1}^{N_\alpha} w_{i1}, \dots, 2 \sum_{i=1}^{N_\alpha} w_{ip}; \phi \right) \right] p(\phi, \alpha)
\end{aligned} \tag{S15}$$

where  $\psi(t_1, \dots, t_p)$  is the multivariate Laplace exponent, which can be evaluated numerically.

## F. Simulation from the tilted truncated random measure

We aim at sampling points of a Poisson point process with (finite) mean measure

$$\begin{aligned} & \alpha e^{-w_0 \sum_{k=1}^p \lambda_k \beta_k} f(\beta_1, \dots, \beta_p) \rho_0(w_0) \mathbb{1}_{w_0 > \varepsilon} \\ &= \frac{e^{-w_0 \sum_{k=1}^p \lambda_k \beta_k} f(\beta_1, \dots, \beta_p)}{M(-\lambda_1 w_0, \dots, -\lambda_p w_0)} \times \alpha M(-\lambda_1 w_0, \dots, -\lambda_p w_0) \rho_0(w_0) \mathbb{1}_{w_0 > \varepsilon} \end{aligned}$$

where we recall that  $M$  is the Laplace transform of the scores. One can first sample points  $(w_{0i})$  from a Poisson point process with mean measure

$$\alpha M(-\lambda_1 w_0, \dots, -\lambda_p w_0) \rho_0(w_0) \mathbb{1}_{w_0 > \varepsilon}, \quad (\text{S16})$$

then sample the scores

$$(\beta_{i1}, \dots, \beta_{ip}) | w_{i0} \stackrel{\text{ind}}{\sim} H(\cdot | w_{i0}) \quad (\text{S17})$$

with  $H$  an exponentially tilted version of  $F$

$$H(d\beta_1, \dots, d\beta_p | w_0) = \frac{e^{-w_0 \sum_{k=1}^p \gamma_k \beta_k} F(d\beta_1, \dots, d\beta_p)}{\int_{\mathbb{R}_+^p} e^{-w_0 \sum_{k=1}^p \gamma_k \tilde{\beta}_k} F(d\tilde{\beta}_1, \dots, d\tilde{\beta}_p)}. \quad (\text{S18})$$

In the case where  $F$  is a product of gamma distributions,  $H$  is also a product of gamma distributions. The mean measure (S16) is a tilted version of  $\rho_0$ , where the tilting function is monotone decreasing. We explain in the next section how to sample from this point process using adaptive thinning. More generally, the mean measure may be of the form of a tilted stable measure (with potentially heavier tails than exponential). One could use in this case the generic series representation developed by [Rosiński \(2007\)](#) for tempered/tilted stable processes.

**Simulation from a tilted truncated generalized gamma process** We want to sample points from a Poisson process with truncated mean measure

$$\rho^\varepsilon(dw) = h(w) w^{-1-\sigma} e^{-\tau w} \mathbb{1}_{w > \varepsilon} dw \quad (\text{S19})$$

where  $h$  is a monotone decreasing and bounded function, and  $(\tau, \sigma)$  verify either  $\tau \geq 0$  and  $\sigma \in (0, 1)$ , or  $\tau > 0$  and  $\sigma \in (-1, 0]$ . We will resort to the method of adaptive thinning ([Lewis and Shedler, 1979](#); [Ogata, 1981](#); [Favaro and Teh, 2013](#)).

For  $\tau > 0$ , consider the family of adaptive bounds

$$g_t(s) = h(t) t^{-1-\sigma} \exp(-\tau s)$$

with  $g_t(s) > \rho(s)$  for  $s > t$ . We have,

$$\begin{aligned} G_t(s) &= \int_t^s g_t(s') ds' \\ &= \frac{h(t)}{\tau} t^{-1-\sigma} (\exp(-\tau t) - \exp(-\tau s)) \end{aligned}$$

and

$$G_t^{-1}(r) = -\frac{1}{\tau} \log \left( \exp(-\tau t) - \frac{r\tau}{t^{-1-\sigma} h(t)} \right).$$

For  $\tau = 0$ , we consider bounds

$$g_t(s) = h(t) s^{-1-\sigma}$$

and we obtain

$$G_t(s) = \frac{h(t)}{\sigma}(t^{-\sigma} - s^{-\sigma})$$

$$G_t^{-1}(r) = \left[ t^{-\sigma} - \frac{r\sigma}{h(t)} \right]^{-1/\sigma}.$$

The adaptive thinning sampling scheme is as follows:

1. Set  $N = \emptyset, t = \varepsilon$ .
2. Iterate until termination:
  - (a) Draw  $r \sim \text{Exp}(1)$ .
  - (b) If  $r > G_t(\infty)$ , terminate; else set  $t' = G_t^{-1}(r)$ .
  - (c) With probability  $\rho^\varepsilon(t')/g_t(t')$  accept sample  $t'$  and set  $N = N \cup \{t'\}$ .
  - (d) Set  $t = t'$  and continue.
3. Return  $N$ .

$N$  obtained by the sampling scheme above is a draw from the Poisson random measure with intensity  $\rho^\varepsilon$  on  $[\varepsilon, +\infty)$ . The efficiency of this approach depends on the acceptance probability. For  $\tau > 0$  and  $s > t$  this is given by the ratio

$$\frac{\rho^\varepsilon(s)}{g_t(s)} = \frac{h(s)s^{-1-\sigma}}{h(t)t^{-1-\sigma}} < 1.$$

## G. Bipartite networks

It is possible to use a construction similar to that of Section 2 to model bipartite graphs, and extend the model of Caron (2012). A bipartite graph is a graph with two types of nodes, where only connections between nodes of different types are allowed. Nodes of the first type are embedded at locations  $\theta_i \in \mathbb{R}_+$ , and nodes of the second type at locations  $\theta'_j \in \mathbb{R}_+$ . The bipartite graph is represented by a (non-symmetric) point process

$$Z = \sum_{i,j} z_{ij} \delta_{(\theta_i, \theta'_j)}, \quad (\text{S20})$$

where  $z_{ij} = 1$  if there is an edge between node  $i$  of type 1 and node  $j$  of type 2.

**Statistical Model.** We consider the model

$$W_1, \dots, W_p \sim \text{CRM}(\rho, \lambda)$$

$$W'_1, \dots, W'_p \sim \text{CRM}(\rho', \lambda)$$

and for  $k = 1, \dots, p$ ,

$$D_k | W_k, W'_k \sim \text{Poisson}(W_k \times W'_k)$$

$$D_k = \sum_{i,j} n_{ijk} \delta_{(\theta_i, \theta'_j)}.$$

In (S20),  $z_{ij}$  is given by  $z_{ij} = \min(1, \sum_{k=1}^p n_{ijk})$ .

**Posterior inference.** We derive here the inference algorithm when  $(W_1, \dots, W_p)$  and  $(W'_1, \dots, W'_p)$  are compound CRMs with  $F$  and  $\rho_0$  taking the form (17) and (18) respectively.

Assume that we observe a set of connections  $z = (z_{ij})_{i=1, \dots, N_\alpha; j=1, \dots, N'_\alpha}$ . For  $k = 1, \dots, p$ , we introduce latent variables  $n_{ijk}$ , for  $1 \leq i \leq N_\alpha$ ,  $1 \leq j \leq N'_\alpha$ ,

$$(n_{ij1}, \dots, n_{ijp}) | w, w', z \sim \begin{cases} \delta_{(0, \dots, 0)} & \text{if } z_{ij} = 0 \\ \text{tPoisson}(w_{i1}w'_{j1}, \dots, w_{ip}w'_{jp}) & \text{if } z_{ij} = 1. \end{cases}$$

The objective is to approximate the posterior density

$$p\left((w_{10}, \dots, w_{N_\alpha 0}), (\beta_{1k}, \dots, \beta_{N_\alpha k}, w_{*k})_{k=1, \dots, p}, (w'_{10}, \dots, w'_{N'_\alpha 0}), (\beta'_{1k}, \dots, \beta'_{N'_\alpha k}, w'_{*k})_{k=1, \dots, p}, \phi, \alpha, \phi', \alpha' | z\right).$$

Let  $m_{ik} = \sum_{j=1}^{N'_\alpha} n_{ijk}$  and  $m_i = \sum_{k=1}^p m_{ik}$ . The MCMC algorithm iterates as follows:

1. Update  $(\alpha, \phi) | \text{rest}$  using a Metropolis-Hastings step.
2. Update

$$w_{i0} | \text{rest} \sim \text{Gamma}\left(m_i - \sigma, \tau + \sum_{k=1}^p \beta_{ik} \left[ \gamma_k + \left( \sum_{j=1}^{N'_\alpha} w'_{jk} \right) + w'_{*k} \right]\right).$$

3. Update

$$\beta_{ik} | \text{rest} \sim \text{Gamma}\left(a_k + m_{ik}, b_k + w_{i0} \left[ \gamma_k + \left( \sum_{j=1}^{N'_\alpha} w'_{jk} \right) + w'_{*k} \right]\right).$$

4. Update  $(w_{*1}, \dots, w_{*p}) | \text{rest}$ .
5. Update the latent variables  $n_{ijk} | \text{rest}$ .
6. Repeat steps 1-4 to update  $(\alpha', \phi'), (w'_{10}, \dots, w'_{N'_\alpha 0}), (\beta'_{1k}, \dots, \beta'_{N'_\alpha k})_{k=1, \dots, p}$  and  $(w'_{*1}, \dots, w'_{*p})$ .

## H. Gaussian approximation of the sum of small jumps

**Theorem S1** Consider the multivariate random variable  $X_\varepsilon \in \mathbb{R}_+^p$  with moment generating function

$$\mathbb{E}[e^{-t^T X_\varepsilon}] = \exp\left[-\alpha \int_{\mathbb{R}_+^p} \left(1 - e^{-\sum_{k=1}^p t_k w_k}\right) \rho_\varepsilon(dw_1, \dots, dw_p)\right],$$

where  $\alpha > 0$  and

$$\rho_\varepsilon(dw_1, \dots, dw_p) = e^{-\sum_{k=1}^p \gamma_k w_k} \int_0^\varepsilon w_0^{-p} F\left(\frac{dw_1}{w_0}, \dots, \frac{dw_p}{w_0}\right) \rho_0(dw_0),$$

with  $\varepsilon > 0$ ,  $\rho_0$  a Lévy measure on  $\mathbb{R}_+$  and  $F$  a probability distribution on  $\mathbb{R}_+^p$ .  $F$  admits the density  $f$  verifying the conditions

$$\int_0^\infty f(zu_1, \dots, zu_p) dz > 0 \text{ } U\text{-almost everywhere}$$

and

$$\int_{\mathbb{R}_+^p} \|\beta_{1:p}\|^2 f(\beta_1, \dots, \beta_p) d\beta_{1:p} < \infty,$$

where  $U$  is the uniform distribution on the unit sphere  $S^{p-1}$ . Then, if  $\rho_0$  is a regularly varying Lévy measure with exponent  $\sigma \in (0, 1)$ , i.e.

$$\int_x^\infty \rho_0(dw_0) \stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x)$$

where  $\ell : (0, \infty) \rightarrow (0, \infty)$  is a slowly varying function, then

$$\Sigma_\varepsilon^{-1/2}(X_\varepsilon - \mu_\varepsilon) \xrightarrow{d} \mathcal{N}(0, I_p)$$

as  $\varepsilon \rightarrow 0$ , where

$$\begin{aligned} \mu_\varepsilon &= \alpha \int_{\mathbb{R}_+^p} w \rho_\varepsilon(dw_1, \dots, dw_p) \\ \Sigma_\varepsilon &= \alpha \int_{\mathbb{R}_+^p} w w^T \rho_\varepsilon(dw_1, \dots, dw_p) \end{aligned}$$

with

$$\begin{aligned}\mu_\varepsilon &\sim \alpha \mathbb{E}[\beta] \frac{\sigma}{1-\sigma} \varepsilon^{1-\sigma} \ell(1/\varepsilon) \text{ as } \varepsilon \rightarrow 0 \\ \Sigma_\varepsilon &\sim \alpha \mathbb{E}[\beta \beta^T] \frac{\sigma}{2-\sigma} \varepsilon^{2-\sigma} \ell(1/\varepsilon) \text{ as } \varepsilon \rightarrow 0\end{aligned}$$

where the distribution of  $\beta$  is  $F$ .

**Proof.** We write the model in spherical form. Let  $r = \sqrt{\sum w_k^2}$  and  $u_k = \frac{w_k}{r}$ , for  $k = 1, \dots, p-1$ . The determinant of the Jacobian is  $\frac{r^{p-1}}{\sqrt{1-\sum_{k=1}^{p-1} u_k^2}}$  and therefore

$$\begin{aligned}\tilde{\rho}_\varepsilon(r, u_1, \dots, u_{p-1}) &= \frac{r^{p-1}}{u_p} e^{-r \sum_{k=1}^p \gamma_k u_k} \int_0^\varepsilon w_0^{-p} f\left(\frac{ru_1}{w_0}, \dots, \frac{ru_p}{w_0}\right) \rho_0(dw_0) dr du_{1:p-1} \\ &:= \mu_\varepsilon(dr|u_{1:p-1}) U(du_{1:p-1})\end{aligned}$$

where  $u_p = \sqrt{1 - \sum_{k=1}^{p-1} u_k^2}$ ,  $\mu_\varepsilon(dr|u) = r^{p-1} e^{-r \sum_{k=1}^p \gamma_k u_k} \int_0^\varepsilon w_0^{-p} f\left(\frac{ru_1}{w_0}, \dots, \frac{ru_p}{w_0}\right) \rho_0(dw_0) dr$  and  $U(du) = \frac{1}{u_p} du_{1:p}$  is the uniform distribution on the unit sphere  $S^{p-1}$ .

In order to apply Theorem 2.4 in [Cohen and Rosinski \(2007\)](#) (see also [Asmussen and Rosiński, 2001](#)), we need to show that there exists a function  $b_\varepsilon : (0, 1] \rightarrow (0, +\infty)$  such that

$$\lim_{\varepsilon \rightarrow 0} \frac{\sigma_\varepsilon(u)}{b_\varepsilon} > 0, \text{ } U\text{-almost everywhere} \quad (\text{S21})$$

where

$$\sigma_\varepsilon^2(u) = \int_0^\infty r^2 \mu_\varepsilon(dr|u)$$

and for every  $\kappa > \varepsilon$

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{b_\varepsilon^2} \int_{\|w_{1:p}\| > \kappa b_\varepsilon} \|w_{1:p}\|^2 \rho_\varepsilon(dw_1, \dots, dw_p) = 0. \quad (\text{S22})$$

Assume that  $\int_0^\infty f(zu_1, \dots, zu_p) dz > 0$   $U$ -almost everywhere. With the change of variable  $z = \frac{r}{w_0}$ , and the dominated convergence theorem we obtain

$$\begin{aligned}\sigma_\varepsilon^2(u) &= \int_0^\infty z^{p+1} f(zu_1, \dots, zu_p) \left[ \int_0^\varepsilon e^{-zw_0 \sum_{k=1}^p \gamma_k u_k} w_0^2 \rho_0(dw_0) \right] dz \\ &\sim \left( \int_0^\infty z^{p+1} f(zu_1, \dots, zu_p) dz \right) \left( \int_0^\varepsilon w_0^2 \rho_0(dw_0) \right) \text{ as } \varepsilon \rightarrow 0 \\ &\sim \left( \int_0^\infty z^{p+1} f(zu_1, \dots, zu_p) dz \right) \frac{\sigma}{2-\sigma} \varepsilon^{2-\sigma} \ell(1/\varepsilon) \text{ as } \varepsilon \rightarrow 0.\end{aligned}$$

Take  $b_\varepsilon = \varepsilon^{1-\sigma/2} \sqrt{\ell(1/\varepsilon)}$ , we then obtain

$$\lim_{\varepsilon \rightarrow 0} \frac{\sigma_\varepsilon^2(u)}{b_\varepsilon^2} = \left( \int_0^\infty z^{p+1} f(zu_1, \dots, zu_p) dz \right) \frac{\sigma}{2-\sigma} > 0, \text{ } U\text{-almost everywhere.} \quad (\text{S23})$$

Now consider for any  $\kappa > 0$ ,

$$\begin{aligned}I_\varepsilon &= \int_{\|w_{1:p}\| > \kappa b_\varepsilon} \|w_{1:p}\|^2 \nu_\varepsilon(dw_1, \dots, dw_p) \\ &= \int_0^\varepsilon \int_{\|\beta_{1:p}\| > \frac{\kappa b_\varepsilon}{w_0}} w_0^2 \|\beta_{1:p}\|^2 e^{-w_0 \sum_{k=1}^p \gamma_k \beta_k} f(\beta_1, \dots, \beta_p) \rho_0(dw_0) d\beta_{1:p}.\end{aligned}$$



For  $w_0 \in (0, \varepsilon)$ , we have  $\frac{\kappa b_\varepsilon}{w_0} \geq \frac{\kappa b_\varepsilon}{\varepsilon} = \varepsilon^{-\sigma/2} \ell(1/\varepsilon) > \kappa_2 \varepsilon^{-\sigma/4}$  for  $\varepsilon$  small enough as  $t^\delta \ell(t) \rightarrow 0$  for any  $\delta > 0$  as  $t \rightarrow \infty$ . So for  $\varepsilon$  small enough

$$\begin{aligned} I_\varepsilon &< \int_0^\varepsilon \int_{\|\beta_{1:p}\| > \kappa_2 \varepsilon^{-\sigma/4}} w_0^2 \|\beta_{1:p}\|^2 e^{-w_0 \sum_{k=1}^p \gamma_k \beta_k} f(\beta_1, \dots, \beta_k) \rho_0(dw_0) d\beta_{1:p} \\ &< \left[ \int_{\|\beta_{1:p}\| > \kappa_2 \varepsilon^{-\sigma/4}} \|\beta_{1:p}\|^2 f(\beta_1, \dots, \beta_k) d\beta_{1:p} \right] \left[ \int_0^\varepsilon w_0^2 \rho_0(dw_0) \right]. \end{aligned}$$

As  $[\int_0^\varepsilon w_0^2 \rho_0(dw_0)] \sim \frac{\sigma}{2-\sigma} b_\varepsilon^2$  when  $\varepsilon \rightarrow 0$ , we conclude that

$$\lim_{\varepsilon \rightarrow 0} \frac{I_\varepsilon}{b_\varepsilon^2} \leq \lim_{\varepsilon \rightarrow 0} \frac{\sigma}{2-\sigma} \int_{\|\beta_{1:p}\| > \kappa_2 \varepsilon^{-\sigma/4}} \|\beta_{1:p}\|^2 f(\beta_1, \dots, \beta_k) d\beta_{1:p} = 0. \quad (\text{S24})$$

Eq. (S23) and (S24) with Theorem 2.4 of [Cohen and Rosinski \(2007\)](#) yield

$$\Sigma_\varepsilon^{-1/2} (X_\varepsilon - \mu_\varepsilon) \xrightarrow{d} \mathcal{N}(0, I_p)$$

as  $\varepsilon \rightarrow 0$ , where

$$\begin{aligned} \mu_\varepsilon &= \alpha \int_{\mathbb{R}_+^p} w_{1:p} \rho_\varepsilon(dw_1, \dots, dw_p) \\ &= \alpha \int_{\mathbb{R}_+^p} \int_0^\varepsilon w_0 \beta_{1:p} e^{-w_0 \sum_{k=1}^p \gamma_k \beta_k} \rho_0(dw_0) f(\beta_1, \dots, \beta_p) d\beta_{1:p} \\ &\sim \alpha \mathbb{E}[\beta_{1:p}] \frac{\sigma}{1-\sigma} \varepsilon^{1-\sigma} \ell(1/\varepsilon) \text{ as } \varepsilon \rightarrow 0 \end{aligned}$$

and

$$\begin{aligned} \Sigma_\varepsilon &= \alpha \int_{\mathbb{R}_+^p} w_{1:p} w_{1:p}^T \rho_\varepsilon(dw_1, \dots, dw_p) \\ &\sim \alpha \mathbb{E}[\beta_{1:p} \beta_{1:p}^T] \frac{\sigma}{2-\sigma} \varepsilon^{2-\sigma} \ell(1/\varepsilon) \text{ as } \varepsilon \rightarrow 0 \end{aligned}$$

using the dominated convergence theorem and lemmas [S3](#) and [S4](#). ■

## I. Technical lemmas

**Proposition S2** *Let  $\nu$  be a Lévy measure defined by Eq. (5) and (13) and let  $\psi$  be its multivariate Laplace exponent. Assume that  $\bar{\rho}_0$  is a regularly varying function with exponent  $\sigma \in (0, 1)$*

$$\bar{\rho}_0 \stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x). \quad (\text{S25})$$

*Then  $\psi$  is (multivariate) regularly varying ([Resnick, 2013](#)), with exponent  $\sigma$ . More precisely, for any  $(x_1, \dots, x_p) \in (0, \infty)^p$ , we have*

$$\begin{aligned} \psi(tx_1, \dots, tx_p) &= \int_{\mathbb{R}_+^p} \left(1 - e^{-t \sum_{k=1}^p x_k w_k}\right) \nu(dw_1, \dots, dw_p) \\ &\stackrel{t \uparrow \infty}{\sim} t^\sigma \Gamma(1-\sigma) \ell(t) \mathbb{E} \left[ \left( \sum_{k=1}^p x_k \beta_k \right)^\sigma \right]. \end{aligned}$$

**Proof.**

$$\begin{aligned} \psi(tx_1, \dots, tx_p) &= \int_{\mathbb{R}_+^p} \left(1 - e^{-t \sum_{k=1}^p x_k w_k}\right) \nu(dw_1, \dots, dw_p) \\ &= \int_{\mathbb{R}_+^p} \left(1 - e^{-t \sum_{k=1}^p x_k w_k}\right) \nu(dw_1, \dots, dw_p) \\ &= \int_{\mathbb{R}_+^p} f(\beta_1, \dots, \beta_p) \left[ \int_0^\infty \left(1 - e^{-w_0 t \sum_{k=1}^p x_k \beta_k}\right) e^{-w_0 \sum_{k=1}^p \gamma_k \beta_k} \rho_0(dw_0) \right] d\beta_{1:p}. \end{aligned}$$

Using Lemmas S3, S4, and the dominated convergence theorem we get

$$\psi(tx_1, \dots, tx_p) \stackrel{t \uparrow \infty}{\sim} t^\sigma \Gamma(1 - \sigma) \ell(t) \int_{(0, \infty)^p} \left( \sum_{k=1}^p x_k \beta_k \right)^\sigma f(\beta_1, \dots, \beta_p) d\beta_{1:p}.$$

■

**Lemma S3** *If*

$$\int_x^\infty \rho(dw) \stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x)$$

where  $\sigma \in (0, 1)$  and  $\ell$  is a slowly varying function (at infinity), then

$$\int_x^\infty e^{-cw} \rho(dw) \stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x).$$

**Proof.**

$$\begin{aligned} \int_x^\infty e^{-cw} \rho(dw) &= \int_x^\infty \rho(dw) - \int_x^\infty (1 - e^{-cw}) \rho(dw) \\ &\stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x) \end{aligned}$$

as  $\int_0^\infty (1 - e^{-cw}) \rho(dw) < \infty$  for any  $c > 0$ . ■

**Lemma S4** (Gnedin et al., 2007; Bingham et al., 1989). Let  $\rho$  be a Lévy measure with regularly varying tail Lévy intensity

$$\int_x^\infty \rho(dw) \stackrel{x \downarrow 0}{\sim} x^{-\sigma} \ell(1/x) \quad (\text{S26})$$

where  $\sigma \in (0, 1)$  and  $\ell$  is a slowly varying function (at infinity). Then (S26) is equivalent to the following

$$\int_0^x w^k \rho(dw) \stackrel{x \downarrow 0}{\sim} \frac{\sigma}{k - \sigma} x^{k - \sigma} \ell(1/x) \quad (\text{S27})$$

$$\int_0^\infty (1 - e^{-tw}) \rho(dw) \stackrel{t \uparrow \infty}{\sim} \Gamma(1 - \sigma) t^\sigma \ell(t) \quad (\text{S28})$$

for any  $k \geq 1$ .

## J. Empirical analysis of the Cohen and Rosinski's Gaussian approximation

We provide here some empirical analysis of the Gaussian approximation described in Section 4 in the main paper. We use the parameters  $p = 1, a = 0.2, b = 1/2, \tau = 1$ . We simulate graphs of different sizes  $\alpha = 5, 10, 100, 200$  for  $\sigma = 0.2$  and  $\sigma = 0.5$ , and set for each graph  $\lambda_k = w_{*k} + 2 \sum_i w_{ik}$ . We then simulate the points of a Poisson process with tilted measure (42), using a small truncation threshold  $T = 10^{-8}$ . For different thresholds  $\varepsilon = 10^{-3}, 10^{-1}, 1, 100$ , we compare the true distribution of  $X_\varepsilon$  to its (truncated) Gaussian approximation. Results are reported in Figures S1-S3. While the theoretical results reported in Section 4 indicate that the Gaussian approximation holds for a fixed  $\alpha$  as  $\varepsilon \rightarrow 0$ , it appears from the simulations that the distribution of  $X_\varepsilon$  is well approximated by a Gaussian distribution for large values of  $\alpha$ , and this whatever the value of  $\varepsilon$ . This is supported by the empirical analysis of the convergence of the Markov chain for different values of  $\varepsilon$  in Section K.1.

## K. Study on the accuracy of the MCMC sampler

### K.1. Sensitivity on truncation level $\varepsilon$

This section shows that the MCMC sampler targets the same distribution under different threshold values (or truncation levels)  $\varepsilon$ . Recall that the threshold is the truncation level for simulating  $w_{*1,p}$  in the infinite-activity case. We perform a sensitivity analysis on the impact of the threshold  $\varepsilon$  for which we use the values

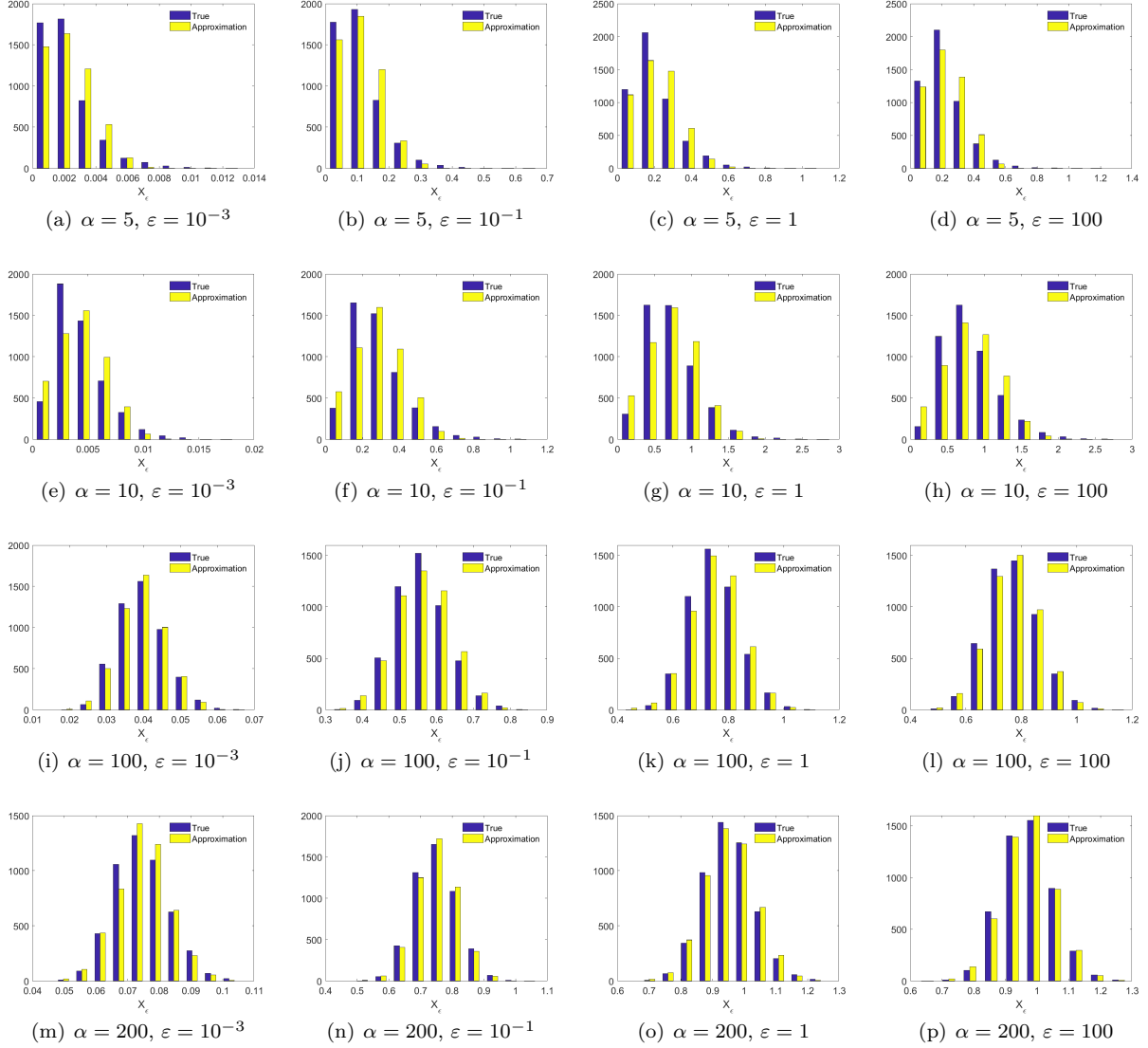


FIG S1. Comparison of the Gaussian approximation for  $\sigma = 0.01$  for different network sizes  $\alpha = 5, 10, 100, 200$  and different thresholds  $\varepsilon = 10^{-3}, 10^{-1}, 1, 100$ .

$100, 0.1, 10^{-4}$  and  $10^{-7}$ . We first generate an undirected graph with  $p = 2$  communities and parameters  $\alpha = 50$ ,  $\sigma = 0.5$ ,  $\tau = 1$ ,  $b_k = b = 0.1$ ,  $a_k = a = 0.2$ . The sampled graph has 1799 nodes and 13,416 edges. For the inference, we consider that  $b$  is known and we assume a vague prior  $\text{Gamma}(0.01, 0.01)$  on the unknown parameters  $\alpha$  and  $\phi = (1 - \sigma, \tau, a)$ . We run 3 parallel MCMC chains and set  $\mathcal{L} = 10$  leapfrog steps in the HMC. The stepsizes of both the HMC and the random walk MH on  $(\log(1 - \sigma), \log \tau, \log a)$  are adapted during the first 50,000 iterations so that to target acceptance ratios of 0.65 and 0.23 respectively. We initialize the chains from the parameter values we use for generating the graph, i.e. from the stationary regime. This is because the objective is to assess the mixing of the chains under different thresholds and see whether chains admit the same limiting distribution. We run 200,000 iterations using our model with  $p$  communities. We repeat this procedure 4 times, using each time one of the truncation values  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ .

The trace plots for the parameters  $\log \tilde{\alpha}$ ,  $\sigma$ ,  $\tilde{b}$ ,  $a$  and  $\bar{w}_*$  are shown in Figure S4. The dotted green line corresponds to the value of the model parameters used for generating the graph. It is clear that the chains are very close to each other and they exhibit a good mixing. The plots in Figure S5 show the trace and autocorrelation function of the approximate log-posterior density (up to a constant). There is no noticeable

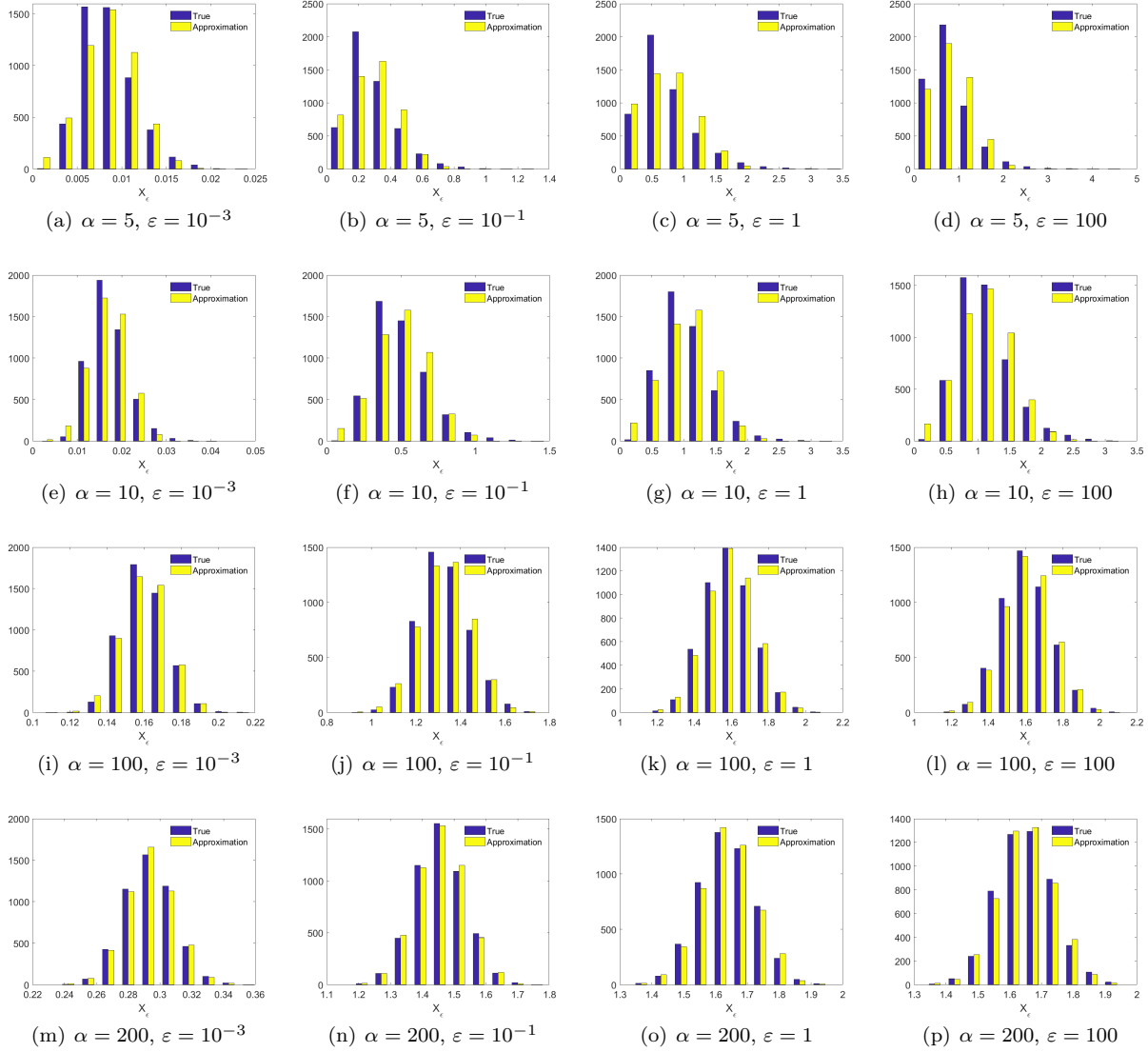


FIG S2. Comparison of the Gaussian approximation for  $\sigma = 0.2$  for different network sizes  $\alpha = 5, 10, 100, 200$  and different thresholds  $\varepsilon = 10^{-3}, 10^{-1}, 1, 100$ .

difference in the limiting distribution nor mixing times for different values of  $\varepsilon$ .

We repeat the sensitivity experiment on inference for another graph generated with  $\sigma = 0.2$ . We generate an undirected graph with  $p = 2$  communities and model parameters  $\alpha = 200$ ,  $\sigma = 0.2$ ,  $\tau = 1$ ,  $b_k = b = 0.5$ ,  $a_k = a = 0.2$ . The sampled graph has 1268 nodes and 15,811 edges. For the inference, we place the same assumptions for the priors and run 3 MCMC chains in parallel with  $\mathcal{L} = 10$  leapfrog steps in the HMC. We repeat the same procedure for each of the values of the threshold  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ . Note that in both experiments ( $\sigma = 0.5$  and  $\sigma = 0.2$ ) we show the results for one of the 3 chains for each threshold since the behaviour was very similar across chains.

The MCMC trace plots for the parameters  $\log \tilde{\alpha}$ ,  $\sigma$ ,  $\tilde{b}$ ,  $a$  and  $\bar{w}_*$  are shown in Figure S6. As in the case of  $\sigma = 0.2$ , the MCMC trace is very similar for the different thresholds in all parameters. The trace and autocorrelation of the approximate log-posterior are shown in Figure S7.

Note that this type of sensitivity test concerns only the infinity-activity case which corresponds to the positive  $\sigma$  values. We have shown the plots of the MCMC output for values of  $\sigma = 0.5$  and  $\sigma = 0.2$  however the same behaviour was observed for other positive  $\sigma$  values too.

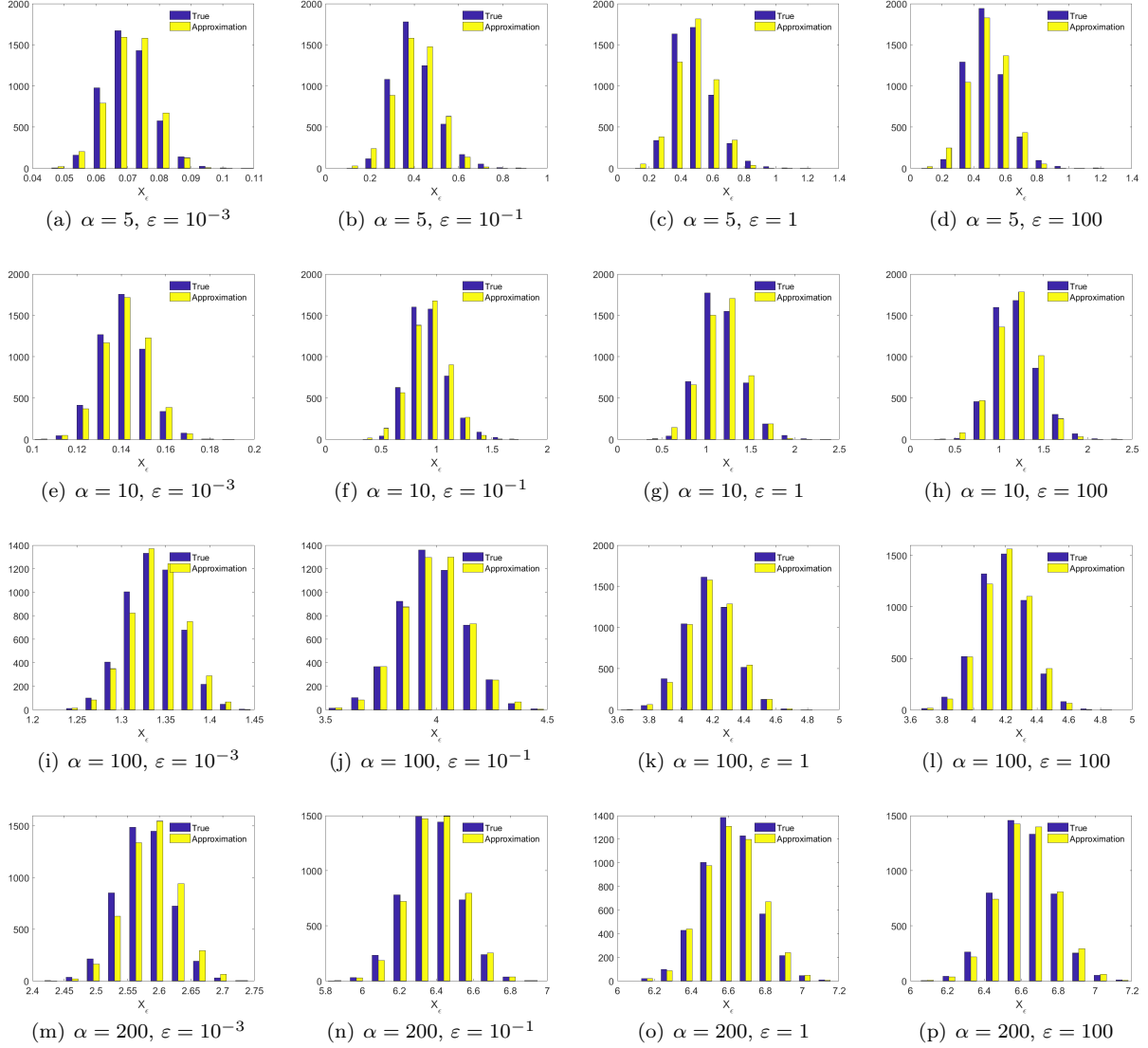


FIG S3. Comparison of the Gaussian approximation for  $\sigma = 0.5$  for different network sizes  $\alpha = 5, 10, 100, 200$  and different thresholds  $\varepsilon = 10^{-3}, 10^{-1}, 1, 100$ .

## K.2. Sensitivity on leapfrog steps $\mathcal{L}$

This section shows a sensitivity analysis on the impact of the number of leapfrog steps  $\mathcal{L}$  on the MCMC sampler which we vary  $\mathcal{L} = 1, 2, 10, 20$ . Recall that the leapfrog steps is the number of steps in the HMC which updates the parameters of the nodes. We generate an undirected graph with  $p = 2$  communities and parameters  $\alpha = 200, \sigma = 0.2, \tau = 1, b_k = b = 0.5, a_k = a = 0.2$ . The sampled graph has 1799 nodes and 13,416 edges. For the inference, we assume as before a vague prior  $\text{Gamma}(0.01, 0.01)$  on the unknown parameters  $\alpha$  and  $\phi = (1 - \sigma, \tau, a)$ . We run 3 parallel MCMC chains with different initial values. Each chain starts with 10,000 iterations using our model with only one community where the scores  $\beta$  are fixed to 1, which is equivalent to the model of [Caron and Fox \(2017\)](#). The chains use 1,000,000 iterations each and a truncation level  $\varepsilon = 10^{-3}$ . We repeat the same procedure (i.e. run the sampler with these settings and 3 parallel chains) 4 times; one for each value of leapfrog steps  $\mathcal{L}$ . Note that here we show the results for one chain since the behaviour was very similar cross chains. The trace plot of the parameters is shown in Figure S8. As shown, the chains exhibit a good mixing and they target the same posterior distribution.

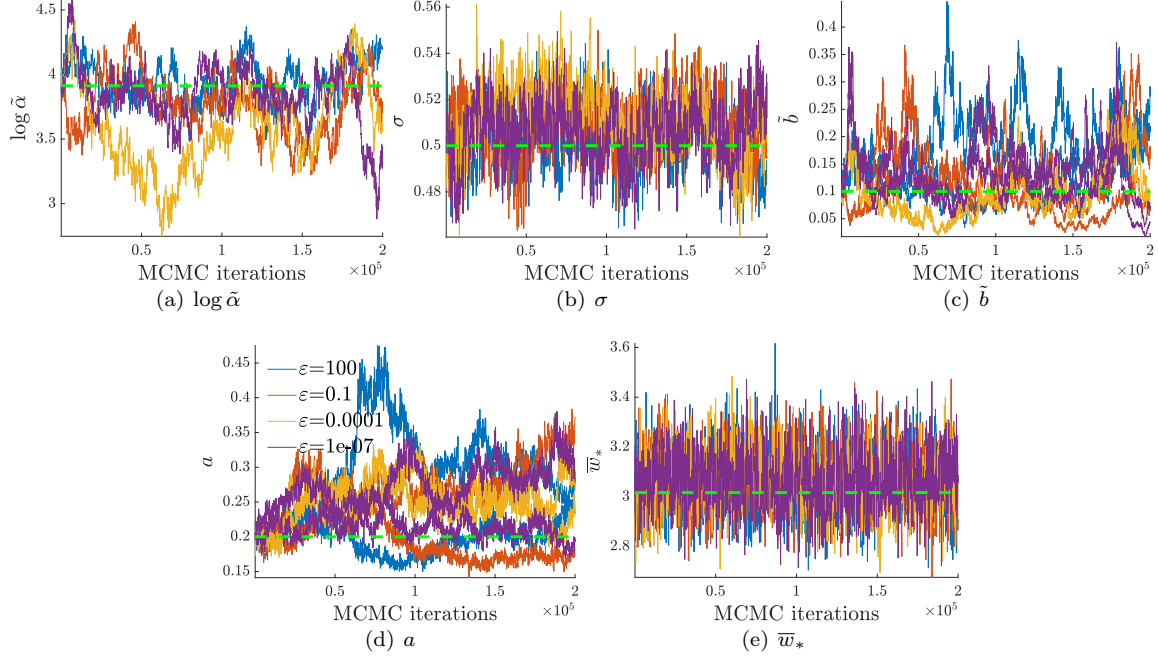


FIG S4. *MCMC trace plots for parameters (a)  $\log \tilde{\alpha}$ , (b)  $\sigma$ , (c)  $\tilde{b}$  and (d)  $a$  and (e)  $\bar{w}_*$  when performing inference with truncation levels  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 50$ ,  $\sigma = 0.5$ ,  $\tau = 1$ ,  $b_k = b = 0.1$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the model parameters used for generating the graph.*

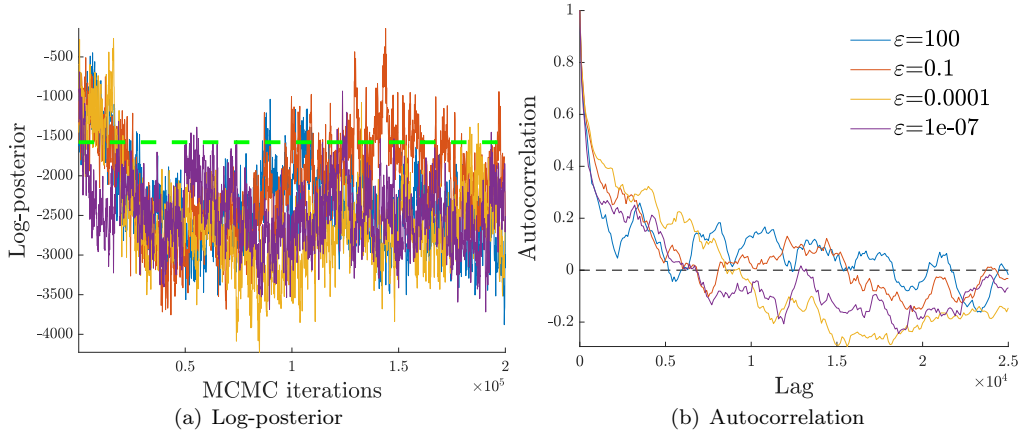


FIG S5. *Trace and autocorrelation of the approximate log-posterior probability density (up to a constant) when performing inference with truncation levels  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 50$ ,  $\sigma = 0.5$ ,  $\tau = 1$ ,  $b_k = b = 0.1$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the approximate log-posterior using the model parameters used for generating the graph.*

Clearly the parameters converge near the samples values used for generating the graph. Figure S9 shows the trace and autocorrelation plot of the log-posterior probability distribution. Unsurprisingly, chains need more time to converge when using 1 or 2 leapfrog steps compared to 10 and 20. However, as shown, there is no difference between 10 and 20, that is why we decide to use 10 as our default value and we also set it so in our real datasets experiments.

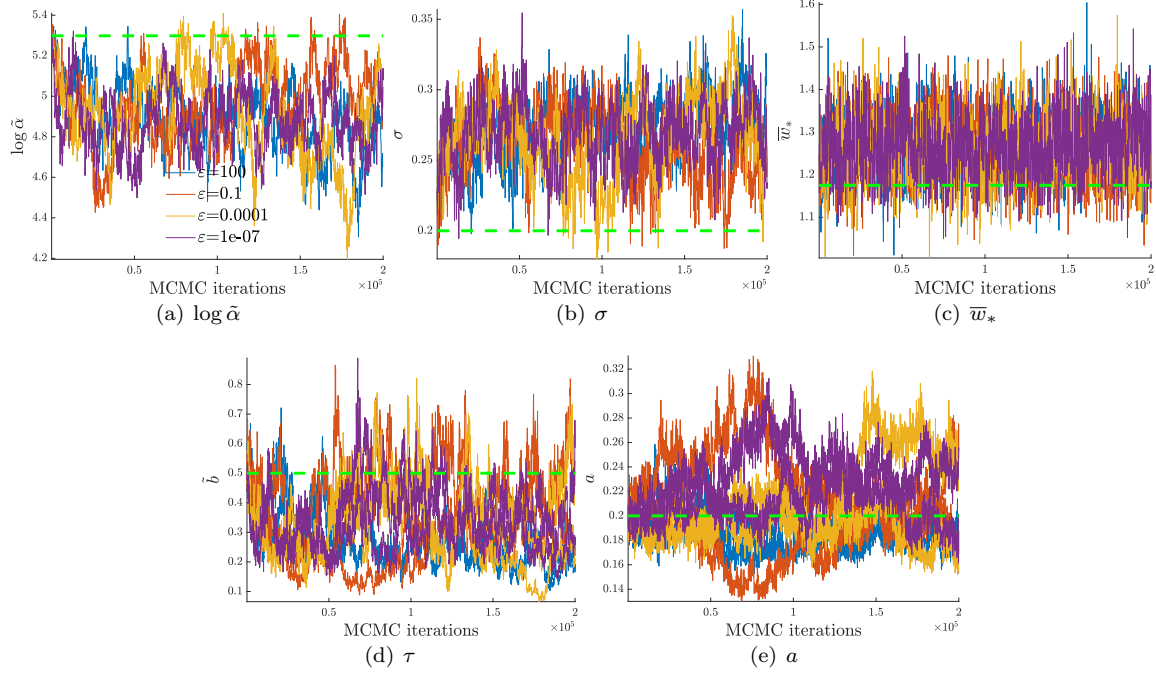


FIG S6. MCMC trace plots for parameters (a)  $\log \tilde{\alpha}$ , (b)  $\sigma$ , (c)  $\tilde{w}$ , (d)  $\tau$  and (e)  $\bar{w}_*$  when performing inference with truncation levels  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 50$ ,  $\sigma = 0.2$ ,  $\tau = 1$ ,  $b_k = b = 0.1$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the model parameters used for generating the graph.

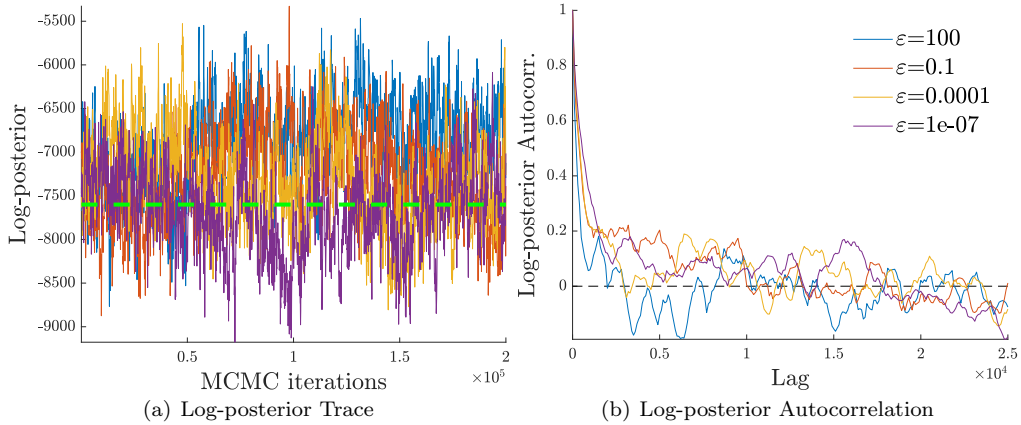


FIG S7. Trace and autocorrelations of the approximate log-posterior probability density (up to a constant) when performing inference with truncation levels  $\varepsilon = 100, 0.1, 10^{-4}, 10^{-7}$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 50$ ,  $\sigma = 0.2$ ,  $\tau = 1$ ,  $b_k = b = 0.1$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the approximate log-posterior using the model parameters used for generating the graph.

## L. Experiments on finite-activity simulated data

In the main paper we only show results for infinite-activity model, i.e. sparser graphs generated with positive values of  $\sigma$ . In this section we demonstrate the applicability of our methods to the case of denser graphs which correspond to finite-activity and equivalently to negative values of  $\sigma$ . In the dense regime ( $\sigma < 0$ ), our

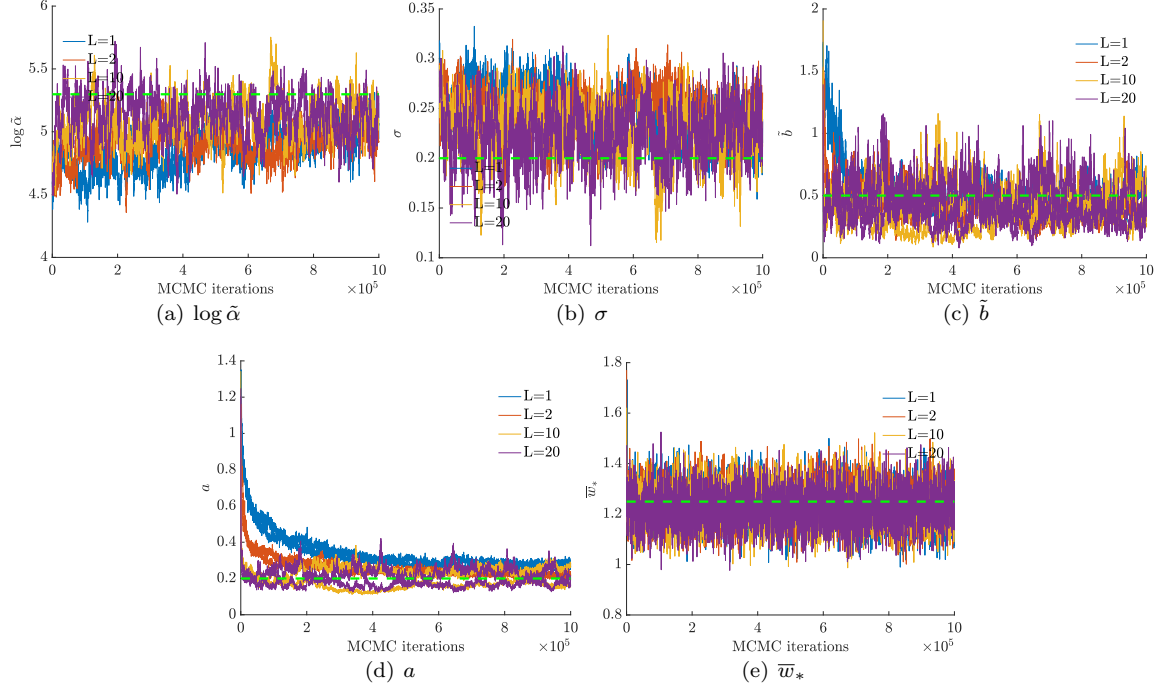


FIG S8. MCMC trace plots for parameters (a)  $\log \tilde{\alpha}$ , (b)  $\sigma$ , (c)  $\tilde{b}$ , (d)  $a$  and (e)  $\bar{w}_*$  when performing inference with number of leapfrog steps  $L = 1, 2, 10, 20$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 200$ ,  $\sigma = 0.2$ ,  $\tau = 1$ ,  $b_k = b = 0.5$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the model parameters used for generating the graph.

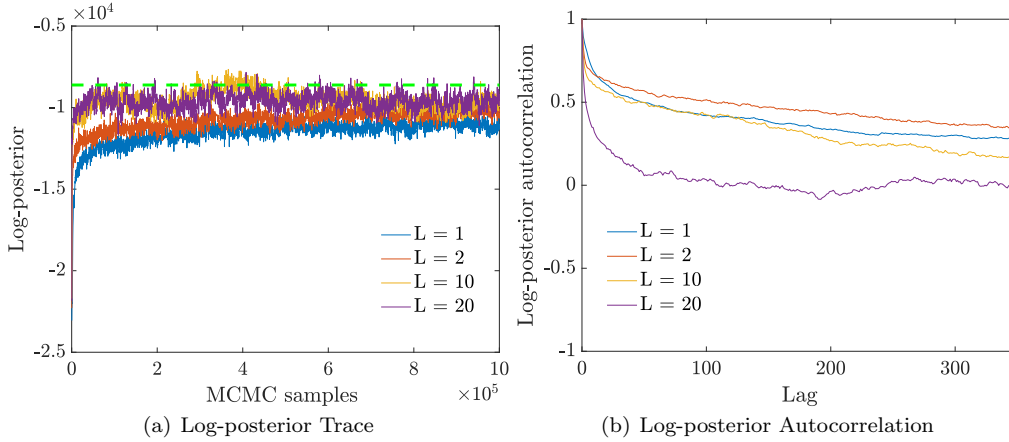


FIG S9. Trace and autocorrelations of the approximate log-posterior probability density (up to a constant) when performing inference with number of leapfrog steps  $L = 1, 2, 10, 20$ . The graph is undirected and generated with  $p = 2$  communities and parameters  $\alpha = 200$ ,  $\sigma = 0.2$ ,  $\tau = 1$ ,  $b_k = b = 0.5$ ,  $a_k = a = 0.2$ . The dotted green line, corresponds to the value of the approximate log-posterior using the model parameters used for generating the graph.

model with mean measure

$$\nu(w_0, \beta_{1:p}) = \frac{\alpha}{\Gamma(1-\sigma)} w_0^{-1-\sigma} e^{-\tau} \prod_{k=1}^p \text{Gamma}(\beta_k; a_k, b_k),$$

has a finite number of atoms  $K_\alpha \geq N_\alpha$ . These atoms have iid weights  $(w_{0i}, \beta_{i,1:p})_{i=1,\dots,K_\alpha}$  given  $K_\alpha$ , where  $w_{0i} \sim \text{Gamma}(-\sigma, \tau)$  and  $\beta_k \sim \text{Gamma}(a_k, b_k)$ . The following parameters are more interpretable in the



dense regime. Firstly, the number of atoms  $K_\alpha$  has mean

$$\varsigma_1 = \int \nu(w_0, \beta_{1:p}) = \frac{-\alpha\tau^\sigma}{\sigma}. \quad (\text{S29})$$

The mean weights are, for  $k = 1, \dots, p$

$$\varsigma_{2,k} = \mathbb{E}[w_{0i}\beta_{ik}] = \frac{-\sigma}{\tau} \frac{a_k}{b_k}. \quad (\text{S30})$$

The variance of the weights is, for  $k = 1, \dots, p$

$$\varsigma_{3,k} = \text{Var}[w_{0i}\beta_{ik}] = \frac{a_k\sigma(\sigma - a_k - 1)}{\tau^2 b_k^2}. \quad (\text{S31})$$

In all our experiments we use the same parameter values  $b = b_k$ ,  $a = a_k$  for all  $k = 1 \dots p$ , and therefore we use simpler notation of  $\varsigma_2 = \varsigma_{2,k}$  and  $\varsigma_3 = \varsigma_{3,k}$ .

### ***L.1. Finite-activity graph with $\sigma = -2$***

We generate an undirected graph with  $p = 2$  communities and parameters  $\alpha = 3000$ ,  $\sigma = -2$ ,  $\tau = 1$ ,  $b_k = b = 3$ ,  $a_k = a = 0.2$ . The sampled graph has 1318 nodes and 101,110 edges. For the inference, we consider that  $b$  is known and we assume a vague prior  $\text{Gamma}(0.01, 0.01)$  on the unknown parameters  $\alpha$  and  $\phi = (1 - \sigma, \tau, a)$ . We run 3 parallel MCMC chains with different initial values and use the exact same settings as in the simulations in Section 5.1 of the main paper. Figure S10 gives the trace plots of the identifiable parameters which are (a)  $\varsigma_1$ , (b)  $\varsigma_2$ , (c)  $\varsigma_3$  and (d)  $\bar{w}_*$  where  $\varsigma_1$  is as given in Eq. (S29),  $\varsigma_2 = \varsigma_{2,k}$  as in Eq. (S30) and  $\varsigma_3 = \varsigma_{3,k}$ , as in Eq. (S31). The dotted green line, corresponds to the value of the model parameters used for generating the graph. Figure S10(e)-(f) shows the trace and autocorrelation plots of the approximate log-posterior distribution (up to a normalizing constant). The dotted green line corresponds to the value of the approximate log-posterior computed using the model parameters.

Our model is able to accurately recover the mean parameters of both low and high degree nodes and to provide reasonable credible intervals, as shown in Figure S11(a)-(b). By generating 500 graphs from the posterior predictive we assess the model fit to the degree distribution and show in Figure S11(c) that our model's posterior predictive degree distribution fits well the empirical degree distribution of the dense generated graph.

### ***L.2. Finite-activity graph with $\sigma = -1$***

This section repeats the same procedure as above but for  $\sigma = -1$ . We generate an undirected graph with  $p = 2$  communities and parameters  $\alpha = 1500$ ,  $\sigma = -1$ ,  $\tau = 1$ ,  $b_k = b = 2$ ,  $a_k = a = 0.2$ . The sampled graph has 1219 nodes and 57,537 edges. We use the same priors and initialization procedure as in Section L.1. We give the trace plots of the identifiable parameters in Figure S12(a)-(d). The dotted green line as before corresponds to the value of the model parameters used for generating the graph. Posterior samples clearly converge around the sampled value. Figure S12(e)-(f) shows the trace and autocorrelation plots of the log-posterior probability density function (up to a normalizing constant). The dotted green line corresponds to the value of the approximate log-posterior computed using the model parameters. The log-posterior trace indicates convergence of the chains and the autocorrelation shows that the correlation of samples decreases quickly with increasing lag.

In Figure S13(a)-(b) we give the credible intervals of the mean parameters of both high and low degree nodes. The model is able to accurately recover the mean parameters. We also generate 500 graphs from the posterior predictive and show that our model fits the empirical degree distribution of the dense generated graph as shown in Figure S13(c).

## **M. Additional plots for experiments on real world data**

In this section we provide additional MCMC trace plots for the two real world datasets experiments we perform in the main paper `USairport` and `polblogs`.

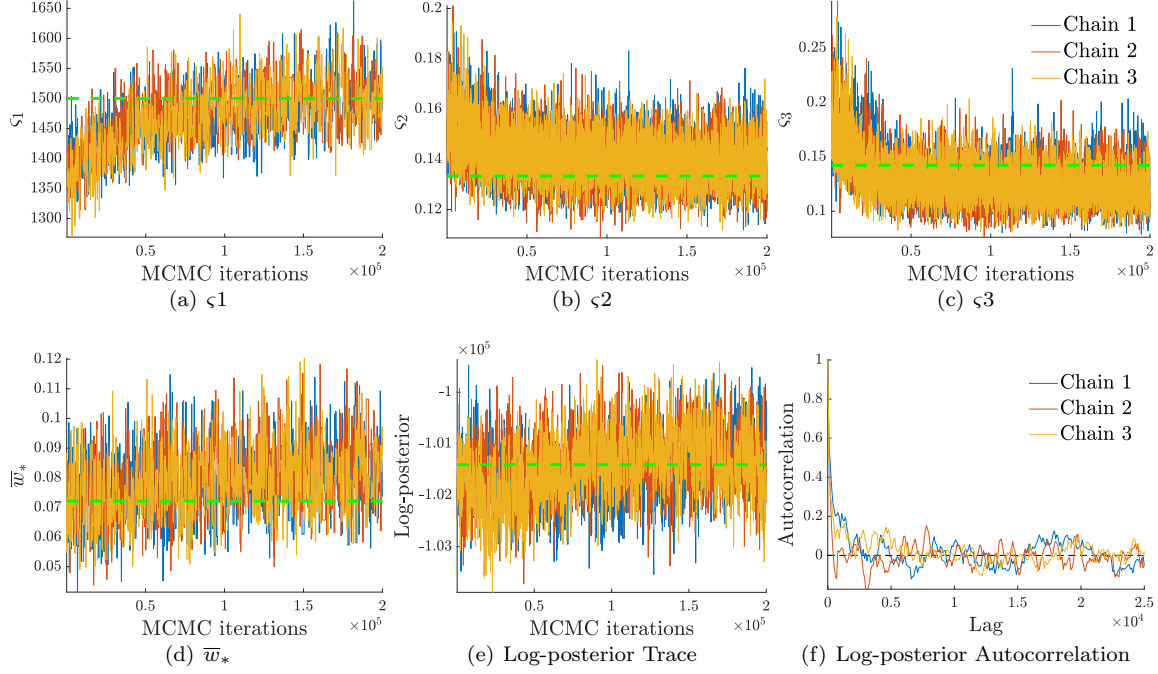


FIG S10. MCMC trace plots of parameters (a)  $\zeta_1$ , (b)  $\zeta_2$ , (c)  $\zeta_3$ , (d)  $\bar{w}_*$ , (e) trace plot of the log-posterior, (f) autocorrelation of the log-posterior for a graph generated with parameters  $p = 2$ ,  $\alpha = 3000$ ,  $\sigma = -2$ ,  $\tau = 1$ ,  $b = 3$ ,  $a = 0.2$ . The dotted green line in (a) corresponds to the value of  $\bar{w}_*$  of the sampled graph, in (b) to the value of the approximate log-posterior (up to a constant) under the model parameters used to generate the graph.

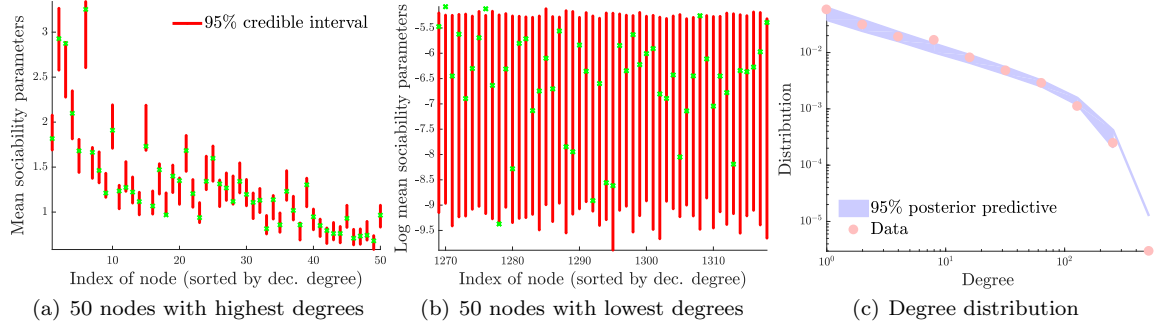


FIG S11. 95% posterior credible intervals and true values of (a) the mean parameters  $\bar{w}_i = \frac{1}{p} \sum_{k=1}^p w_{ik}$  of the 50 nodes with highest degrees and (b) the log mean parameters  $\log \bar{w}_i$  of the 50 nodes with lowest degrees. (c) Empirical degree distribution and 95% posterior predictive credible interval. Results obtained for a graph generated with parameters  $p = 2$ ,  $\alpha = 3000$ ,  $\sigma = -2$ ,  $\tau = 1$ ,  $b = 3$ ,  $a = 0.2$ .

### M.1. Polblogs

For the `polblogs` dataset we provide trace plots for the identifiable parameters  $\zeta_1$ ,  $\zeta_2$ ,  $\zeta_3$  as given above in section L in Eq. (S29), (S30) and (S31). As shown in the trace plots, the MCMC chains exhibit a good mixing and the graphs suggest convergence of the sampler. Similarly, the approximate log-posterior probability distribution, shown in Figure 11 of the main paper also demonstrates convergence of the sampler.

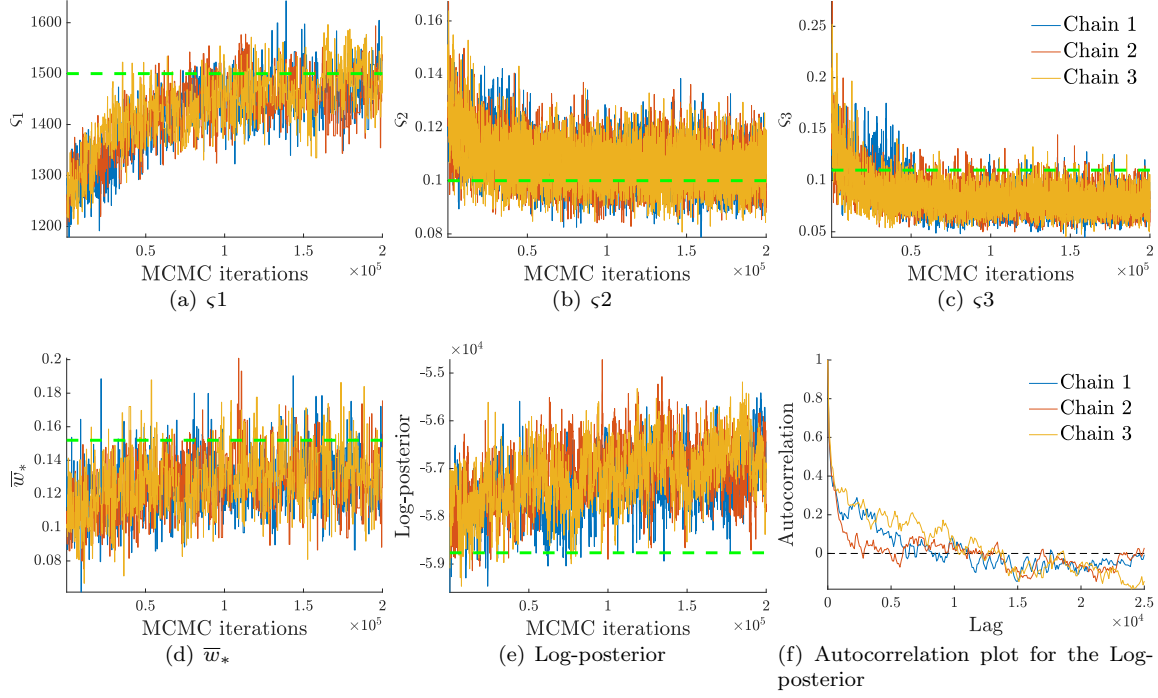


FIG S12. *MCMC trace plots of parameters (a)  $\varsigma_1$ , (b)  $\varsigma_2$ , (c)  $\varsigma_3$ , (d)  $\bar{w}_*$ , (e) trace plot of the log-posterior, (f) autocorrelation of the log-posterior for a graph generated with parameters  $p = 2$ ,  $\alpha = 1500$ ,  $\sigma = -1$ ,  $\tau = 1$ ,  $b = 2$ ,  $a = 0.2$ . The dotted green line in (a) corresponds to the value of  $\bar{w}_*$  of the sampled graph, in (b) to the value of the approximate log-posterior (up to a constant) under the model parameters used to generate the graph.*

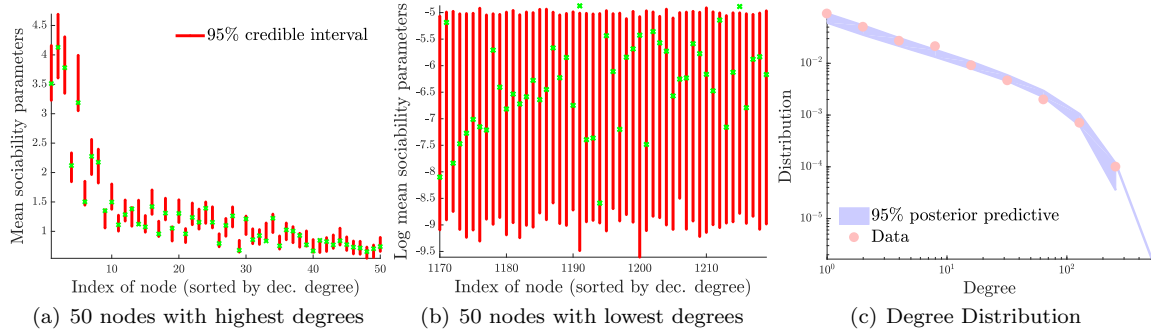


FIG S13. *95% posterior credible intervals and true values of (a) the mean parameters  $\bar{w}_i = \frac{1}{p} \sum_{k=1}^p w_{ik}$  of the 50 nodes with highest degrees and (b) the log mean parameters  $\log \bar{w}_i$  of the 50 nodes with lowest degrees. (c) Empirical degree distribution and 95% posterior predictive credible interval. Results obtained for a graph generated with parameters  $p = 2$ ,  $\alpha = 1500$ ,  $\sigma = -1$ ,  $\tau = 1$ ,  $b = 2$ ,  $a = 0.2$ .*

## M.2. USairport

For the **USairport** dataset we provide trace plots for the parameters  $\log \tilde{\alpha}$ ,  $\sigma$ ,  $\tilde{b}$ ,  $a$  and  $\bar{w}_*$  shown in Figure S15(a)-(e). Figure S15 suggests that the posterior distribution is multimodal as one chain seems to be in a different mode than the other two. We ran longer chains, but this did not lead to improvement in the mixing for this dataset, and we acknowledge that any credible interval for the hyperparameters on this dataset should be considered with cautiousness. Note however that the two modes correspond to parameters with similar approximate log-posterior density (see Figure 11 in the main paper). Similarly, the credible intervals for the sociability parameters, community structure and posterior predictive within each individual chain are again

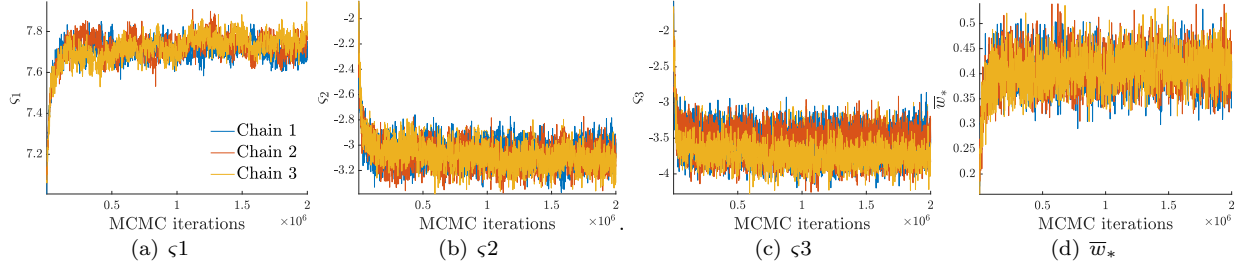


FIG S14. MCMC trace plots of parameters (a)  $\varsigma_1$ , (b)  $\varsigma_2$ , (c)  $\varsigma_3$  and (d)  $\bar{w}_*$  for *polblogs*.

quantitatively very close to those given in the main paper when considering the whole chain.

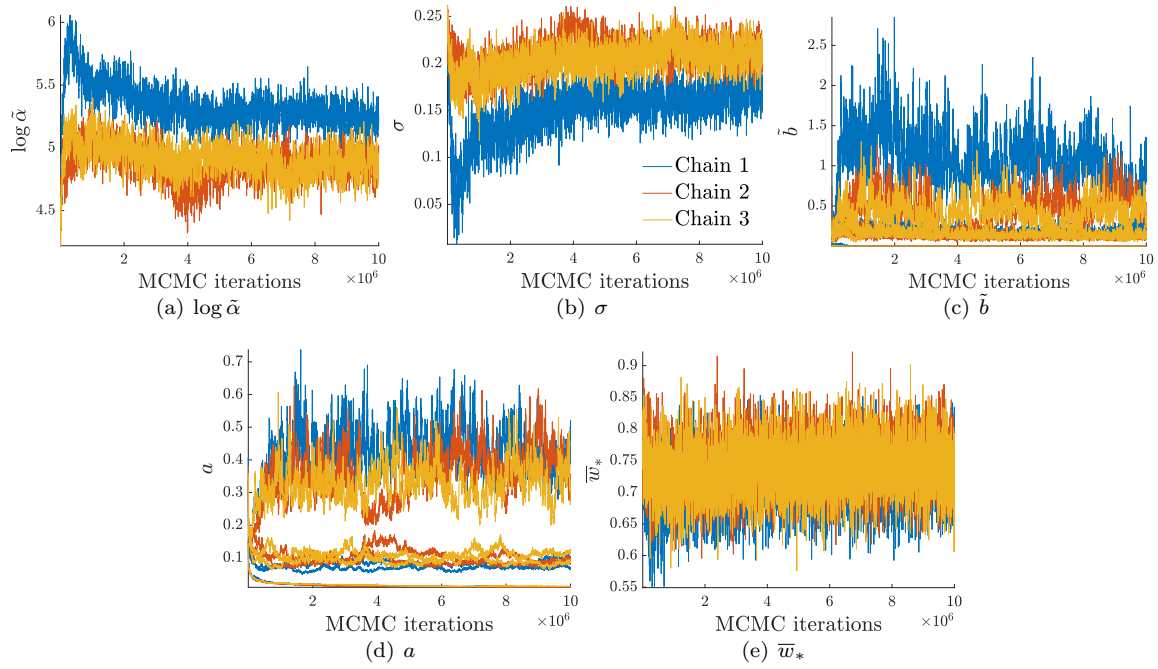


FIG S15. MCMC trace plots of parameters (a)  $\log \tilde{\alpha}$ , (b)  $\sigma$ , (c)  $\tilde{b}$ , (d)  $a$  and (e)  $\bar{w}_*$  for *USairport*.

## References

- S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability*, pages 482–493, 2001.
- O. E. Barndorff-Nielsen, J. Pedersen, and K.-I. Sato. Multivariate subordination, self-decomposability and stability. *Advances in Applied Probability*, 33:160–187, 2001.
- N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- F. Caron. Bayesian nonparametric models for bipartite graphs. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2051–2059. Curran Associates, Inc., 2012.
- F. Caron and E. B. Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 2017.

- S. Cohen and J. Rosinski. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.
- R. Cont and P. Tankov. *Financial modelling with jump processes*, volume 2. CRC press, 2003.
- D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Springer Verlag, 2nd edition edition, 2008.
- I. Epifani and A. Lijoi. Nonparametric priors for vectors of survival functions. *Statistica Sinica*, pages 1455–1484, 2010.
- S. Favaro and Y.W. Teh. MCMC for normalized random measure mixture models. *Statistical Science*, 28(3):335–359, 2013.
- A. Gneden, B. Hansen, and J. Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv*, 4(146-171):88, 2007.
- J. E. Griffin, M. Kolossiaty, and M. F. J. Steel. Comparing distributions by using dependent normalized random-measure mixtures. *Journal of the Royal Statistical Society: Series B*, 75(3):499–529, 2013.
- J. Kallsen and P. Tankov. Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572, 2006.
- J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.
- J.F.C. Kingman. *Poisson processes*. Oxford University Press, USA, 1993.
- F. Leisen and A. Lijoi. Vectors of two-parameter Poisson–Dirichlet processes. *Journal of Multivariate Analysis*, 102(3):482–495, 2011.
- F. Leisen, A. Lijoi, and D. Spanó. A vector of Dirichlet processes. *Electronic Journal of Statistics*, 7:62–90, 2013.
- P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.
- A. Lijoi and I. Prünster. Models beyond the Dirichlet process. In N. L. Hjort, C. Holmes, P. Müller, and S.G. Walker, editors, *Bayesian nonparametrics*, volume 28, page 80. Camb. Ser. Stat. Probab. Math, 2010.
- A. Lijoi, B. Nipoti, and I. Prünster. Bayesian inference with dependent normalized completely random measures. *Bernoulli*, 20(3):1260–1291, 2014.
- J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.
- Y. Ogata. On Lewis’ simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.
- S. Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.
- J. Rosiński. Tempering stable processes. *Stochastic processes and their applications*, 117(6):677–707, 2007.
- A. V. Skorohod. *Random processes with independent increments*, volume 47. Springer, 1991.
- P. Tankov. Dependence structure of spectrally positive multidimensional Lévy processes. *Unpublished manuscript*, 2003.