



**HAL**  
open science

# Rapidly Varying Completely Random Measures for Modeling Extremely Sparse Networks

Valentin Kilian, Benjamin Guedj, François Caron

► **To cite this version:**

Valentin Kilian, Benjamin Guedj, François Caron. Rapidly Varying Completely Random Measures for Modeling Extremely Sparse Networks. 2025. <hal-05080529>

**HAL Id: hal-05080529**

**<https://inria.hal.science/hal-05080529v1>**

Preprint submitted on 22 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Rapidly Varying Completely Random Measures for Modeling Extremely Sparse Networks

Valentin Kilian<sup>1</sup> Benjamin Guedj<sup>2,3</sup> François Caron<sup>1</sup>

<sup>1</sup>Department of Statistics, University of Oxford

<sup>2</sup>Department of Computer Science, University College London

<sup>3</sup>Inria, France

kilian@stats.ox.ac.uk, b.guedj@ucl.ac.uk, caron@stats.ox.ac.uk

Completely random measures (CRMs) are fundamental to Bayesian non-parametric models, with applications in clustering, feature allocation, and network analysis. A key quantity of interest is the Laplace exponent, whose asymptotic behavior determines how the random structures scale. When the Laplace exponent grows nearly linearly — known as rapid variation — the induced models exhibit approximately linear growth in the number of clusters, features, or edges with sample size or network nodes. This regime is especially relevant for modeling sparse networks, yet existing CRM constructions lack tractability under rapid variation. We address this by introducing a new class of CRMs with index of variation  $\alpha \in (0, 1]$ , defined as mixtures of stable or generalized gamma processes. These models offer interpretable parameters, include well-known CRMs as limiting cases, and retain analytical tractability through a tractable Laplace exponent and simple size-biased representation. We analyze the asymptotic properties of this CRM class and apply it to the Caron–Fox framework for sparse graphs. The resulting models produce networks with near-linear edge growth, aligning with empirical evidence from large-scale networks. Additionally, we present efficient algorithms for simulation and posterior inference, demonstrating practical advantages through experiments on real-world sparse network datasets.

## 1 Introduction

Over the past twenty-five years, completely random measures (CRMs) have emerged as fundamental components of modern Bayesian methodology, with applications in clustering, feature modeling, species discovery, survival analysis, finance, and network analysis (Brix, 1999; Regazzini et al., 2003; James, 2002; Nieto-Barajas et al., 2004; Cont and Tankov, 2004; Thibaux and Jordan, 2007; Lijoi et al., 2007; Teh and Gorur, 2009; James et al., 2009; Lijoi and Prünster, 2010; Broderick et al., 2012; Favaro et al., 2013; Cai et al., 2016; Caron and Fox, 2017; Camerlenghi et al., 2019). A key quantity associated with a CRM

$G$  defined on a space  $\Theta$  is its Laplace exponent, the Bernstein function  $\psi : [0, \infty) \rightarrow [0, \infty)$  given by

$$\psi(t) = -\log \mathbb{E} \left[ e^{-tG(\Theta)} \right], \quad t \geq 0. \quad (1)$$

The asymptotic behavior of  $\psi$  as  $t$  grows to infinity determines critical properties of the CRM and of the random structures built upon it. In many cases of practical interest,  $\psi$  exhibits regular variation at infinity:

$$\psi(t) \underset{t \rightarrow \infty}{\sim} t^\alpha \ell(t), \quad (2)$$

for some  $\alpha \in [0, 1]$  and a slowly varying function  $\ell(t)$ , which is typically either a positive constant or a power of  $\log(t)$ .

This asymptotic form is central in determining the growth rate of observable quantities in CRM-based models. For example, in normalized CRM mixture models, latent feature models, or species sampling schemes, the number of clusters, features, or distinct species  $K_n$  scales as  $n^\alpha \ell(n)$  as the sample size  $n \rightarrow \infty$  (Pitman, 2003; Gnedin et al., 2007; Broderick et al., 2012). Similarly, in the Caron–Fox family of sparse graph models, which are based on CRMs (Caron and Fox, 2017), the number of nodes  $N$  scales as  $E^{(1+\alpha)/2} \ell(\sqrt{E})$ , where  $E$  denotes the number of edges (Caron et al., 2023).

The parameter  $\alpha \in [0, 1]$  is referred to as the *index of variation*. The limiting cases  $\alpha = 0$  and  $\alpha = 1$  correspond to *slow* and *rapid* variation respectively (Gnedin et al., 2007). Classical CRMs with favorable analytical properties – such as conjugacy – typically fall in the range  $\alpha \in [0, 1)$ , and include the gamma, beta, stable, generalized gamma, and stable beta processes (Hougaard, 1986; Hjort, 1990; Brix, 1999; Lijoi et al., 2007; Teh and Gorur, 2009); we collect a summary in Table 1.

Despite its importance, the extreme case  $\alpha = 1$ , corresponding to rapid variation, lacks a tractable CRM construction in the literature. In clustering and feature allocation models, it yields regimes where the number of clusters or features grows approximately linearly in  $n$ , modulo a logarithmic correction. Within the Caron–Fox framework for random graphs, the case  $\alpha = 1$  produces networks in which the number of edges grows nearly linearly with the number of nodes—a property widely regarded as more consistent with empirical observations of real-world sparse networks (Van Der Hofstad, 2024, Figure 1.1).

In this article, we introduce a novel class of completely random measures with index of variation  $\alpha \in (0, 1]$ , thereby filling the gap in the existing literature for the rapid variation regime. The proposed CRM is constructed as a mixture of stable, or more generally, generalized gamma CRMs. It features interpretable parameters, includes the stable and generalized gamma processes as limiting cases, and admits a simple and tractable Laplace exponent. Moreover, it supports a size-biased construction that facilitates inference and simulation.

We apply this model to the construction of sparse random graphs. Within the Caron–Fox framework, the resulting graphs exhibit edge growth that is approximately linear in the number of nodes. We develop algorithms for both simulation and posterior inference, leveraging the mixture structure of the model, and demonstrate its empirical validity on two real-world sparse networks.

The remainder of the paper is structured as follows. In Section 2, we introduce our novel CRM family and detail its construction. In Section 3, we study its theoretical properties and describe appropriate simulation methods. In Section 4, we apply our model to network data, describe graph properties, and present an efficient approach to posterior inference. In Section 5, we present numerical experiments on real-world networks, and we gather concluding remarks in Section 6. We defer to the supplementary material

Table 1: Index of variation and slowly varying function associated to some CRMs.

CRM	$\alpha$	$\ell_\alpha$
Gamma	0	log
Generalized gamma	(0,1)	constant
Beta	0	log
Stable beta	(0,1)	constant
Stable	(0,1)	constant
Mixed Stable (ours)	(0,1]	$\log^{-1}$
Mixed Generalized Gamma (ours)	(0,1]	$\log^{-1}$

proofs, implementation details, additional experiments, and some background on regular variation.

**Asymptotic notation.** Throughout this article, we write  $f(x) \underset{x \rightarrow a}{\sim} g(x)$  to denote that  $\lim_{x \rightarrow a} f(x)/g(x) = 1$ , where  $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ . If  $f(x)$  and  $g(x)$  are random variables, the notation indicates that the equivalence holds almost surely. This usage of  $\sim$  should not be confused with the standard distributional notation  $X \sim \mathcal{D}$ , which means that the random variable  $X$  has distribution  $\mathcal{D}$ .

## 2 Rapidly Varying Completely Random Measures

### 2.1 Background on homogeneous CRMs

An homogeneous CRM on some Polish space  $\Theta$  with no deterministic component nor fixed components takes the form (Kingman, 1967, 1993; Lijoi and Prünster, 2010)

$$G = \sum_{j \geq 1} W_j \delta_{\theta_j}, \quad (3)$$

where  $\{(W_j, \theta_j)\}_{j \geq 1}$  are realisations of a Poisson point process on  $(0, \infty) \times \Theta$  with mean measure  $\nu(dw)H(d\theta)$  where  $H$  is some probability distribution on  $\Theta$ , and  $\nu$  is a Lévy measure on  $(0, \infty)$  that satisfies

$$\int_0^\infty (1 - e^{-w})\nu(dw) < \infty. \quad (4)$$

We assume here that  $\nu$  admits a density  $\rho$  with respect to the Lebesgue measure, with  $\nu(dw) = \rho(w)dw$ . We write  $G \sim \text{CRM}(\rho, H)$ . Note that, if  $\int_0^\infty \rho(w)dw = \infty$ , then  $0 < G(\Theta) = \sum_{j \geq 1} W_j < \infty$  almost surely. The Laplace exponent is related to the Lévy intensity  $\rho$  through

$$\psi(t) = \int_0^\infty (1 - e^{-wt})\rho(w)dw. \quad (5)$$

Two standard CRMs are the stable CRM, with Lévy intensity and Laplace exponent

$$\rho_{\text{St}}(w; \alpha) = \frac{\alpha}{\Gamma(1 - \alpha)} w^{-1-\alpha}, \quad \psi_{\text{St}}(t; \alpha) = t^\alpha$$

where  $\alpha \in (0, 1)$  is the index of variation, and the generalized gamma (GG) CRM (also known as exponentially tilted stable CRM), whose Lévy intensity and Laplace exponent

are given by

$$\begin{aligned}\rho_{\text{GG}}(w; \alpha, \beta) &= \frac{\alpha}{\Gamma(1-\alpha)} w^{-1-\alpha} e^{-\beta w}, \\ \psi_{\text{GG}}(t; \alpha, \beta) &= \psi_{\text{St}}(t + \beta; \alpha) - \psi_{\text{St}}(\beta; \alpha) = (t + \beta)^\alpha - \beta^\alpha\end{aligned}$$

where  $\alpha \in (0, 1)$  is the index of variation and  $\beta \geq 0$ . For  $\beta = 0$ , one recovers the stable CRM.

## 2.2 A novel class CRMs with rapid variation

We now introduce a novel class of CRMs whose index of variation  $\alpha \in (0, 1]$ , including the rapid variation case  $\alpha = 1$ . This class of CRMs is defined as a mixture of generalized gamma CRMs, where the mixing is done with respect to the index of variation. We will refer to them as mixed Generalized Gamma (mGG). For ease of presentation, we first present the canonical model, based on a mixture of Stable CRMs (mSt) with parameters  $\alpha$  and  $\tau$ . We then present the full model, with five parameters, whose properties can be derived from those of the canonical model.

**Canonical model: Mixed Stable.** The Lévy intensity of the canonical model is obtained by mixing over the index of variation in a stable model. It is given by

$$\rho_{\text{mSt}}(w; \alpha, \tau) = \frac{1}{\alpha - \tau} \int_{\tau}^{\alpha} \rho_{\text{St}}(w; s) ds = \frac{1}{\alpha - \tau} \int_{\tau}^{\alpha} \frac{s}{\Gamma(1-s)} w^{-1-s} ds \quad (6)$$

where  $0 \leq \tau < \alpha \leq 1$ . Although the intensity  $\rho_{\text{mSt}}$  does not have a closed-form expression, the corresponding Laplace exponent has a remarkably simple expression:

$$\psi_{\text{mSt}}(t; \alpha, \tau) = \frac{1}{\alpha - \tau} \int_{\tau}^{\alpha} \psi_{\text{St}}(t; s) ds = \begin{cases} \frac{t^{\alpha-t\tau}}{(\alpha-\tau)\log t} & t \in (0, 1) \cup (1, \infty), \\ 1 & t = 1, \\ 0 & t = 0. \end{cases} \quad (7)$$

In this model,  $\alpha \in (0, 1]$  is the index of variation, as will be shown in [Proposition 3.1](#). Note that the function  $\psi$  is monotonically increasing and continuous on  $[0, \infty)$ . Additionally, for any  $\alpha \in (0, 1)$ ,

$$\lim_{\tau \rightarrow \alpha} \rho_{\text{mSt}}(w; \alpha, \tau) = \rho_{\text{St}}(w; \alpha), \quad (8)$$

hence the stable model is recovered as a limiting case. Of the values for  $\tau, \alpha$ , we will be particularly interested in the rapid variation case  $\alpha = 1, \tau = 0$ . In this case, the Laplace exponent has a particularly tractable form, given by

$$\psi_{\text{mSt}}(t; 1, 0) = \begin{cases} \frac{t-1}{\log t} & t > 0, t \neq 1, \\ 1 & t = 1, \\ 0 & t = 0. \end{cases} \quad (9)$$

**Full model: Mixed Generalized Gamma.** The full five-parameters model is obtained by scaling and exponentially tilting the density of the mixed Stable CRM. Its Lévy intensity can alternatively be expressed as a mixture of Generalized Gamma CRMs, hence the name mixed Generalized Gamma (mGG). Its Lévy intensity and Laplace exponent are given by

$$\rho_{\text{mGG}}(w; \alpha, \tau, \beta, c, \eta) = \frac{\eta}{c} \rho_{\text{mSt}}\left(\frac{w}{c}; \alpha, \tau\right) e^{-\beta w/c} = \frac{\eta}{\alpha - \tau} \int_{\tau}^{\alpha} \frac{sc^s}{\Gamma(1-s)} w^{-1-s} e^{-\beta w/c} ds, \quad (10)$$

$$\psi_{\text{mGG}}(t; \alpha, \tau, \beta, c, \eta) = \eta (\psi_{\text{mSt}}(\beta + ct; \alpha, \tau) - \psi_{\text{mSt}}(\beta; \alpha, \tau)). \quad (11)$$

The five parameters of the model have the following interpretation:

- $\alpha \in (0, 1]$  is the index of variation;
- $\tau \in [0, \alpha)$  is a power-law exponent that tunes the decay of large weights, see (17);
- $\beta$  is the exponential tilting parameter; if  $\beta > 0$ , it tunes the exponential decay of large weights, see (17);
- $c$  is a scaling parameter: if  $G \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, 1, \eta), H)$ , then  $cG \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta), H)$ ;
- $\eta$  is a rate parameter: if  $G_1 \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta_1), H)$  and  $G_2 \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta_2), H)$ , then  $G_1 + G_2 \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta_1 + \eta_2), H)$ .

### 3 Properties and Simulation

In this section, we conduct a theoretical analysis of the properties of the mGG CRM. Specifically, we prove in Proposition 3.1 that the parameter  $\alpha \in (0, 1]$  in this model is the index of variation, and the model may therefore exhibit rapid variation when  $\alpha = 1$ . All these results will serve as essential tools for our analysis of the asymptotic properties of random graphs based on this model in Section 4.

#### 3.1 Asymptotic properties of the Lévy intensity and Laplace exponent

We derive here some asymptotic properties of the CRM. The behavior of the Laplace exponent, as explained in the introduction, tunes important asymptotic properties related to the behavior of small weights of the CRM. The behaviour of the Lévy intensity at infinity is also of interest for some applications, as it relates to the behavior of large weights.

**Proposition 3.1.** *We have, for the canonical model*

$$\psi_{\text{mSt}}(t; \alpha, \tau) \underset{t \rightarrow \infty}{\sim} \frac{t^\alpha}{(\alpha - \tau) \log t} \quad (12)$$

$$\rho_{\text{mSt}}(w; \alpha, \tau) \underset{w \rightarrow 0}{\sim} \begin{cases} \frac{1}{1-\tau} \frac{w^{-2}}{\log^2(1/w)} & \text{if } \alpha = 1, \\ \frac{\alpha}{(\alpha-\tau)\Gamma(1-\alpha)} \frac{w^{-1-\alpha}}{\log(1/w)} & \text{if } \alpha < 1, \end{cases} \quad (13)$$

$$\rho_{\text{mSt}}(w; \alpha, \tau) \underset{w \rightarrow \infty}{\sim} \begin{cases} \frac{1}{\alpha-\tau} \frac{w^{-1-\tau}}{\log^2(w)} & \text{if } \tau = 0, \\ \frac{\tau}{(\alpha-\tau)\Gamma(1-\tau)} \frac{w^{-1-\tau}}{\log(w)} & \text{if } \tau > 0. \end{cases} \quad (14)$$

For the full model,

$$\psi_{\text{mGG}}(t; \alpha, \tau, \beta, c, \eta) \underset{t \rightarrow \infty}{\sim} \eta c^\alpha \psi_{\text{mSt}}(t; \alpha, \tau), \quad (15)$$

$$\rho_{\text{mGG}}(w; \alpha, \tau, \beta, c, \eta) \underset{w \rightarrow 0}{\sim} \eta c^\alpha \rho_{\text{mSt}}(w; \alpha, \tau), \quad (16)$$

$$\rho_{\text{mGG}}(w; \alpha, \tau, \beta, c, \eta) \underset{w \rightarrow \infty}{\sim} \eta c^\tau e^{-\beta w/c} \rho_{\text{mSt}}(w; \alpha, \tau). \quad (17)$$

### 3.2 Moments

Derivatives of the Laplace exponent  $\psi$  of a CRM are of interest to derive moments or other properties. For  $n \geq 1$  and  $t > 0$ , let

$$\kappa(n, t) = (-1)^{n+1} \psi^{(n)}(t) = \int_0^\infty w^n e^{-wt} \rho(w) dw. \quad (18)$$

If  $G \sim \text{CRM}(\rho, H)$ , then  $\mathbb{E}[G(\Theta)] = \kappa(1, 0)$  and  $\mathbb{V}[G(\Theta)] = \kappa(2, 0)$ . We provide below the expression of the  $\kappa$  function for the mGG CRM, and analytic expressions for the expectation and variance of  $G(\Theta)$  under this model.

**Proposition 3.2.** *Let  $m \geq 1$  and  $z \geq 0$ . Define*

$$\kappa_{\text{mGG}}(m, z; \alpha, \tau, \beta, c, \eta) = \int_0^\infty w^m e^{-zw} \rho(w; \alpha, \tau, \beta, c, \eta) dw$$

and  $\kappa_{\text{mSt}}(m, z; \alpha, \tau) = \int_0^\infty w^m e^{-zw} \rho_{\text{mSt}}(w; \alpha, \tau) dw$ . Then

$$\begin{aligned} \kappa_{\text{mSt}}(m, z; \alpha, \tau) &= \frac{z^{-m}}{\alpha - \tau} \int_\tau^\alpha s z^s \frac{\Gamma(m-s)}{\Gamma(1-s)} ds, \\ \kappa_{\text{mGG}}(m, z; \alpha, \tau, \beta, c, \eta) &= \eta c^m \kappa_{\text{mSt}}(m, \beta + cz; \alpha, \tau). \end{aligned}$$

For  $m = 1$ ,  $z > 0$ , we have

$$\kappa_{\text{mSt}}(1, z; \alpha, \tau) = \begin{cases} \frac{z^{-1}}{\alpha - \tau} \frac{z^\tau - z^\alpha + (\alpha z^\alpha - \tau z^\tau) \log z}{(\log z)^2} & z \neq 1, \\ \frac{\alpha + \tau}{2} & z = 1. \end{cases}$$

If  $G$  is a CRM with parameters  $\alpha, \tau, \beta > 0, c, \eta$ , then the total mass  $G(\Theta) = \sum_{j \geq 1} W_j$  has expectation  $\mathbb{E}[G(\Theta)] = \eta c \kappa_{\text{mSt}}(1, \beta; \alpha, \tau)$  and variance  $\mathbb{V}[G(\Theta)] = \eta c^2 \kappa_{\text{mSt}}(2, \beta; \alpha, \tau)$ . In particular, if  $\alpha = 1$  and  $\tau = 0$ , we obtain

$$\begin{aligned} \mathbb{E}[G(\Theta)] &= \begin{cases} \eta c \frac{1 - \beta + \beta \log(\beta)}{\beta \log^2(\beta)} & \beta \neq 1, \\ \frac{\eta c}{2} & \beta = 1, \end{cases} \\ \mathbb{V}[G(\Theta)] &= \begin{cases} \eta c^2 \frac{(\beta+1) \log \beta + 2(1-\beta)}{\beta^2 \log^3 \beta} & \beta \neq 1, \\ \frac{\eta c^2}{6} & \beta = 1. \end{cases} \end{aligned}$$

### 3.3 Simulation of the weights and total mass

Among the key properties expected from a tractable statistical model is the ability to sample from it. However, since CRMs are infinite objects that cannot be fully represented on a finite computer, only approximate sampling is possible. Several approaches for approximating CRM sampling exist, and here we propose a size-biased method, which is advantageous due to its ease of implementation. Additionally, we provide an estimate for the total mass lost in the approximation, which is proportional to  $1/\log(n)$ , where  $n$  is the number of size-biased samples. In certain applications, only the total mass of the CRM is required. For such cases, we introduce an approximate method for sampling the total mass based on a Riemann sum, which converges at a rate of  $1/n$ . This convergence rate outperforms the alternative approach of computing individual weights and summing them.

### 3.3.1 Size-biased sampling of the CRM

We follow here the strategy of [Perman et al. \(1992\)](#) to sample the size-biased weights, see [Appendix S6](#) in the supplementary material for details.

**Proposition 3.3.** *Let  $\xi_1, \xi_2, \dots$  be the ordered points of a unit-rate Poisson process on  $(0, \infty)$ ; that is,  $\xi_1, \xi_2 - \xi_1, \xi_3 - \xi_2, \dots$  are iid unit-rate exponential random variable. We have the following size-biased construction for  $G \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta), H)$ . For  $j \geq 1$ ,*

$$\begin{aligned} T_j &= \psi_{\text{mSt}}^{-1} \left( \frac{\xi_j}{\eta} + \psi_{\text{mSt}}(\beta; \alpha, \tau); \alpha, \tau \right) - \beta, \\ S_j | \{T_j = t\} &\sim p(s|t) \propto s(t + \beta)^s \mathbf{1}_{\{s \in (\tau, \alpha)\}}, \\ W'_j | \{T_j = t, S_j = s\} &\sim \text{Gamma}(1 - s, t + \beta), \\ W_j &= cW'_j, \end{aligned}$$

where  $\psi_{\text{mSt}}^{-1}$  is the inverse of  $\psi_{\text{mSt}}$ .

One can sample exactly from  $p(s|t)$  using the inverse CDF, which has a tractable expression in terms of standard functions, see [Proposition S2](#) in the supplementary material. When  $\tau = 0$ , the inverse  $\psi_{\text{mSt}}^{-1}$  can be expressed using standard functions, see [Proposition S3](#). The following proposition describes the asymptotic L1 error of the size-biased approximation. Its proof, similar to that in [Lee et al. \(2023\)](#), is given in [Appendix S1](#).

**Proposition 3.4.** *Let  $\alpha = 1$  and  $\tau = 0$ . Let  $(W_j)_{j \geq 1}$  be the sequence of weights sampled via the size-biased algorithm in [Proposition 3.3](#). Then*

$$R_n = \sum_{j > n}^{\infty} W_j \underset{n \rightarrow \infty}{\sim} \frac{c\eta}{\log(n)} \quad \text{almost surely.}$$

### 3.3.2 Approximate sampling of the total mass

One can use the mixture representation of the CRM in order to simulate random variables that are approximately distributed as  $G(\Theta) = \sum_{j \geq 1} W_j$ . The approximation relies on a weighted sum of exponentially tilted positive stable random variables, for which efficient simulation algorithms exist ([Devroye, 2009](#); [Hofert, 2011](#)).

**Proposition 3.5.** *Let  $n \geq 1$ . For  $i = 1, \dots, n$  let  $s_i = \tau + \frac{(\alpha - \tau)(i - 1)}{n}$  and let  $X_i$  be an exponentially tilted stable random variables, with stable parameter  $s_i$  and exponential tilting parameter  $\beta_i = \left(\frac{\eta}{n}\right)^{\frac{1}{s_i}} \beta$ . That is, for any  $t \geq 0$ ,*

$$\mathbb{E}[e^{-tX_i}] = \exp(-((t + \beta_i)^{s_i} - \beta_i^{s_i})).$$

Then

$$S_n = c \sum_{i=1}^n \left(\frac{\eta}{n}\right)^{\frac{1}{s_i}} X_i \rightarrow G(\Theta)$$

in distribution as  $n \rightarrow \infty$ .

## 4 Sparse network modelling

We consider in this section the use of the mGG CRM on  $\Theta = \mathbb{R}_+$ , with  $\alpha = 1$ ,  $\tau = 0$ ,  $\beta > 0$ ,  $c > 0$ ,  $\eta > 0$ , within the Caron-Fox graph construction in order to obtain extremely sparse sequences of graph. We describe the model, its asymptotic properties, and a Markov chain Monte Carlo method for posterior inference of the model parameters.

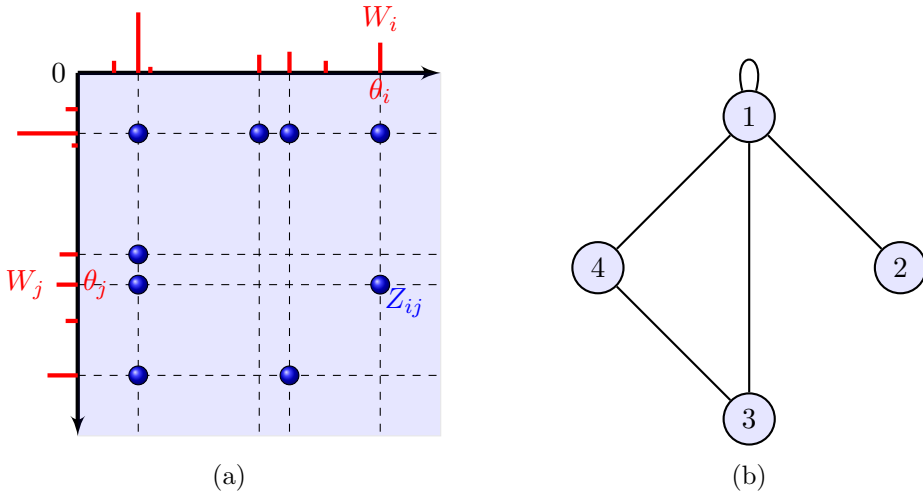


Figure 1: Point process representation of a random graph. Each node  $i$  is embedded in  $\mathbb{R}_+$  at some location  $\theta_i$  and is associated with a sociability parameter  $W_i$ . An edge between nodes  $\theta_i$  and  $\theta_j$  is represented by a point at locations  $(\theta_i, \theta_j)$  and  $(\theta_j, \theta_i)$  in  $\mathbb{R}_+^2$  (Figure 1, Caron and Fox, 2017). The corresponding graph is plotted in (b).

#### 4.1 mGG graph model

Caron and Fox (2017) introduced a novel class of graph models based on CRMs which falls within the broader framework of graphex processes – a generalisation of dense graphons to the sparse setting (Caron and Fox, 2017; Veitch and Roy, 2015; Borgs et al., 2018). For a CRM

$$G = \sum_{j \geq 1} W_j \delta_{\theta_j} \sim \text{CRM}(\rho_{\text{mGG}}(\cdot; \alpha, \tau, \beta, c, \eta), \lambda),$$

where  $\lambda$  is the Lebesgue measure, we consider the atomic measure on  $\mathbb{R}_+^2$

$$U = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} Z_{i,j} \delta_{(\theta_i, \theta_j)}, \quad (19)$$

where the binary random variables  $Z_{ij} \in \{0, 1\}$  indicate if there is an edge between node  $i$  and node  $j$ . They are defined, for  $i < j$ , by

$$Z_{ij} \mid (W_k)_{k \geq 1} \sim \text{Bernoulli}(1 - e^{-2W_i W_j}),$$

$Z_{ji} = Z_{ij}$ , and  $Z_{ii} \mid (W_k)_{k \geq 1} \sim \text{Bernoulli}(1 - e^{-W_i^2})$ . The weight  $W_i$  may be interpreted as a sociability parameter of node  $i$ ; the larger  $W_i$ , the more likely node  $i$  is to connect to other nodes. From  $Z$  we derive a growing family of graphs  $(\mathcal{G}_t)_{t \geq 0}$ , where  $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$  is the graph of size  $t$ , with vertex set  $\mathcal{V}_t$  and edge set  $\mathcal{E}_t$ , defined as

$$\begin{aligned} \mathcal{V}_t &= \{\theta_i \mid \theta_i \leq t \text{ and } \exists \theta_k \leq t \text{ s.t. } Z_{ik} = 1\}, \\ \mathcal{E}_t &= \{\{\theta_i, \theta_j\} \mid t, \theta_j \leq t \text{ and } Z_{ij} = 1\}. \end{aligned}$$

## 4.2 Properties

For a graph  $\mathcal{G}_t$ , let  $N_t = |\mathcal{V}_t|$  be the number of nodes,  $N_t^{(e)} = |\mathcal{E}_t|$  the number of edges and  $N_{t,j}$  the number of nodes of degree  $j$ ,  $j \geq 1$ .

The following sparsity results follow from [Caron et al. \(2023, Proposition 11\)](#). It states that the graphs are extremely sparse, with the number of edges scaling approximately linearly (up to a log factor) with respect to the number of nodes. [Figure 3](#) shows a random graph drawn from the model compared to a (dense) Erdős–Rényi graph.

**Proposition 4.1** (Asymptotic number of nodes and edges). *We have almost surely that*

$$N_t \underset{t \rightarrow \infty}{\sim} t^2 \frac{C}{\log(t)}, \quad N_t^{(e)} \underset{t \rightarrow \infty}{\sim} \frac{t^2}{2} \overline{W} \quad (20)$$

where  $C = (\eta c)^2$  if  $\beta = 1$  and  $C = 2(\eta c)^2 \frac{\beta^{-1} - 1 + \log \beta}{(\log \beta)^2}$  otherwise, and

$$\overline{W} = \int_0^\infty \psi_{\text{mGG}}(2w; 1, 0, \beta, c, \eta) \rho_{\text{mGG}}(w; 1, 0, \beta, c, \eta) dw.$$

**Corollary 4.2** (Extreme sparsity). *We have almost surely:*

$$N_t^{(e)} \underset{t \rightarrow \infty}{\sim} \frac{\overline{W}}{4C} N_t \log(N_t).$$

Hence  $\frac{N_t^{(e)}}{N_t} \rightarrow \infty$  and for any  $\epsilon > 0$  we have  $\frac{N_t^{(e)}}{N_t^{1+\epsilon}} \rightarrow 0$ . In other words the graph is almost extremely sparse.

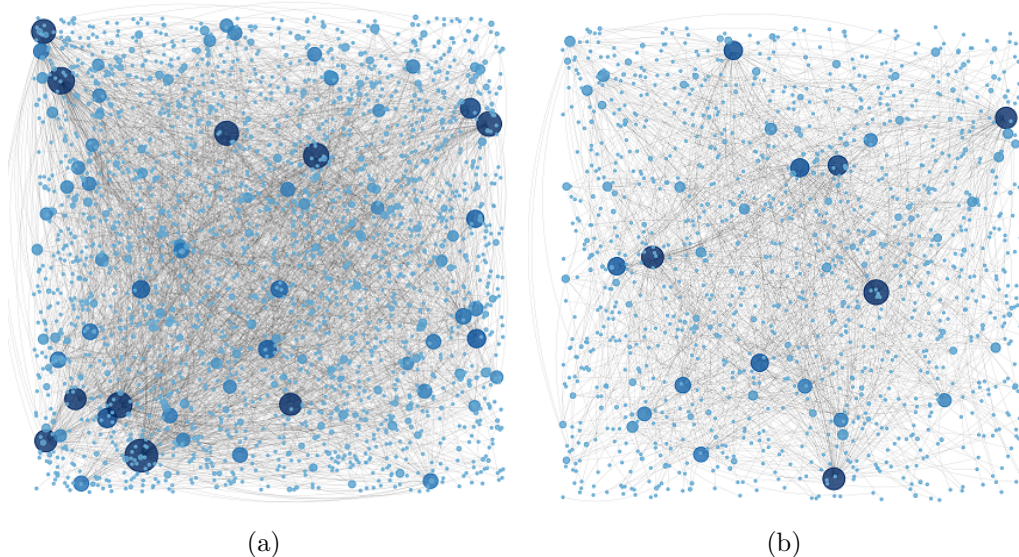


Figure 2: Sampled mGG graphs (a) with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $\beta = 1$ ,  $c = 1$ , and  $\eta = 100$ , 2122 nodes, and 2583 edges; (b) with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $\beta = 1.5$ ,  $c = 5$ , and  $\eta = 20$ , 1249 nodes, and 1366 edges. The node size is proportional to its degree. The graphs were generated using our size-biased method and displayed using NetworkX.

Another consequence of ([Caron et al., 2023, Proposition 11](#)) is the asymptotic degree distribution.

**Proposition 4.3** (Asymptotic degree distribution). *We have that, almost surely,*

$$N_{t,j} \underset{t \rightarrow \infty}{\sim} \begin{cases} t^2 \frac{C}{\log(t)} & j = 1, \\ \frac{t^2}{j(j-1)} \frac{C}{\log^2(t)} & j \geq 2. \end{cases}$$

Consequently, the nodes of degree 1 dominate in the graphs as  $N_{t,j} = o(N_{t,1})$  for all  $j \geq 2$ . Let  $\tilde{N}_{t,2} = \sum_{k \geq 2} N_{t,k}$  be the number of nodes of degree at least 2. Then, almost surely for all  $j \geq 2$

$$\frac{N_{t,j}}{\tilde{N}_{t,2}} \underset{t \rightarrow \infty}{\rightarrow} \frac{1}{j(j-1)}. \quad (21)$$

This corresponds, for nodes of degree larger than 2, to a degree distribution with power-law of exponent 2.

In Figure 3 we compare our methods to the Generalized Gamma without mixture of Caron and Fox (2017) and with the Barabási–Albert model (Barabási and Albert, 1999). All these models are known to lead to sparse sequences of graphs presenting power law degree distribution. The expected asymptotic behaviour of our model is given in Corollary 4.2 and (21). For the Generalized Gamma, here with  $\alpha = 0.5$ , we have asymptotically  $N_t^{(e)} \underset{t \rightarrow \infty}{\sim} N_t^{\frac{4}{3}}$  and a degree distribution following a power law of exponent 1.5. For the Barabási–Albert model, the number of edges is linear in the number of nodes and presents an asymptotic degree distribution following a power law of exponent 3.

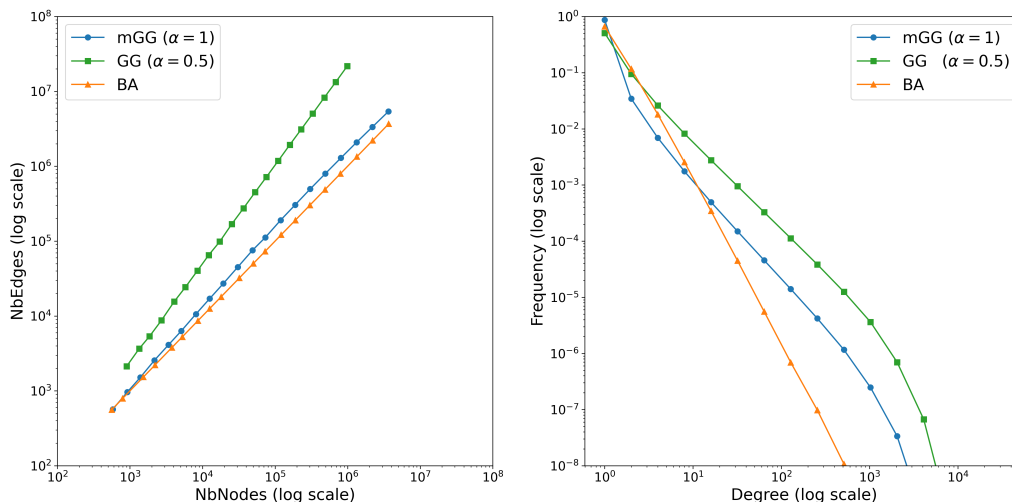


Figure 3: Examination of the mGG graph properties (●) with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $c = 1$ , and  $\beta = 1$  for various values of  $\eta$  ranging from 50 to 6000, resulting in graphs of different sizes. Comparison with the Generalized Gamma CRM (■) with parameters  $\tau = 1$  and  $\sigma = 0.5$ , and with the Barabási–Albert model (▲). For every configuration we simulate 20 graph samples and plot the mean of the quantity of interest.

### 4.3 Posterior inference

Here we briefly describe the MCMC algorithm to approximate the posterior density over the model parameters. Assume that we have observed a graph  $\mathcal{G}$ , which we assume was sampled from a mGG graph model. We aim to infer the sociability parameter  $W_i$  of each

---

**Algorithm 1** Posterior inference

---

**Step 1:** update the weights  $w_i$  and the latent space variables  $s_i$  given the rest, using a Hamiltonian Monte Carlo step.

**Step 2:** update  $(\phi, w_*)$  given the rest, using a Gibbs sampler update.

**Step 3:** update the latent count  $(\tilde{q}_{ij})$  given the rest using (23).

---

nodes and the parameters  $\phi = (\beta, c, \eta)$  of the mGG model. We consider the following improper priors:

$$p(\beta) \propto \frac{1}{\beta}, \quad p(c) \propto \frac{1}{c}, \quad p(\eta) \propto \frac{1}{\eta}. \quad (22)$$

We introduced two sets of auxiliary variables. First, as in Caron and Fox (2017), we introduce latent count variables  $\tilde{q}_{ij}$  (an interpretation of this  $\tilde{q}_{ij}$  in term of directed multigraph is given in Appendix S2.1), with conditional distribution, for  $i < j$ ,

$$\tilde{q}_{ij} \mid Z, W \sim \begin{cases} \delta_0 & \text{if } Z_{ij} = 0, \\ \text{tPoisson}(2W_i W_j) & \text{if } Z_{ij} = 1, i \neq j, \\ \text{tPoisson}(W_i^2) & \text{if } Z_{ij} = 1, i = j, \end{cases} \quad (23)$$

where  $\text{tPoisson}(\lambda)$  is the zero-truncated Poisson distribution, and with  $\tilde{q}_{ji} = \tilde{q}_{ij}$ . Second, to take advantage of the mixture representation of the Lévy intensity, we introduce, as in the size-biased algorithm, local indices of variations  $S_i \in (0, 1)$  for each node  $i$ . In the following, we also write  $w_*$  for the sum of the sociabilities of nodes with no connection in  $\mathcal{G}$ . The algorithm is presented in Algorithm 1.

## 5 Numerical experiments on networks

### 5.1 Synthetic data

We first examine the convergence of the MCMC algorithm on simulated data, where the graph is generated from the mGG graph model described in Section 4.1. We simulate a mGG graph with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $\beta = 1$ ,  $c = 2$  and  $\eta = 130$  using size-biased sampling with  $N = 10^5$  weights. The resulting graph contains 11,613 nodes and 17,427 edges. We run three MCMC chains, each consisting of 6 million iterations and initialized from different starting values. Half of the iterations are discarded as burn-in. The hyperparameters  $\alpha$  and  $\tau$  are fixed to their true values, while the remaining parameters are estimated. Details of the MCMC parameterization are provided in Appendix S3.2.

Trace plots for the parameters  $\beta$ ,  $c$ , and  $\eta$  are shown in Figure 4, while trace plots for four of the 11,613 weights are presented in Figure 5. To further illustrate convergence of the weights (or sociability parameters)  $w_i$ , we also include in Figure 4(d) the trace plot of the total sum of the weights. In addition, Figure 6 displays posterior credible intervals for the sociability parameters  $w_i$  of the 50 nodes with the highest degrees, as well as for the log-sociability parameters  $\log(w_i)$  of the 50 nodes with the lowest degrees. These results highlight the ability of our method to accurately recover sociability parameters for both high- and low-degree nodes.

We use our sampling method to simulate 200 graphs from the posterior predictive distribution. Figure 7(a) compares the degree distribution of a graph generated using the true parameter values with those of the simulated graphs. We then repeat the procedure with  $\eta$  rescaled by a factor of 2: we generate a new reference graph using the true parameters

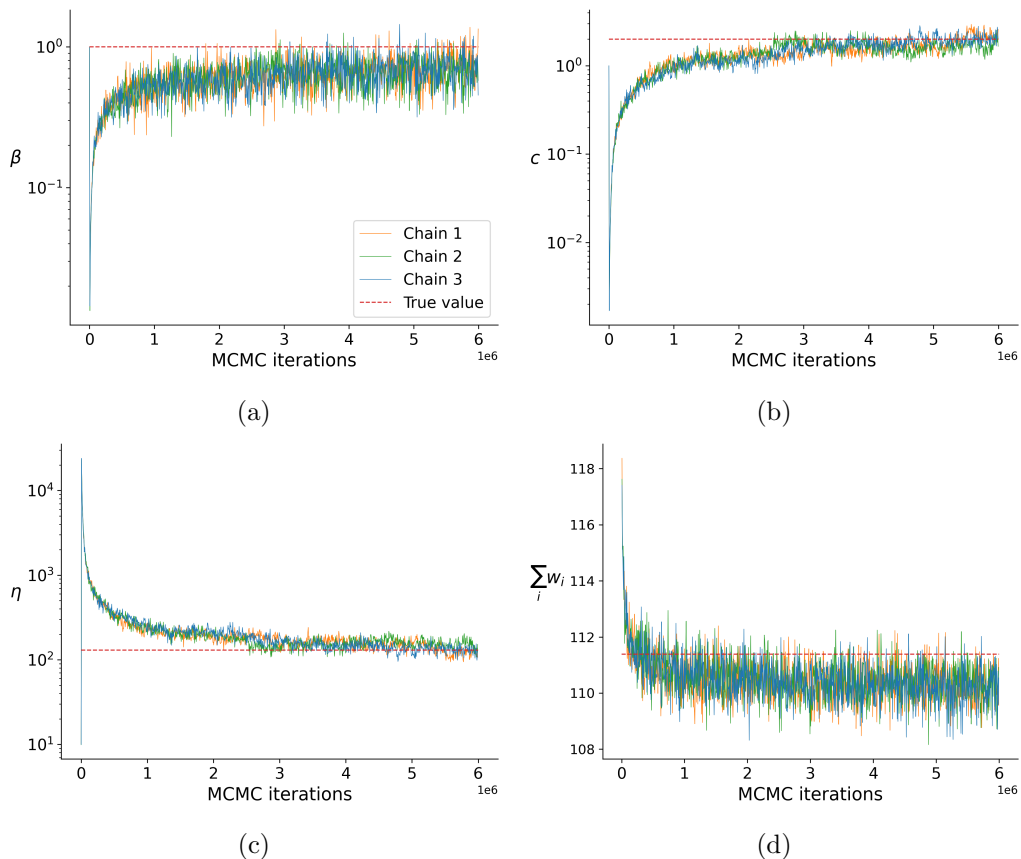


Figure 4: MCMC traceplot of parameters (a)  $\beta$ , (b)  $c$ , (c)  $\eta$  and (d) the total sum of the weights  $w_i$  for a mGG graph with model parameters  $\alpha = 1, \tau = 0, \beta = 1, c = 2$  and  $\eta = 130$ . Three chains are displayed and the true value is in red dashed line.

but with  $\eta' = 2\eta$ , and simulate 200 posterior predictive graphs using the same MCMC run, but with the rescaled  $\hat{\eta}'$ . The resulting degree distributions are again compared in Figure 7(b).

We compute the Gelman–Rubin convergence diagnostic (Gelman and Rubin, 1992) for the four key parameters, all of which yield values below 1.05 (see Table S4 for exact figures). To assess convergence across all parameters –  $\beta$ ,  $c$ ,  $\eta$ ,  $(w_i)$ , and  $(s_i)$  – we use the multivariate Gelman–Rubin diagnostic introduced by Vats and Knudson (2021), see (S37), obtaining a value of 1.001, which strongly suggests convergence.

## 5.2 Real World Data

We now evaluate our method on real-world networks. We consider three datasets, all consisting of undirected graphs, downloaded from Kunegis (2013) and Rossi and Ahmed (2015).

- **Flickr:** A crawl of Flickr, a photo and video sharing platform. The dataset contains all links between users. (<https://socialnetworks.mpi-sws.org/data-imc2007.html>, Mislove et al., 2007).
- **Douban:** A crawl of Douban, an online social network that provides user reviews and recommendations for movies, books, and music. The dataset contains all links

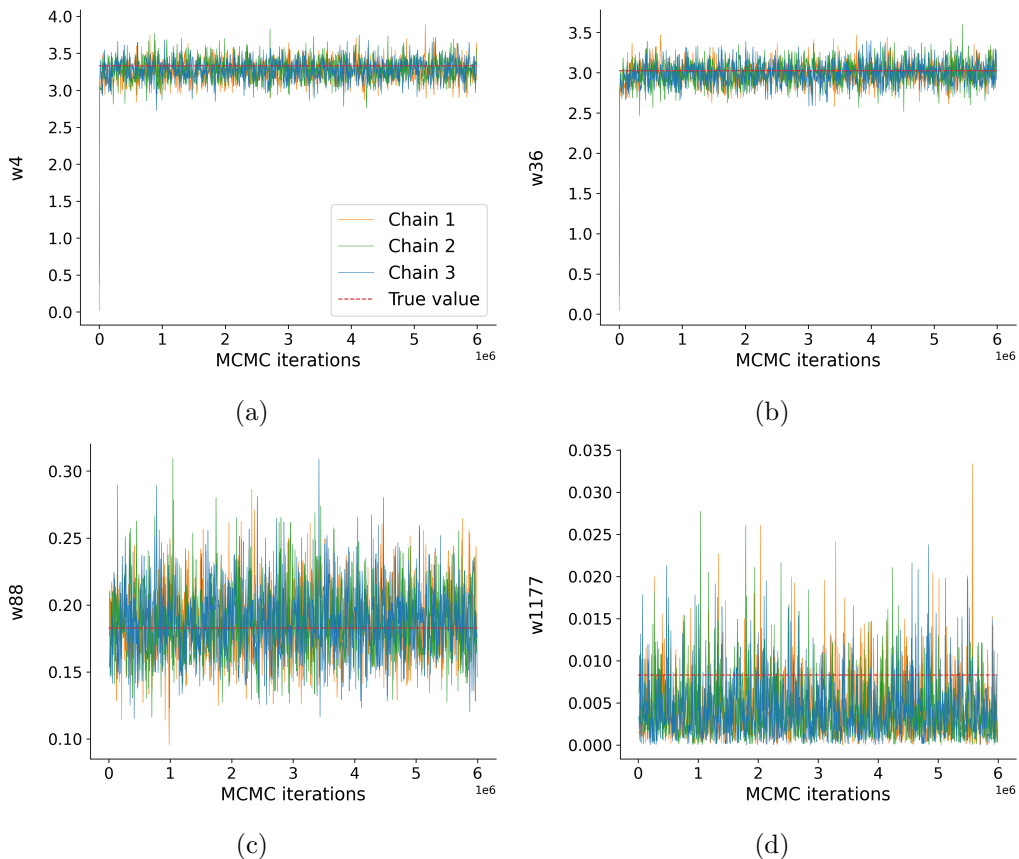


Figure 5: MCMC traceplot of four weights for a graph generated from a mGG graph model with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $\beta = 1$ ,  $c = 2$  and  $\eta = 130$ . Three chains are displayed and the true value is in red dashed line. The degree of the corresponding node is (a) 536, (b) 496, (c) 49 and (d) 2.

between users (<https://networkrepository.com/soc-douban.php>, Zafarani and Liu, 2014).

- **TwitterCrawl:** A crawl of Twitter, where nodes represent users and edges correspond to retweets collected from various social and political hashtags (<https://networkrepository.com/rt-retweet-crawl.php>, Rossi et al., 2014).

The sizes of the datasets are shown in Table 2. To train our model we extracted a subgraph containing approximately 5% of the nodes from each network using  $p$ -sampling (Veitch and Roy, 2019). The corresponding values of  $p$  and subgraph statistics are reported in Table 3. Further details and additional plots are provided in Appendix S3.3.

Table 2: Sizes of the real-world datasets

Dataset	Nodes	Edges	Max degree	Mean degree
Flickr	1,861,232	155,55,041	54,472	33.4
Douban	154,908	327,162	574	8.4
TwitterCrawl	1,112,702	2,278,852	10,140	8.2

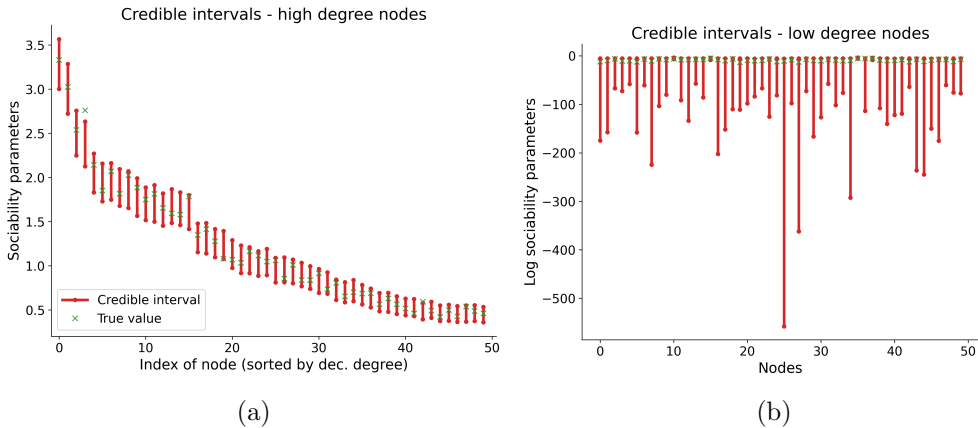


Figure 6: 95% posterior intervals of (a) the sociability parameters  $w_i$  of the 50 nodes with highest degree and (b) the log-sociability parameters  $\log(w_i)$  of the 50 nodes with lowest degree, for a sample from a mGG graph model with parameters  $\alpha = 1, \tau = 0, \beta = 1, c = 2$  and  $\eta = 130$ . The true values are in green.

Table 3: Sizes of the subgraphs used for inference

Dataset	Nodes	Edges	Max degree	Mean degree	p
Flickr	85,613	299,358	2890	13.9	0.14
Douban	7775	9586	82	4.9	0.17
TwitterCrawl	70,944	87,914	618	5.0	0.2

To assess model fit, we apply  $p$ -sampling to split each dataset into a training set (approximately 5% of the nodes) and a test set (the remaining nodes). We then compare the empirical degree distributions of the test sets with the predictive degree distributions generated by our model, using hyperparameters inferred from the training data. Specifically, we follow the procedure described in [Section 3.3](#) to simulate 200 graphs from the posterior predictive distribution for each dataset. Parameters are drawn from the post-burn-in MCMC samples, and we scale  $\hat{\eta}$  by  $(1 - p)/p$  to match the expected size of the target graph.

[Figure 8](#) shows the resulting comparisons between the empirical and predictive degree distributions. Overall, the model provides a good fit across the datasets, successfully capturing the heavy-tailed behavior characteristic of real-world networks.

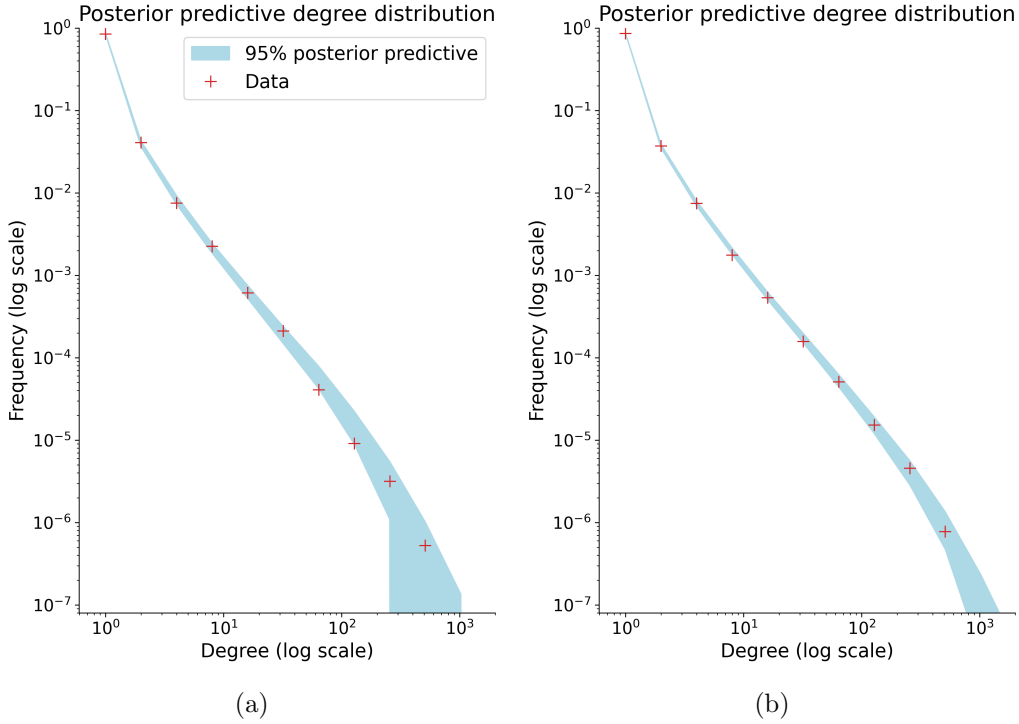


Figure 7: (a) Posterior predictive degree distributions from our MCMC run, compared to those of a graph—distinct from the training graph—generated using the true parameters. (b) Same comparison, but with a reference graph that is twice as large; the corresponding  $\hat{\eta}$  values from the MCMC run have been rescaled accordingly.

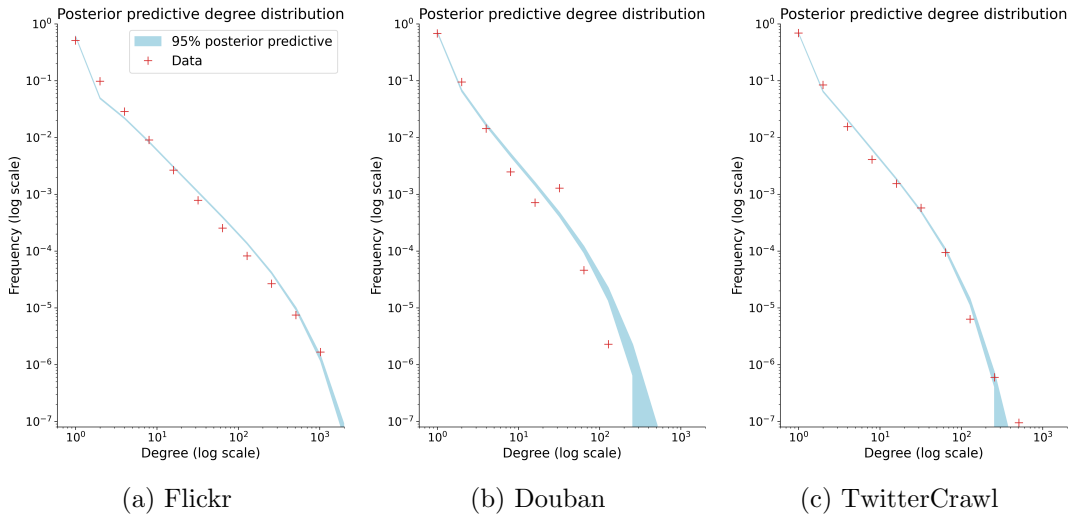


Figure 8: Posterior predictive degree distributions for the real-world datasets. The predictive distribution is shown in blue, and the empirical distribution of the test set is shown in red.

## 6 Discussion

In this work, we propose a novel class of CRMs by mixing the Lévy intensity of a stable or GG CRM over the stability parameter. The use of mixtures of existing Lévy measures to obtain novel univariate or multivariate CRMs has a long history (Cont and Tankov, 2004; Griffin and Leisen, 2017, 2018; Ayed et al., 2019; Todeschini et al., 2020; Naik et al., 2021; Ayed et al., 2024). A key novelty here is that the mixing is performed with respect to the index of variation of the base model, in order to obtain the desired rapid varying properties. The Lévy intensities we consider have the form

$$\int_0^1 g(s)h(w)w^{-1-s}ds, \quad (24)$$

where  $g(s) = \frac{s}{\Gamma(1-s)}$ ,  $h(w) = 1$  for mST (6) and  $g(s) = \frac{\eta sc^s}{\Gamma(1-s)}$ ,  $h(w) = e^{-\beta w/c}$  for mGG (10). In both cases,

$$g(s) \underset{s \rightarrow 1}{\sim} C_1(1-s), \quad h(w) \underset{w \rightarrow 0}{\sim} C_2 \quad (25)$$

for some constants  $C_1$  and  $C_2$ . These asymptotics imply, via Karamata’s theorem, the behavior of the Laplace exponent at  $\infty$  and Lévy intensity at 0 described in Proposition 3.1, and lead to extreme sparsity in the associated random graph. The specific forms of  $g$  and  $h$  were chosen to balance flexibility with analytical tractability: the Laplace exponents (7) and (11) and associated moments admit simple expressions. This tractability also enables implementation of the size-biased sampling algorithm presented in Section 3.3.1 without requiring numerical integration. More broadly, any mixture of the form in (24) where  $g$  and  $h$  satisfy the asymptotics in (25) could be used. For example, one can define a rapidly varying version of the stable beta process (Teh and Gorur, 2009) by taking  $g(s) = \frac{\eta s}{\Gamma(1-s)}$  and  $h(w) = (1-w)^{\xi-1}1_{w \in (0,1)}$  for some  $\eta, \xi > 0$ .

While we have focused on the application to sparse graph modeling, other domains are of significant interest. For instance, in clustering or partition models, such CRMs would yield random partitions with the number of clusters  $K_n$  scaling as  $n/\log(n)$ . This extreme behavior may be desirable in problems such as entity resolution and deduplication (Betancourt et al., 2016; Di Benedetto et al., 2021; Betancourt et al., 2022), where most records correspond to unique individuals and thus  $K_n$  is expected to grow nearly linearly with  $n$ .

In Section 4, we introduced a class of extremely sparse random graph models based on the graphex framework. Another family of models – edge-exchangeable networks (Crane and Dempsey, 2018; Cai et al., 2016; Janson, 2018) – also employs CRMs as key components. Incorporating the CRMs developed in this paper within such models yields edge-exchangeable multigraphs whose number of edges also scales approximately linearly (up to a logarithmic factor) with the number of nodes.

## Bibliography

- Ayed, F., Lee, J., and Caron, F. (2019). Beyond the Chinese Restaurant and Pitman-Yor processes: Statistical Models with Double Power-law Behavior. In *International Conference on Machine Learning*, volume 97, pages 395–404.
- Ayed, F., Lee, J., and Caron, F. (2024). The Normal-Generalised Gamma-Pareto Process: A Novel Pure-Jump Lévy Process with Flexible Tail and Jump-Activity Properties. *Bayesian Analysis*, 19(1):123–152.

- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Betancourt, B., Zanella, G., Miller, J., Wallach, H., Zaidi, A., and Steorts, R. (2016). Flexible models for microclustering with application to entity resolution. *Advances in Neural Information Processing Systems*, 29.
- Betancourt, B., Zanella, G., and Steorts, R. C. (2022). Random Partition Models for Microclustering Tasks. *Journal of the American Statistical Association*, 117(539):1215–1227.
- Borgs, C., Chayes, J. T., Cohn, H., and Holden, N. (2018). Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research*, 18(210):1–71.
- Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, pages 929–953.
- Broderick, T., Jordan, M. I., and Pitman, J. (2012). Beta processes, stick-breaking and power laws. *Bayesian analysis*, 7(2):439–476.
- Cai, D., Campbell, T., and Broderick, T. (2016). Edge-exchangeable graphs and sparsity. *Advances in Neural Information Processing Systems*, 29.
- Camerlenghi, F., Dunson, D. B., Lijoi, A., Prünster, I., Rodríguez, A., et al. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Analysis*, 14(4):1303–1356.
- Caron, F. and Fox, E. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79:1295–1366.
- Caron, F., Panero, F., and Rousseau, J. (2023). On sparsity, power-law, and clustering properties of graphex processes. *Advances in Applied Probability*, 55(4):1211–1253.
- Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*, volume 2. CRC press.
- Crane, H. and Dempsey, W. (2018). Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, 113:1311–1326.
- Devroye, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation*, 19(4):1–20.
- Di Benedetto, G., Caron, F., and Teh, Y. W. (2021). Nonexchangeable random partition models for microclustering. *The Annals of Statistics*, 49(4):1931–1957.
- Favaro, S., Lijoi, A., and Prünster, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *The Annals of Applied Probability*, 23(5):1721–1754.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.
- Gnedin, A., Hansen, B., and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: General asymptotics and power laws. *Probab. Surv*, 4(146-171):88.
- Griffin, J. and Leisen, F. (2018). Modelling and Computation Using NCoRM Mixtures for Density Regression. *Bayesian Analysis*, 13(3):897–916.

- Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(2):525–545.
- Hjort, N. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294.
- Hofert, M. (2011). Sampling exponentially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):3.
- Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396.
- James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. *arXiv preprint math/0205093*.
- James, L. F., Lijoi, A., and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, 36(1):76–97.
- Janson, S. (2018). On Edge Exchangeable Random Graphs. *Journal of Statistical Physics*, 173(3):448–484.
- Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.
- Kingman, J. (1993). *Poisson Processes*, volume 3. Oxford University Press, USA.
- Kunegis, J. (2013). KONECT – The Koblenz Network Collection. In *Proc. Int. Conf. on World Wide Web Companion*, pages 1343–1350.
- Lee, J., Miscouridou, X., and Caron, F. (2023). A unified construction for series representations and finite approximations of completely random measures. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 29(3):2142–2166.
- Lijoi, A., Mena, R. H., and Prünster, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):715–740.
- Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G., editors, *Bayesian Nonparametrics*. Cambridge University Press.
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 29–42, San Diego California USA. ACM.
- Naik, C., Caron, F., and Rousseau, J. (2021). Sparse networks with core-periphery structure. *Electronic Journal of Statistics*, 15(1):1814–1868.
- Nieto-Barajas, L. E., Prünster, I., and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics*, 32(6):2343–2360.
- Perman, M., Pitman, J., and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39.

- Pitman, J. (2003). Poisson-kingman partitions. In Goldstein, D. R., editor, *Statistics and Science: A Festschrift for Terry Speed*, volume Volume 40 of *Lecture Notes–Monograph Series*, pages 1–34. Institute of Mathematical Statistics, Beachwood, OH.
- Regazzini, E., Lijoi, A., and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31(2):560–585.
- Rossi, R. A. and Ahmed, N. K. (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.
- Rossi, R. A., Gleich, D. F., Gebremedhin, A. H., and Patwary, M. M. A. (2014). Fast maximum clique algorithms for large graphs. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pages 365–366, New York, NY, USA. Association for Computing Machinery.
- Teh, Y. and Gorur, D. (2009). Indian buffet processes with power-law behavior. *Advances in Neural Information Processing systems*, 22.
- Thibaux, R. and Jordan, M. I. (2007). Hierarchical beta processes and the Indian buffet process. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS'07)*, pages 564–571.
- Todeschini, A., Miscouridou, X., and Caron, F. (2020). Exchangeable random measures for sparse and modular graphs with overlapping communities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2):487–520.
- Van Der Hofstad, R. (2024). *Random Graphs and Complex Networks*, volume 2. Cambridge University Press, 1 edition.
- Vats, D. and Knudson, C. (2021). Revisiting the Gelman–Rubin Diagnostic. *Statistical Science*, 36(4):518–529.
- Veitch, V. and Roy, D. M. (2015). The Class of Random Graphs Arising from Exchangeable Random Measures. *arXiv:1512.03099*.
- Veitch, V. and Roy, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *The Annals of Statistics*, 47(6):3274 – 3299.
- Zafarani, R. and Liu, H. (2014). Users Joining Multiple Sites: Distributions and Patterns. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):635–638.

# Rapidly Varying Completely Random Measures for Modeling Extremely Sparse Networks Supplementary material

The supplementary material is organized as follows: [Appendix S1](#) contains the proofs of the propositions and corollaries from [Section 3](#) and [Section 4](#) of the main text. [Appendix S2](#) provides full details of the posterior inference algorithm. [Appendix S3](#) presents additional experiments. [Appendices S4 to S6](#) provide background on the Lambert W function, regularly varying functions, and size-biased sampling for CRMs, respectively. For clarity, all sections, theorems, propositions, lemmas in the supplementary material are prefixed with “S” to distinguish them from those in the main text.

All the code used in this work is available at: [https://github.com/ValentinKil/rapidly\\_varying\\_crm](https://github.com/ValentinKil/rapidly_varying_crm).

## S1 Proofs

### S1.1 Proofs of [Section 3](#)

#### S1.1.1 Proof of [Proposition 3.1](#)

The asymptotics for the Laplace exponent are obtained trivially. For the Lévy intensity, consider

$$\begin{aligned} (\alpha - \tau)x^{\tau+1} \rho_{\text{mSt}}(x; \alpha, \tau) &= \int_{\tau}^{\alpha} \frac{s}{\Gamma(1-s)} x^{\tau-s} ds \\ &= \int_0^{\alpha-\tau} \frac{u+\tau}{\Gamma(1-u-\tau)} e^{-u \log x} du \end{aligned}$$

with the change of variables  $u = s - \tau$ . Let  $f(z) = \int_0^{\alpha-\tau} \frac{u+\tau}{\Gamma(1-u-\tau)} e^{-uz} du$ . We have, as  $u \rightarrow 0$ ,

$$\frac{u+\tau}{\Gamma(1-u-\tau)} \sim \begin{cases} u & \text{if } \tau = 0, \\ \frac{\tau}{\Gamma(1-\tau)} & \text{if } \tau > 0. \end{cases}$$

Using the Tauberian [Theorem S8](#), we obtain, as  $z \rightarrow \infty$

$$f(z) \sim \begin{cases} z^{-2} & \text{if } \tau = 0, \\ \frac{z^{-1}\tau}{\Gamma(1-\tau)} & \text{if } \tau > 0. \end{cases}$$

Hence, as  $x \rightarrow \infty$ ,

$$(\alpha - \tau)x^{\tau+1} \rho_{\text{mSt}}(x; \alpha, \tau) = f(\log x) \sim \begin{cases} \log^{-2}(x) & \text{if } \tau = 0, \\ \frac{\log^{-1}(x)\tau}{\Gamma(1-\tau)} & \text{if } \tau > 0. \end{cases}$$

The asymptotics of the full model follow. For the behaviour at 0, we proceed similarly. Consider

$$\begin{aligned} (\alpha - \tau)x^{\alpha+1} \rho_{\text{mSt}}(x; \alpha, \tau) &= \int_{\tau}^{\alpha} \frac{s}{\Gamma(1-s)} x^{\alpha-s} ds \\ &= \int_0^{\alpha-\tau} \frac{\alpha-u}{\Gamma(1+u-\alpha)} e^{-u \log(1/x)} du \end{aligned}$$

using the change of variables  $u = \alpha - s$ . Let  $g(z) = \int_0^{\alpha-\tau} \frac{\alpha-u}{\Gamma(1+u-\alpha)} e^{-uz} du$ . We have, as  $u \rightarrow 0$ ,

$$\frac{\alpha-u}{\Gamma(1+u-\alpha)} \sim \begin{cases} u & \text{if } \alpha = 1, \\ \frac{u^\alpha}{\Gamma(1-\alpha)} & \text{if } \alpha < 1. \end{cases}$$

Using the Tauberian [Theorem S8](#), we obtain, as  $z \rightarrow \infty$

$$g(z) \sim \begin{cases} z^{-2} & \text{if } \alpha = 1, \\ \frac{z^{-1-\alpha}}{\Gamma(1-\alpha)} & \text{if } \alpha < 1. \end{cases}$$

It follows that, as  $x \rightarrow 0$ ,

$$(\alpha - \tau)x^{\alpha+1} \rho_{\text{mSt}}(x; \alpha, \tau) = g(\log 1/x) \sim \begin{cases} \log^{-2}(1/x) & \text{if } \alpha = 1, \\ \frac{\log^{-1}(1/x)^\alpha}{\Gamma(1-\alpha)} & \text{if } \alpha < 1. \end{cases}$$

Hence the result for the full model.

### S1.1.2 Proof of [Proposition 3.2](#)

For  $m \geq 1$  and  $z \geq 0$ , recall that

$$\begin{aligned} \kappa_{\text{mGG}}(m, z; \alpha, \tau, \beta, c, \eta) &= \int_0^\infty w^m e^{-zw} \rho_{\text{mGG}}(w; \alpha, \tau, \beta, c, \eta) dw, \\ \kappa_{\text{mSt}}(m, z; \alpha, \tau) &= \int_0^\infty w^m e^{-zw} \rho_{\text{mSt}}(w; \alpha, \tau) dw. \end{aligned}$$

Note that

$$\kappa_{\text{mGG}}(m, z; \alpha, \tau, \beta, c, \eta) = \eta c^m \kappa_{\text{mSt}}(m, \beta + cz; \alpha, \tau). \quad (\text{S26})$$

We have, for  $z > 0$

$$\begin{aligned} \kappa_{\text{mSt}}(m, z; \alpha, \tau) &= \int_0^\infty w^m e^{-zw} \frac{1}{\alpha - \tau} \int_\tau^\alpha \frac{s}{\Gamma(1-s)} w^{-1-s} ds dw \\ &= \frac{1}{\alpha - \tau} \int_\tau^\alpha \frac{s}{\Gamma(1-s)} \frac{\Gamma(m-s)}{z^{m-s}} ds \\ &= \frac{z^{-m}}{\alpha - \tau} \int_\tau^\alpha s z^s \frac{\Gamma(m-s)}{\Gamma(1-s)} ds \end{aligned}$$

and  $\kappa_{\text{mSt}}(m, 0; \alpha, \tau) = \infty$ . For  $m = 1$ , we obtain

$$\kappa_{\text{mSt}}(1, z; \alpha, \tau) = \frac{z^{-1}}{\alpha - \tau} \int_\tau^\alpha s z^s ds. \quad (\text{S27})$$

For  $z = 1$ , we have  $\kappa_{\text{mSt}}(1, 1; \alpha, \tau) = \frac{1}{\alpha - \tau} \frac{\alpha^2 - \tau^2}{2} = \frac{\alpha + \tau}{2}$ . For  $z \neq 1$ , using the change of variable  $u = z^s$ ,  $du = z^s \log z ds$ , we obtain

$$\begin{aligned} \kappa_{\text{mSt}}(1, z; \alpha, \tau) &= \frac{z^{-1}}{\alpha - \tau} \int_{z^\tau}^{z^\alpha} s z^s ds = \frac{z^{-1}}{\alpha - \tau} \frac{1}{\log^2 z} \int_{z^\tau}^{z^\alpha} \log u du \\ &= \frac{z^{-1}}{\alpha - \tau} \frac{1}{\log^2 z} (z^\alpha \log z^\alpha - z^\alpha - [z^\tau \log z^\tau - z^\tau]) \\ &= \frac{z^{-1}}{\alpha - \tau} \frac{z^\tau - z^\alpha + (\alpha z^\alpha - \tau z^\tau) \log z}{(\log z)^2}. \end{aligned}$$

For  $m = 2$ , we have

$$\kappa_{\text{mSt}}(2, z; \alpha, \tau) = \frac{z^{-2}}{\alpha - \tau} \int_{\tau}^{\alpha} s(1-s)z^s ds. \quad (\text{S28})$$

For  $z = 1$ , this gives  $\kappa_{\text{mSt}}(2, 1; \alpha, \tau) = \frac{1}{\alpha - \tau} \left[ \frac{\alpha^2}{2} - \frac{\alpha^3}{3} - \left( \frac{\tau^2}{2} - \frac{\tau^3}{3} \right) \right]$ . For  $z \neq 1$ ,

$$\begin{aligned} \kappa_{\text{mSt}}(2, z; \alpha, \tau) &= \frac{z^{-2}}{\alpha - \tau} \int_{\tau}^{\alpha} s(1-s)z^s ds \\ &= z^{-1} \kappa_{\text{mSt}}(1, z; \alpha, \tau) - \frac{z^{-2}}{\alpha - \tau} \int_{\tau}^{\alpha} s^2 z^s ds \\ &= z^{-1} \kappa_{\text{mSt}}(1, z; \alpha, \tau) - \frac{z^{-2}}{\alpha - \tau} \int_{z\tau}^{z\alpha} \frac{\log^2 u}{\log^3 z} du. \end{aligned}$$

using  $\int \log^2 u du = u \log^2 u - 2u \log u + 2u + C$  gives the result.

### S1.1.3 Proof of Proposition 3.3

The proof follows directly from [Perman et al. \(1992\)](#), whose construction is recalled in [Appendix S6](#). Indeed, the conditional density  $p(w|t)$  in [Appendix S6](#) takes the form

$$\begin{aligned} p(w|t) &= \frac{w e^{-(t+\beta)w} \rho_{\text{mSt}}(w)}{\kappa_{\text{mSt}}(1, t + \beta)} \\ &= \frac{\int_{\tau}^{\alpha} s / \Gamma(1-s) w^{-s} e^{-(t+\beta)w} ds}{\int_{\tau}^{\alpha} s'(t+\beta)^{s'-1} ds'} \\ &= \int_{\tau}^{\alpha} p(s|t) p(w|t, s) ds \end{aligned}$$

where  $p(s|t) = \frac{s(t+\beta)^s}{\int_{\tau}^{\alpha} s'(t+\beta)^{s'} ds'}$   $\mathbf{1}_{\{s \in (\tau, \alpha)\}}$  and  $p(w|t, s) = w^{-s} e^{-(t+\beta)w} \frac{(t+\beta)^s}{\Gamma(1-s)}$ .

### S1.1.4 Proof of Proposition 3.4

We can follow the same proof strategy as in [Lee et al. \(2023\)](#). Let  $\alpha = 1$ ,  $\tau = 0$ . Without loss of generality, set  $c = 1$ . According to [Lee et al. \(2023, Proposition 5.1\)](#), we have

$$\begin{aligned} \mathbb{E}[R_n | T_{n+1}] &= \int_0^{\infty} w e^{-w T_{n+1}} \rho_{\text{mGG}}(w; 1, 0, \beta, 1, \eta) dw \\ &= \kappa_{\text{mGG}}(1, T_{n+1}; 1, 0, \beta, 1, \eta) \\ &= \eta \kappa_{\text{mSt}}(1, \beta + T_{n+1}; 1, 0) \end{aligned}$$

with  $T_{n+1} = \psi_{\text{mSt}}^{-1} \left( \frac{\xi_{n+1}}{\eta} + \psi_{\text{mSt}}(\beta) \right) - \beta$  where  $\xi_{n+1}$  is a  $\text{Gamma}(n+1, 1)$  random variable. Similarly,

$$\begin{aligned} \mathbb{V}[R_n | T_{n+1}] &= \int_0^{\infty} w^2 e^{-w T_{n+1}} \rho_{\text{mGG}}(w; 1, 0, \beta, 1, \eta) dw \\ &= \kappa_{\text{mGG}}(2, T_{n+1}; 1, 0, \beta, 1, \eta) \\ &= \eta \kappa_{\text{mSt}}(2, \beta + T_{n+1}; 1, 0). \end{aligned}$$

Using [Proposition 3.1](#) and [Appendix S5.3](#) we get

$$\psi_{\text{mSt}}^{-1}(x) \sim x \log(x) \text{ as } x \rightarrow +\infty.$$

And so

$$z_n := \beta + T_{n+1} \sim \frac{n}{\eta} \log(n) \text{ as } n \rightarrow \infty.$$

In particular,  $T_{n+1} \rightarrow +\infty$ , so as  $n \rightarrow \infty$ ,

$$\mathbb{E}[R_n|T_{n+1}] = \eta \kappa_{\text{mGG}}(1, z_n; 1, 0) = \frac{\eta}{z_n} \frac{1 - \beta - T_{n+1} + (z_n) \log(z_n)}{\log^2(T_{n+1})} \sim \frac{\eta}{\log(z_n)} \sim \frac{\eta}{\log(n)}$$

and

$$\begin{aligned} \mathbb{V}[R_n|T_{n+1}] &= \eta \kappa_{\text{mGG}}(2, z_n; 1, 0) = \frac{1}{z_n} \mathbb{E}[R_n|T_{n+1}] - \frac{\eta}{z_n^2} \int_1^{z_n} \frac{\log^2(u) du}{\log^3(z_n)} \\ &\sim \frac{\eta^2}{n \log^2(n)} - \frac{\eta^2 n \log^3(n)}{n^2 \log^5(n)} + \frac{2\eta^2 n \log^2(n)}{n^2 \log^5(n)} \sim \frac{2\eta^2}{n \log^3(n)}. \end{aligned} \quad (\text{S29})$$

Finally, for all  $\epsilon > 0$ , using Chebyshev's inequality, we obtain

$$\mathbb{P}\left(\left|\frac{R_n}{\mathbb{E}[R_n|T_{n+1}]} - 1\right| > \epsilon | T_{n+1}\right) \leq \frac{\mathbb{V}[R_n|T_{n+1}]}{\epsilon^2 \mathbb{E}[R_n|T_{n+1}]^2}$$

and as

$$\frac{\mathbb{V}[R_n|T_{n+1}]}{\mathbb{E}[R_n|T_{n+1}]^2} \sim \frac{2}{n \log(n)},$$

then by the Borel-Cantelli lemma, almost surely as  $n \rightarrow +\infty$ ,

$$R_n \sim \mathbb{E}[R_n|T_{n+1}] \sim \frac{\eta}{\log(n)}.$$

### S1.1.5 Proof of [Proposition 3.5](#)

The total mass  $G(\Theta)$  has Laplace transform, for  $t \geq 0$ ,

$$\mathbb{E}[e^{-tG(\Theta)}] = \exp(-\psi_{\text{mGG}}(t; \alpha, \tau, \beta, c, \eta)).$$

We have

$$\mathbb{E}[e^{-tc(\frac{n}{n})^{\frac{1}{s_i}} X_i}] = \exp\left(-\frac{\eta}{n}((ct + \beta)^{s_i} - \beta^{s_i})\right).$$

Hence

$$\mathbb{E}[e^{-tS_n}] = \exp\left(-\frac{\eta}{n} \sum_{i=1}^n ((ct + \beta)^{s_i} - \beta^{s_i})\right).$$

Additionally, as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n t^{s_i} = \psi_{\text{mSt}}(t; \alpha, \tau), \quad (\text{S30})$$

it follows that  $\mathbb{E}[e^{-tS_n}] \rightarrow \mathbb{E}[e^{-tG(\Theta)}]$  as  $n \rightarrow \infty$ . As  $\psi_{\text{mGG}}(t; \alpha, \tau, \beta, c, \eta)$  is continuous, the result follows.

## S1.2 Proofs of Section 4

Let

$$\bar{\rho}_{\text{mGG}}(x; \alpha, \tau, \beta, c, \eta) := \int_x^\infty \rho_{\text{mGG}}(w; \alpha, \tau, \beta, c, \eta) dw$$

be the tail Lévy intensity of the mGG CRM. The proofs of Section 4 mostly follow from Caron et al. (2023, Proposition 11, Theorem 2 and Corollary 1), which state the asymptotic properties of a CRM-based Caron-Fox graph model described in Section 4.1, given the behavior of the tail Lévy intensity at 0. Proposition 11 in Caron et al. (2023) requires that  $\bar{\rho}_{\text{mGG}}(x; 1, 0, \beta, c, \eta) \underset{x \rightarrow 0}{\sim} w^{-1} \tilde{\ell}(1/x)$  for some slowly varying function  $\tilde{\ell}$ . We prove in Lemma S1 below that this holds with

$$\tilde{\ell}(1/x) = \frac{\eta c}{1 - \tau \log^2(1/x)} \frac{x^{-1}}{\log^2(1/x)}.$$

The mean

$$m = \int_0^\infty w \rho_{\text{mGG}}(w; 1, 0, \beta, c, \eta) dw = \begin{cases} \eta c \frac{1 - \beta + \beta \log(\beta)}{\beta \log^2(\beta)} & \beta \neq 1, \\ \frac{\eta c}{2} & \beta = 1. \end{cases}$$

is given by Proposition 3.2, and is finite for  $\beta > 0$ . Define

$$\ell(x) = 2m \tilde{\ell}(x) = \frac{C}{\log^2(x)},$$

where

$$C = \begin{cases} 2(\eta c)^2 \frac{\beta^{-1} - 1 + \log \beta}{(\log \beta)^2} & \beta \neq 1, \\ (\eta c)^2 & \beta = 1. \end{cases}$$

Let

$$\ell_1(x) = \int_x^\infty y^{-1} \ell(y) dy = \frac{C}{\log(x)}$$

and define

$$\ell_1^*(y) = \left[ \left\{ \sqrt{\ell_1(\sqrt{y})} \right\}^\# \right]^2.$$

As shown in Lemma S2 below,

$$\ell_1^*(y) = \frac{1}{2C} \log(y).$$

### S1.2.1 Secondary lemmas

**Lemma S1.** *We have*

$$\bar{\rho}_{\text{mGG}}(x; 1, 0, \beta, c, \eta) \underset{x \rightarrow 0}{\sim} x^{-1} \tilde{\ell}(1/x)$$

where

$$\tilde{\ell}(1/x) = \frac{\eta c}{1 - \tau \log^2(1/x)} \frac{x^{-1}}{\log^2(1/x)}.$$

**Proof.** Proposition 3.1 implies that  $\rho_{\text{mGG}}(w; 1, \tau, \beta, c, \eta) = w^{-2} L(1/x)$ , where  $L(1/x) \underset{x \rightarrow 0}{\sim} \frac{1}{(1-\tau) \log^2(1/w)}$ . Using Corollary S4, we obtain

$$\bar{\rho}_{\text{mGG}}(x; 1, \tau, \beta, c, \eta) \underset{x \rightarrow 0}{\sim} x^{-1} \frac{\eta c}{1 - \tau \log^2(x)} \frac{1}{\log^2(x)}.$$

□

**Lemma S2.** Let  $\ell_1(x) = \frac{C}{\log(x)}$  for some constant  $C$ , and define

$$\ell_1^*(y) = \left[ \left\{ \sqrt{\ell_1(\sqrt{y})} \right\}^\# \right]^2,$$

where  $\ell^\#$  denotes the de Bruijn conjugate (see [Appendix S5.3](#)) of the slowly varying function  $\ell$ . Then

$$\ell_1^*(y) = \frac{1}{2C} \log(y).$$

**Proof.** The pair  $(\log(y), \log^{-1}(y))$  is conjugate. Using [Proposition S6](#) the following pairs are also conjugate

$$\begin{aligned} & \left( \frac{1}{4C} \log(y), 4C \log^{-1}(y) \right), \\ & \left( \sqrt{\frac{1}{4C} \log(y^2)}, \sqrt{4C \log^{-1}(y^2)} \right), \\ & \left( \sqrt{\frac{1}{C} \log(\sqrt{y})}, \sqrt{C \log^{-1}(\sqrt{y})} \right), \\ & \left( \sqrt{\frac{1}{2C} \log(y)}, \sqrt{\ell_1(\sqrt{y})} \right). \end{aligned}$$

Hence  $\ell_1^*(y) = \frac{1}{2C} \log(y)$ . □

### S1.2.2 Proof of [Proposition 4.3](#)

Finally, while most of [Proposition 4.3](#) follows from Theorem 2 in [Caron et al. \(2023\)](#), that theorem did not provide an exact asymptotic expression for the proportion of nodes of degree  $j$  for  $j \geq 2$ , which we now derive. Let  $\tilde{N}_{t,j} = \sum_{k \geq j} N_{t,k}$  denote the number of nodes of degree at least  $j$ . From [Caron et al. \(2023, Theorem 2\)](#), we have, almost surely,

$$\tilde{N}_{t,j} \sim \mathbb{E}(\tilde{N}_{t,j}).$$

Then, for  $j \geq 2$ , almost surely, using [Caron et al. \(2023, Theorem 1\)](#),

$$N_{t,j} = \tilde{N}_{t,j} - \tilde{N}_{t,j+1} \sim \mathbb{E}(\tilde{N}_{t,j}) - \mathbb{E}(\tilde{N}_{t,j+1}) = \mathbb{E}(\tilde{N}_{t,j} - \tilde{N}_{t,j+1}) = \mathbb{E}(N_{t,j}) \sim \frac{t^2}{j(j-1)} \frac{C}{\log^2(t)}.$$

The proof is similar for  $j = 1$ . The dominance of nodes of degree 1 in the graph follows directly from

$$\log(t)^{-2} = o(\log(t)^{-1}).$$

For [\(21\)](#), we know from Theorem 2 in [Caron et al. \(2023\)](#) that

$$\tilde{N}_{t,2} \sim \mathbb{E}(\tilde{N}_{t,2}) = \sum_{k \geq 2} \mathbb{E}(N_{t,k}) \sim \sum_{k \geq 2} \frac{t^2}{k(k-1)} \frac{C}{\log^2(t)} = t^2 \frac{C}{\log^2(t)}.$$

Therefore, we obtain

$$\frac{N_{t,j}}{\tilde{N}_{t,2}} \sim \frac{\frac{t^2}{j(j-1)} \frac{C}{\log^2(t)}}{t^2 \frac{C}{\log^2(t)}} = \frac{1}{j(j-1)}.$$

## S2 Directed multigraph posterior and MCMC Algorithm

We give here the details of the algorithm introduced in [Section 4.3](#).

### S2.1 Directed multigraph posterior

Formally, the symmetric 0-1 valued atomic measure [Equation \(19\)](#) can be viewed as a transformation of a integer-valued atomic measure (also called directed multigraph) ; this view is useful when it comes to posterior inference. We note the directed multigraph

$$M = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} Q_{ij} \delta_{\theta_i, \theta_j}, \quad (\text{S31})$$

where  $Q_{ij}$  counts the number of directed edges from node  $i$  to node  $j$ , with location  $\theta_i$  and  $\theta_j$ . Given a CRM  $G \sim \text{CRM}(\rho, \lambda)$  for any Lévy measure  $\rho$ ,  $M$  is simply generated from a Poisson process with intensity given by the product measure  $G \times G$ . To derive the undirected graphs  $U$  we simply put  $Z_{ij} = Z_{ji} = \min(\tilde{q}_{ij}, 1)$  where  $\tilde{q}_{ji} = \tilde{q}_{ij} = Q_{ij} + Q_{ji}$ . This formulation is equivalent to the one introduced in [Section 4.1](#) as explained in [Caron and Fox \(2017\)](#).

For  $t \in \mathbb{R}_+$  we note  $M_t$  the measure  $M$  restrained to the square  $[0, t]^2$ . Similarly we note  $G_t$  the measure  $G$  restrained to  $[0, t]$  and  $G_t^*$  the total masse of  $G_t$ . The directed multigraph posterior given  $(Q_{ij})_{1 \leq i, j \leq N_t}$  is characterise in the following theorem

**Theorem S1** (Theorem 6, [Caron and Fox, 2017](#)). *For  $N_t \geq 1$ , let  $\theta_1 < \dots < \theta_{N_t}$  be the set of support points of the measure  $M_t$ . Let  $w_i = G_t(\{\theta_i\})$  and  $w_* = G_t^* - \sum_{i=1}^{N_t} w_i$ . We have*

$$\begin{aligned} & \mathbb{P} \left[ (w_i \in dw_i)_{1:N_t}, w_* \in dw_* \mid (Q_{ij})_{1 \leq i, j \leq N_t} \right] \\ & \propto \exp \left\{ - \left( \sum_{i=1}^{N_t} w_i + w_* \right)^2 \right\} \left\{ \prod_{i=1}^{N_t} w_i^{m_i} \rho(dw_i) \right\} W_t^*(dw_*) \end{aligned}$$

where  $m_i = \sum_{j=1}^{N_t} \tilde{q}_{ij} > 0$  for  $i = 1, \dots, N_t$  are the node degrees of the multigraph and  $W_t^*$  is the probability distribution of the random variable  $G_t^*$ , with Laplace transform

$$\mathbb{E} [\exp(-xG_t^*)] = \exp(-\psi(x)).$$

Additionally, conditionally on observing an empty graph, i.e.,  $N_t = 0$ , we have

$$\mathbb{P} [w_* \in dw_* \mid N_t = 0] \propto \exp(-w_*^2) W_t^*(dw_*).$$

### S2.2 Our MCMC Algorithm

#### S2.2.1 Step 1 : update of the $w$ and $s$

We propose to use Hamiltonian Monte Carlo to perform the step 1 of our previous algorithm which requires the computation of the gradient of the log-posterior. Following [Theorem S1](#) we can compute the conditional probability.

$$\begin{aligned}
& \mathbb{P}[(w_i \in dw_i)_{0:N_t} | m_{1:N_t}, \phi, w_*] \\
& \propto \exp \left[ - \left( \sum_{i=1}^{N_\alpha} w_i + w_* \right)^2 \right] \left( \prod_{i=1}^{N_\alpha} w_i^{m_i} \rho_{\text{mGG}}(dw_i) \right) \\
& \propto \int \cdots \int_{s_{1:N_t}} \exp \left[ - \left( \sum_{i=1}^{N_t} w_i + w_* \right)^2 \right] \left( \prod_{i=1}^{N_t} w_i^{m_i} \hat{\rho}(w_i, s_i; \phi) dw_i \right) \left( \prod_{i=1}^{N_t} \mathbf{1}_{\{s_i \in (0,1)\}} ds_i \right) \\
& \propto \int \cdots \int_{s_{1:N_t}} \exp \left[ - \left( \sum_{i=1}^{N_t} w_i + w_* \right)^2 \right] \left( \prod_{i=1}^{N_t} w_i^{m_i} \frac{\eta s_i c^{s_i}}{\Gamma(1-s_i)} w_i^{-1-s_i} e^{-\beta w_i/c} dw_i \right) \left( \prod_{i=1}^{N_t} \mathbf{1}_{\{s_i \in (0,1)\}} ds_i \right).
\end{aligned}$$

So we can deduce the conditional PDF

$$\begin{aligned}
p(w_{1:N_t}, s_{1:N_t} | m_{1:N_t}, \phi, w_*) & \propto \exp \left[ - \left( \sum_{i=1}^{N_t} w_i + w_* \right)^2 \right] \left( \prod_{i=1}^{N_t} w_i^{m_i} \frac{\eta s_i c^{s_i}}{\Gamma(1-s_i)} w_i^{-1-s_i} e^{-\beta w_i/c} \mathbf{1}_{\{s_i \in (0,1)\}} \right) \\
& = e^{-(\sum w_i + w_*)^2 - (\sum_i w_i) \frac{\beta}{c}} \eta^N \left( \prod_{i=1}^N w_i^{m_i - s_i - 1} \frac{s_i c^{s_i}}{\Gamma(1-s_i)} \mathbf{1}_{\{s_i \in (0,1)\}} \right).
\end{aligned}$$

Let  $\text{sig}(x) = \frac{1}{1+e^{-x}}$  with inverse  $\text{sig}^{-1}(z) = \log(\frac{z}{1-z})$ . Note  $\text{sig}'(x) = \text{sig}(x)(1-\text{sig}(x))$ . Do the change of variable  $v_i = \log(w_i)$ , and,  $z_i = \text{sig}^{-1}(s_i)$ . Hence  $dw_i = e^{v_i} dv_i = w_i dv_i$  and  $ds_i = \text{sig}(z_i)(1-\text{sig}(z_i)) dz_i = s_i(1-s_i) dz_i$ . It follows that

$$\begin{aligned}
p(v_{1:N_t}, z_{1:N_t} | m_{1:N_t}, \phi, w_*) & \propto e^{-(\sum w_i + w_*)^2 - (\sum_i w_i) \frac{\beta}{c}} \eta^N \left[ \prod_{i=1}^N w_i^{m_i - s_i - 1} \frac{s_i c^{s_i}}{\Gamma(1-s_i)} w_i s_i (1-s_i) \right] \\
& = e^{-(\sum w_i + w_*)^2 - (\sum_i w_i) \frac{\beta}{c}} \eta^N \left[ \prod_{i=1}^N w_i^{m_i - s_i} \frac{s_i^2 (1-s_i)}{\Gamma(1-s_i)} c^{s_i} \right] \\
& = e^{-(\sum w_i + w_*)^2 - (\sum_i w_i) \frac{\beta}{c}} \eta^N \left[ \prod_{i=1}^N w_i^{m_i - s_i} \frac{s_i^2 (1-s_i)^2}{\Gamma(2-s_i)} c^{s_i} \right],
\end{aligned}$$

where the last line comes from  $\Gamma(2-s) = (1-s)\Gamma(1-s)$ . This form is preferred to avoid numerical instabilities when  $s_i$  is close to 1. We obtain (up to a constant independent of  $v_i, z_i$ )

$$\begin{aligned}
\ln(p(v_{1:N_t}, z_{1:N_t} | m_{1:N_t}, \phi, w_*)) & = - \left( \sum w_i + w_* \right)^2 - \left( \sum_i w_i \right) \frac{\beta}{c} \\
& \quad + \sum_{i=1}^N (m_i - s_i) \log w_i + 2 \log s_i + 2 \log(1-s_i) - \log \Gamma(2-s_i) + s_i \log(c) \\
& := g(v_{1:N_t}, z_{1:N_t}).
\end{aligned}$$

Now we can compute the gradient of the log posterior. We have

$$\begin{aligned}
\frac{\partial g(v_{1:N_t}, z_{1:N_t})}{\partial v_i} & = \frac{\partial w_i}{\partial v_i} \frac{\partial g(v_{1:N_t}, z_{1:N_t})}{\partial w_i} \\
& = w_i \times \left[ -2 \left( \sum w_i + w_* \right) - \frac{\beta}{c} + \frac{m_i - s_i}{w_i} \right] \quad (\text{S32}) \\
& = m_i - s_i - w_i \left[ 2 \left( \sum w_i + w_* \right) + \frac{\beta}{c} \right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial g(v_{1:N_t}, z_{1:N_t})}{\partial z_i} &= \frac{\partial s_i}{\partial z_i} \frac{\partial g(v_{1:N_t}, z_{1:N_t})}{\partial s_i} \\
&= s_i(1 - s_i) \times \left[ -\log w_i + \frac{2}{s_i} - \frac{2}{1 - s_i} + \psi(2 - s_i) + \log c \right] \\
&= s_i(1 - s_i) \times [\psi(2 - s_i) - \log w_i + \log c] + 2 - 4s_i,
\end{aligned} \tag{S33}$$

where  $\psi = \frac{\Gamma'}{\Gamma}$  is the digamma function which is implemented in `scipy`. Noting that  $\psi(x + 1) = \psi(x) + \frac{1}{x}$ , the above is equivalent to

$$\frac{\partial g(v_{1:N_t}, z_{1:N_t})}{dz_i} = s_i(1 - s_i) \times [\psi(1 - s_i) - \log w_i + \log c] + 2 - 3s_i$$

but the first expression should be preferred.

In order to perform Step 1, we first update the  $w_{1:N_t}$  with  $s$  fixed and then the  $s_{1:N_t}$  with  $w$  fixed, for both update we use an HMC update with momentum variable  $\mathbf{p}_w$  and  $\mathbf{p}_s$ . We refer the reader to [Neal \(2011\)](#) and [Betancourt \(2018\)](#) for an overview. We start by explaining how we update the  $w$ , the update of the latent  $s$  is similar.

**Update of the  $w$ .** Let  $L \geq 1$  be the number of Leapfrog steps and  $\epsilon_w > 0$  the step size. The potential energy is given by

$$U(w, s, m_{1:N_t}, \phi, w_*) = \ln(p(v_{1:N_t}, z_{1:N_t} | m_{1:N_t}, \phi, w_*)) = g(u, v)$$

whose gradient  $U'_w$  has been given in [S32](#). The algorithm proceeds by first sampling momentum variable as

$$\mathbf{p}_w \sim \mathcal{N}(0, I_{N_t})$$

The Hamiltonian proposal  $q(\tilde{\mathbf{w}}, \tilde{\mathbf{p}}_w | \mathbf{w}, \mathbf{p}_w)$  (we drop out the  $1 : N_t$  for lisibility) is obtained by the following Leapfrog algorithm : simulate  $L$  steps of discretised Hamiltonian via

$$\tilde{\mathbf{p}}_w^{(0)} = \mathbf{p}_w + \frac{\epsilon_w}{2} U'_w(\mathbf{w}, \mathbf{s}, w_*, \phi); \quad \tilde{\mathbf{w}}^{(0)} = \mathbf{w}.$$

And for  $l = 1, \dots, L - 1$ ,

$$\log(\tilde{\mathbf{w}}^{(l)}) = \log(\tilde{\mathbf{w}}^{(l-1)}) + \epsilon_w \tilde{\mathbf{p}}_w^{(l-1)}; \quad \tilde{\mathbf{p}}_w^{(l)} = \tilde{\mathbf{p}}_w^{(l-1)} + \epsilon_w U'_w(\tilde{\mathbf{w}}^{(l)}, \mathbf{s}, w_*, \phi).$$

Finally, set

$$\log(\tilde{\mathbf{w}}) = \log(\tilde{\mathbf{w}}^{(L-1)}) + \epsilon_w \tilde{\mathbf{p}}_w^{(L-1)}; \quad \tilde{\mathbf{p}}_w = \tilde{\mathbf{p}}_w^{(L-1)} + \frac{\epsilon_w}{2} U'_w(\tilde{\mathbf{w}}, \mathbf{s}, w_*, \phi).$$

The proposal  $(\tilde{\mathbf{w}}, \tilde{\mathbf{p}}_w)$  is accepted with probability  $\min(1, r_w)$  with

$$r_w = \exp \left[ -\frac{1}{2} \sum_{i=1}^{N_t} (\tilde{p}_{w,i}^2 - p_{w,i}^2) \right] \frac{p(\tilde{\mathbf{v}} | \mathbf{z}, G, \phi, w_*)}{p(\mathbf{v} | \mathbf{z}, G, \phi, w_*)}.$$

**Update of the latent  $s$ .** Once  $w$  has been updated, the update of  $s$  follows the same principle: the algorithm proceeds by first sampling momentum variable as

$$\mathbf{p}_s \sim \mathcal{N}(0, I_{N_t}).$$

The Hamiltonian proposal  $q(\tilde{\mathbf{s}}, \tilde{\mathbf{p}}_s, |s, \mathbf{p}_s)$  is obtained by  $L$  Leapfrog step:

$$\tilde{\mathbf{p}}_s^{(0)} = \mathbf{p}_s + \frac{\epsilon_s}{2} U'_s(\mathbf{w}, \mathbf{s}, w_*, \phi); \quad \tilde{\mathbf{s}}^{(0)} = \mathbf{s}.$$

And for  $l = 1, \dots, L - 1$ ,

$$\log(\tilde{\mathbf{s}}^{(l)}) = \log(\tilde{\mathbf{s}}^{(l-1)}) + \epsilon_s \tilde{\mathbf{p}}_s^{(l-1)} \quad \tilde{\mathbf{p}}_s^{(l)} = \tilde{\mathbf{p}}_s^{(l-1)} + \epsilon_s U'_s(\mathbf{w}, \tilde{\mathbf{s}}^{(l)}, w_*, \phi);$$

and finally set

$$\text{logit}(\tilde{\mathbf{s}}) = \text{logit}(\tilde{\mathbf{s}}^{(L-1)}) + \epsilon_s \tilde{\mathbf{p}}_s^{(L-1)} \quad \tilde{\mathbf{p}}_s = \tilde{\mathbf{p}}_s^{(L-1)} + \frac{\epsilon_s}{2} U'_s(\mathbf{w}, \tilde{\mathbf{s}}, w_*, \phi).$$

The proposal  $(\tilde{\mathbf{s}}, \tilde{\mathbf{p}}_s)$  is accepted with probability  $\min(1, r_s)$  with

$$r_s = \exp \left[ -\frac{1}{2} \sum_{i=1}^{N_t} (\tilde{p}_{s,i}^2 - p_{s,i}^2) \right] \frac{p(\tilde{\mathbf{z}} | \mathbf{v}, G, \phi, w_*)}{p(\mathbf{z} | \mathbf{v}, G, \phi, w_*)}.$$

### S2.2.2 Step 2 : update of hyperparameters

Let  $G_\phi^*$  be the probability distribution of the random variable of  $G(\Theta)$ . Recall that we have  $\mathbb{E}[e^{-tG(\Theta)}] = e^{-\psi(t)}$  for  $t \geq 0$ . Let  $g_\phi^*$  denote the probability density of  $G_\phi^*$  with respect of the Lebesgue mesure. First we can note the interesting property :

$$g_{\alpha, \tau, \beta_1 + c\beta_2, c, \eta}^*(x) = g_{\alpha, \tau, \beta_1, c, \eta}^*(x) \frac{\exp(-x\beta_2)}{\exp(-\eta \psi_{\text{mGG}}(\beta_2; \alpha, \tau, \beta_1, c, 1))}. \quad (\text{S34})$$

**Proof.** For all  $t \in \mathbb{R}$ ,

$$\begin{aligned} & \int_0^{+\infty} e^{-tx} g_{\alpha, \tau, \beta_1, c, \eta}^*(x) \frac{\exp(-x\beta_2)}{\exp(-\eta \psi_{\text{mGG}}(\beta_2; \alpha, \tau, \beta_1, c, 1))} dx \\ &= \exp(-\eta \psi_{\text{mGG}}(\beta_2; \alpha, \tau, \beta_1, c, 1)) \int_0^{+\infty} e^{-tx} g_{\alpha, \tau, \beta_1, c, \eta}^*(x) \exp(-x\beta_2) dx \\ &= \exp(-\eta \psi_{\text{mGG}}(\beta_2; \alpha, \tau, \beta_1, c, 1)) \int_0^{+\infty} e^{-x(t+\beta_2)} g_{\alpha, \tau, \beta_1, c, \eta}^*(x) dx \\ &= \exp(-\eta \psi_{\text{mGG}}(\beta_2; \alpha, \tau, \beta_1, c, 1)) \exp(-\psi_{\text{mGG}}(t + \beta_2, \alpha, \tau, \beta_1, c, \eta)) \\ &= \exp[\eta \psi_{\text{mSt}}(\beta_1 + c\beta_2, \alpha, \tau) - \eta \psi_{\text{mSt}}(\beta_1, \alpha, \tau) - \eta \psi_{\text{mSt}}(\beta_1 + c\beta_2 + ct, \alpha, \tau) + \eta \psi_{\text{mSt}}(\beta_1, \alpha, \tau)] \\ &= \exp(-\psi_{\text{mGG}}(t, \alpha, \tau, \beta_1 + c\beta_2, c, \eta)). \end{aligned}$$

□

Following the same strategy than in [Caron and Fox \(2017, Appendix F.2\)](#), we propose  $(\tilde{\phi}, \tilde{w}_*)$  from  $q(\tilde{\phi}, \tilde{w}_* | \phi, w_*)$  and accept with probability  $\min(1, r)$  where

$$\begin{aligned} r &= \frac{\exp \left\{ -\left( \sum_{i=1}^N w_i + \tilde{w}_* \right)^2 \right\}}{\exp \left\{ -\left( \sum_{i=1}^N w_i + w_* \right)^2 \right\}} \left\{ \prod_{i=1}^N \frac{\hat{\rho}(w_i, s_i; \tilde{\phi})}{\hat{\rho}(w_i, s_i; \phi)} \right\} \frac{g_\phi^*(\tilde{w}_*) p(\tilde{\phi}) q(\phi, w_* | \tilde{\phi}, \tilde{w}_*)}{g_\phi^*(w_*) p(\phi) q(\tilde{\phi}, \tilde{w}_* | \phi, w_*)} \\ &= \frac{\exp \left\{ -\left( \sum_{i=1}^N w_i + \tilde{w}_* \right)^2 \right\}}{\exp \left\{ -\left( \sum_{i=1}^N w_i + w_* \right)^2 \right\}} \left\{ \left( \frac{\tilde{\eta}}{\eta} \right)^N \left( \frac{\tilde{c}}{c} \right)^{\sum_{i=1}^N s_i} e^{-(\tilde{\beta}/\tilde{c} - \beta/c)(\sum_{i=1}^N w_i)} \right\} \frac{g_\phi^*(\tilde{w}_*) p(\tilde{\phi}) q(\phi, w_* | \tilde{\phi}, \tilde{w}_*)}{g_\phi^*(w_*) p(\phi) q(\tilde{\phi}, \tilde{w}_* | \phi, w_*)}. \end{aligned}$$

We shall use the proposal

$$q(\tilde{\phi}, \tilde{w}_* | \phi, w_*) = q(\tilde{\beta} | \beta) q(\tilde{c} | c) q(\tilde{\eta} | \beta, \tilde{c}, \eta, w_*) q(\tilde{w}_* | \tilde{\phi}, w_*)$$

where

$$\begin{aligned} q(\tilde{\beta} | \beta) &= \text{lognormal}(\tilde{\beta}; \log(\beta), \sigma_\beta^2), \\ q(\tilde{c} | c) &= \text{lognormal}(\tilde{c}; \log(c), \sigma_c^2), \\ q(\tilde{w}_* | \tilde{\phi}, w_*) &= g_{\alpha, \tau, \tilde{\beta} + 2\tilde{c} \sum_{i=1}^N w_i + 2\tilde{c} w_*, \tilde{c}, \tilde{\eta}}^*(\tilde{w}_*). \end{aligned}$$

For  $\eta$ , we alternatively use two proposals

$$q(\tilde{\eta} | \tilde{\beta}, \tilde{c}, \eta, w_*) = \text{Gamma}\left(\tilde{\eta}; N, \psi_\phi^1\left(2 \sum_{i=1}^N w_i + 2w_*\right)\right) \quad (\text{S35})$$

where  $\psi_\phi^1(t) = \psi_{\text{mGG}}(t; \alpha, \tau, \beta, c, 1)$ , or a simple random walk on  $\log \eta$

$$q(\tilde{\eta} | \eta) = \text{lognormal}(\tilde{\eta}; \log(\eta), \sigma_\eta^2). \quad (\text{S36})$$

Note that, under the priors (22),

$$\frac{p(\tilde{\beta})q(\beta | \tilde{\beta})}{p(\beta)q(\tilde{\beta} | \beta)} = \frac{p(\tilde{c})q(c | \tilde{c})}{p(c)q(\tilde{c} | c)} = 1$$

and

$$\frac{g_\phi^*(\tilde{w}_*)}{g_\phi^*(w_*)} \times \frac{q(w_* | \phi, \tilde{w}_*)}{q(\tilde{w}_* | \tilde{\phi}, w_*)} = \frac{e^{-w_* (2 \sum_{i=1}^N w_i + 2\tilde{w}_*)}}{e^{-\tilde{w}_* (2 \sum_{i=1}^N w_i + 2w_*)}} \times \frac{e^{-\tilde{\eta} \psi_\phi^1(2 \sum_{i=1}^N w_i + 2w_*)}}{e^{-\eta \psi_\phi^1(2 \sum_{i=1}^N w_i + 2\tilde{w}_*)}}.$$

Sometimes it is useful to put a gamma prior on  $\beta$  and  $c - 1$ :

$$p(\beta) = \text{Gamma}(\beta; a_\beta, b_\beta), \quad p(c - 1) = \text{Gamma}(c; a_c, b_c).$$

In that case we have

$$\frac{p(\tilde{\beta})q(\beta | \tilde{\beta})}{p(\beta)q(\tilde{\beta} | \beta)} = \left(\frac{\tilde{\beta}}{\beta}\right)^{a_\beta} \exp(-b_\beta(\tilde{\beta} - \beta)), \quad \text{and} \quad \frac{p(\tilde{c})q(c | \tilde{c})}{p(c)q(\tilde{c} | c)} = \left(\frac{\tilde{c}}{c}\right)^{a_c} \exp(-b_c(\tilde{c} - c)).$$

Note that we can recover the improper prior by taking  $a = b = 0$  in both case. It follows that the acceptance ratio is

$$\begin{aligned} r &= \frac{\exp\left[-\left(\sum_{i=1}^N w_i + \tilde{w}_*\right)^2\right]}{\exp\left[-\left(\sum_{i=1}^N w_i + w_*\right)^2\right]} \left\{ \left(\frac{\tilde{c}}{c}\right)^{\sum_{i=1}^N s_i} e^{-(\tilde{\beta}/\tilde{c} - \beta/c + 2w_* - 2\tilde{w}_*)(\sum_{i=1}^N w_i)} \right\} \\ &\times \left(\frac{\tilde{\beta}}{\beta}\right)^{a_\beta} \exp(-b_\beta(\tilde{\beta} - \beta)) \times \left(\frac{\tilde{c}}{c}\right)^{a_c} \exp(-b_c(\tilde{c} - c)) \\ &\times \underbrace{\left(\frac{\tilde{\eta}}{\eta}\right)^{N-1} \frac{e^{-\tilde{\eta} \psi_\phi^1(2 \sum_{i=1}^N w_i + 2w_*)}}{e^{-\eta \psi_\phi^1(2 \sum_{i=1}^N w_i + 2\tilde{w}_*)}} \frac{q(\eta | \beta, c, \tilde{\eta}, \tilde{w}_*)}{q(\tilde{\eta} | \tilde{\beta}, \tilde{c}, \eta, w_*)}}_{r_2}. \end{aligned}$$

Under the gamma proposal (S35) for  $\eta$ ,

$$r_2 = \left( \frac{\psi_{\phi}^1(2 \sum_{i=1}^N w_i + 2\tilde{w}_*)}{\psi_{\phi}^1(2 \sum_{i=1}^N w_i + 2w_*)} \right)^N.$$

Under the log-normal proposal (S36) for  $\eta$ ,

$$r_2 = \left( \frac{\tilde{\eta}}{\eta} \right)^N \frac{e^{-\tilde{\eta}\psi_{\phi}^1(2 \sum_{i=1}^N w_i + 2w_*)}}{e^{-\eta\psi_{\phi}^1(2 \sum_{i=1}^N w_i + 2\tilde{w}_*)}}.$$

### S2.2.3 Step 3 : update of the latent counts

We simply update the latent counts by using the conditional distribution Equation (23).

## S3 Additional experiments

### S3.1 Synthetic Data

#### S3.1.1 Proportion of degree one

Under our graph model, Proposition 4.3 states that the proportion of nodes of degree one converges to 1 as the size of the graph goes to infinity. This phenomenon is illustrated in Figure S9.

#### S3.1.2 Cluster

In order to get a more precise idea of what the cluster structure of the generated networks looks like, we performed additional simulations. These simulations were performed using the Python implementation available on GitHub, running on an Apple M2 chip with 16 GB of RAM. We observe that the generated network consists of one large connected component containing most of the nodes (see Figure S10(a)), along with some very small connected components. As the size of the graph increases (*i.e.*, as  $\eta$  grows), the proportion of nodes in the giant connected component also increases. The second-largest component is already of negligible size (see Figure S10(b)). As shown in Figure S11, apart from the largest component, the other connected components are of small size, with the large majority containing only two nodes.

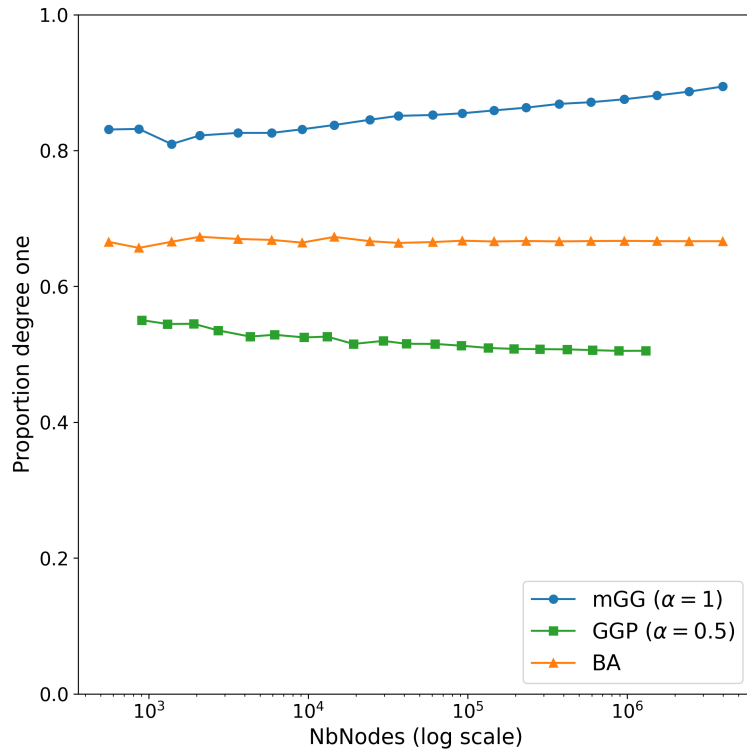


Figure S9: Proportion of nodes of degree one in the mGG model (●) with parameters  $\alpha = 1$ ,  $\tau = 0$ ,  $c = 1$ , and  $\beta = 1$  for various values of  $\eta$  ranging from 50 to 6000, resulting in graphs of different sizes. Comparison with the Generalised Gamma CRM (Caron and Fox, 2017) (■) with parameters  $\tau = 1$  and  $\sigma = 0.5$ , and with the Barabási–Albert model (Barabási and Albert, 1999) (▲). For every configuration we simulate 20 graph samples and plot the mean of the quantity of interest. These simulations were performed using the Python implementation available on Github running on an Apple M2 Chips with 16Gb of RAM.

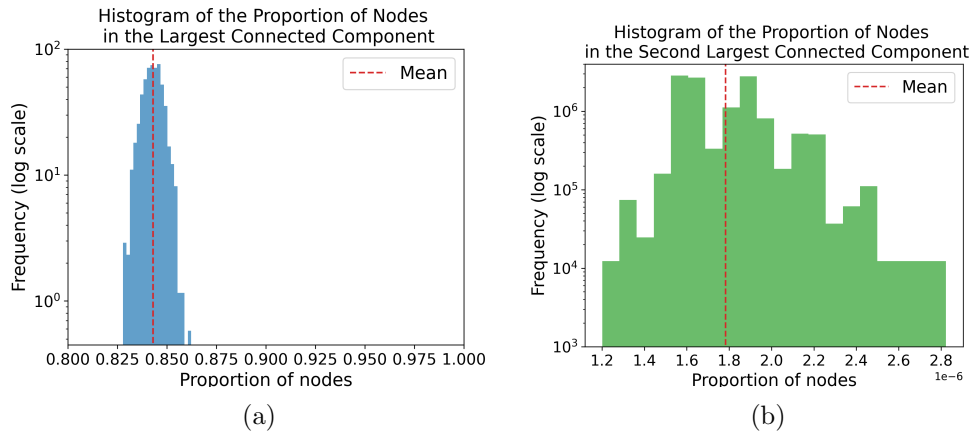


Figure S10: Proportion of nodes in the largest connected component (a), second largest connected component (b) for 1000 samples of graph generate with parameters  $\alpha = 0, \tau = 1, \beta = 1, c = 8$  and  $\eta = 800$ , the typical size of such a graph is 3 574 053 nodes. The mean is 0.842 for the largest connected component and  $1.78 \times 10^{-6}$  for the second largest. These simulations were performed using the Python implementation available on Github running on a standard computer (13th Gen Intel Core i5-1350P).

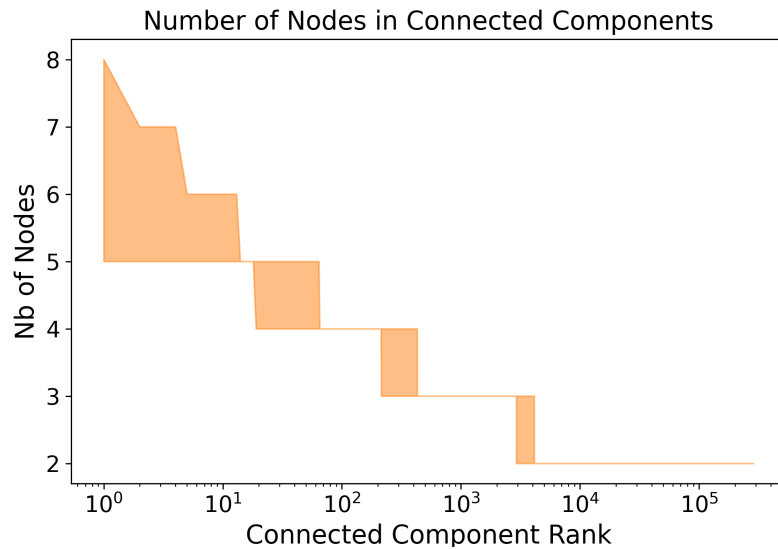


Figure S11: Number of nodes in a connected component with respect of its rank, where the connected components are ranked by decreasing size. The number of nodes in the largest connected component is not displayed. We have superposed the value for the same 1000 graphs than the ones generated for [Figure S10](#).

## S3.2 Inference

### S3.2.1 Details on the inference procedure

In this section, we provide additional details on the inference procedure from [Section 5.1](#). Note that the same procedure is used for real-world data. The weights  $w$  and latent  $s$  are updated via our HMC procedure with  $L = 5$  leapfrog steps. As suggested by [Betancourt \(2018\)](#), the HMC step sizes  $\epsilon_s$  and  $\epsilon_w$  are automatically adapted during the first half of the iterations to target an acceptance rate of 0.65, following the update rule:

$$\epsilon(t + 1) = \epsilon(t) \times \exp(0.005(\text{rate}(t) - 0.65)).$$

For the estimated hyperparameters, the standard deviations of the random walk Metropolis-Hastings steps for  $\log(\beta)$ ,  $\log(c)$  and  $\log(\eta)$  were set to 0.01, 0.01 and 0.02 respectively. The computing time for running each chain was about 14 hours using the Python implementation available on Github on a cluster (CPU : Intel Platinum 8628, 2.90GHz). The latent counts  $\tilde{q}_{ij}$  are updated only once every 30 iterations. Every other iteration, the parameter  $\eta$  is updated using a simple random walk Metropolis-Hastings step. A thinning procedure is applied to retain only 1,000 iterations per chain.

### S3.2.2 Diagnostics of convergence

To assess the convergence of our procedure, we use the Gelman-Rubin convergence diagnostic ([Gelman and Rubin, 1992](#)). The exact GR values for each hyperparameters are provided in [Table S4](#). Given the large number of parameters, we also report a multivariate version of the GR diagnostic, proposed by [Vats and Knudson \(2021\)](#) and defined as

$$R_{multi} = \sqrt{\frac{n_{it} - 1}{n_{it}} + \frac{(\prod_{i=0}^{n_{par}} X_i)^{\frac{1}{n_{par}}}}{n_{it}}}, \quad (\text{S37})$$

where  $n_{it}$  is the number of MCMC iteration after the burning phase,  $n_{par}$  is the number of estimated parameters, and  $X_i = n_{it} \times R_i^2 - n_{it} + 1$ , with  $R_i$  denoting the individual GR diagnostic for parameter  $i$ , where  $1 \leq i \leq n_{par}$ . In our case, we obtain  $R_{multi} = 1.001$ , suggesting the convergence of the algorithm.

Parameter	Value
$\beta$	1.004
$c$	1.045
$\eta$	1.045

Table S4: Values of the Gelman-Rubin convergence diagnostic.

## S3.3 Real World Data

In this section, we provide additional details on the inference performed on real-world data in [Section 5.2](#). To split the network data between a training set and a test set, we use  $p$ -sampling, as described by [Veitch and Roy \(2019\)](#). Statistics of the sampled graphs used for inference are given in [Table 3](#). The inference was performed on a standard computer (13th Gen Intel Core i5-1350P), each chain taking about 24 hours to be computed. For each graph, we run our inference algorithm for 2 million iterations. Apart from the number of iterations, we follow exactly the same procedure as for the synthetic data. We further

illustrate the inference procedure on the three datasets. Unlike in the synthetic case, the true parameter values are unknown. We present the trace plots of the hyperparameters (Figure S12, Figure S15, Figure S18) and selected weights (Figure S13, Figure S16, Figure S19), as well as the credible intervals for the 50 nodes with the highest degree and the 50 nodes with the lowest degree (Figure S14, Figure S17, Figure S20). We compute the same Gelman-Rubin (GR) convergence diagnostic as for the synthetic data for each of our 3 datasets in Table S5. In all three cases, the values suggest the convergence of the algorithm.

Parameter	Value	Parameter	Value	Parameter	Value
$\beta$	1.005	$\beta$	1.058	$\beta$	1.028
$c$	1.009	$c$	1.070	$c$	1.044
$\eta$	1.010	$\eta$	1.069	$\eta$	1.046
multi	1.000	multi	1.003	multi	1.000

(a) Flickr                                      (b) Douban                                      (c) TwitterCrawl

Table S5: Values of the Gelman-Rubin convergence diagnostic for the 3 datasets.

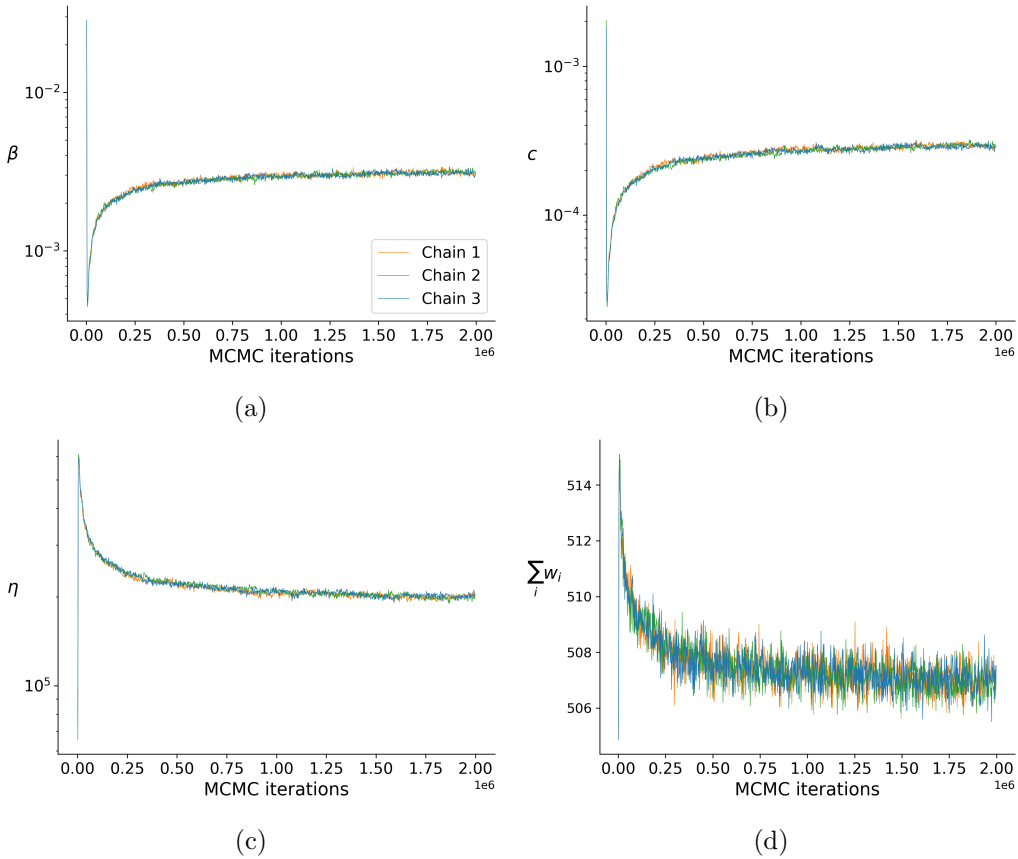


Figure S12: MCMC traceplot of parameters (a)  $\beta$ , (b)  $c$ , (c)  $\eta$  and (d) the total sum of the weights  $w$  for the Flickr subgraph

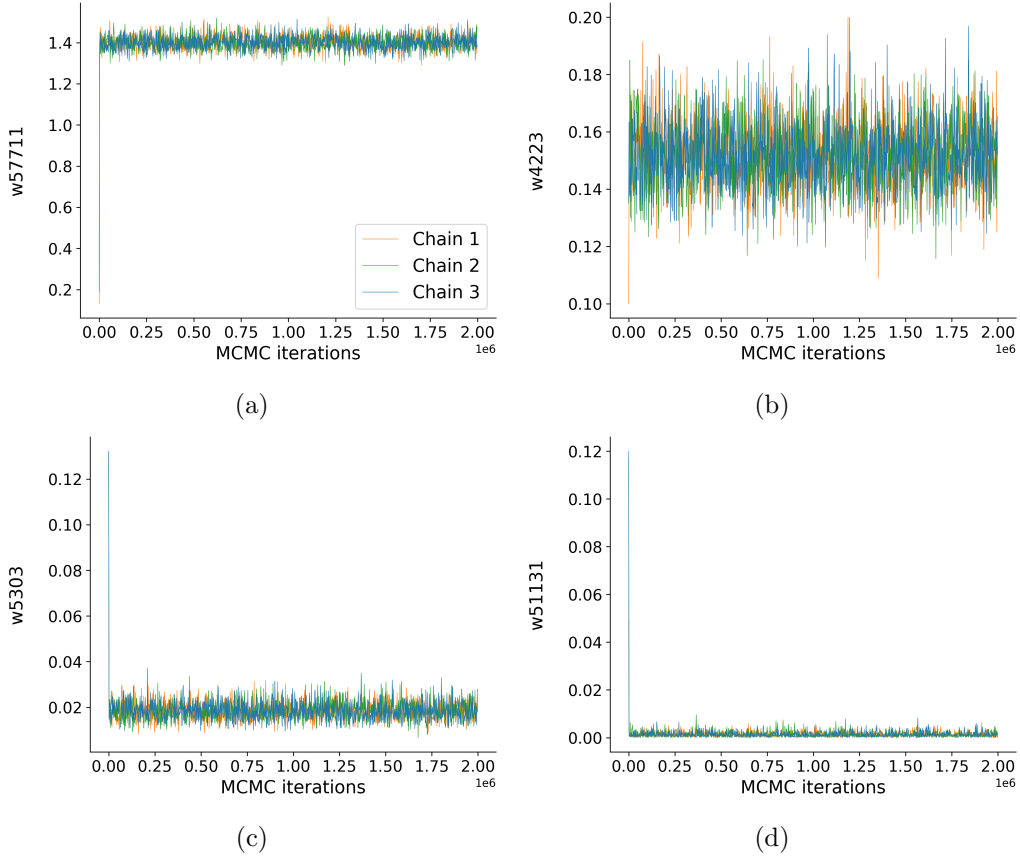


Figure S13: MCMC traceplot of four weights for the Flickr subgraph. The degree of the corresponding node are (a) 1445, (b) 166, (c) 21 and (d) 2.

## S4 Useful identities and results

### S4.1 Gautschi's inequality

For any  $x > 0$ , any  $s \in (0, 1)$ ,

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s}. \quad (\text{S38})$$

Taking  $x \in (0, 1)$  and  $s = 1 - x$ , we obtain, for any  $x \in (0, 1)$

$$\frac{1}{x^{1-x}} < \Gamma(x) < \frac{1}{(x+1)^{1-x}}. \quad (\text{S39})$$

For  $x > 0$ ,

$$x \leq 1 + x \log(x).$$

### S4.2 Lambert function

The LambertW function is defined to be the multivalued inverse of the function  $f(y) = ye^y$ . When dealing with real number only, the two branches suffice: for real number  $x$  and  $y$  the equation  $f(y) = x$  is given by

- $y = \text{LambertW}_0(x)$  for  $x \geq 0$ ,

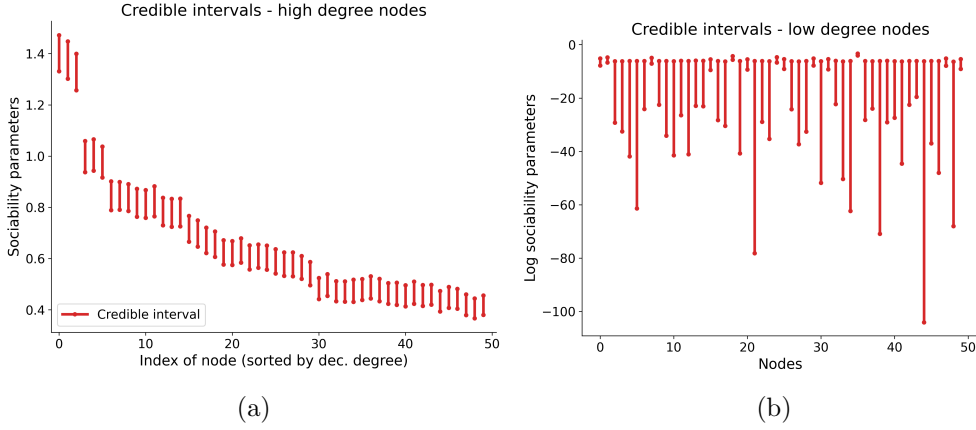


Figure S14: 95% posterior intervals of (a) the sociability parameters  $w_i$  of the 50 nodes with highest degree and (b) the log-sociability parameters  $\log(w_i)$  of the 50 nodes with lowest degree, for the Flickr subgraph.

- $y = \text{LambertW}_0(x)$  and  $y = \text{LambertW}_{-1}(x)$  if  $-\frac{1}{e} < x < 0$ ,

where  $\text{LambertW}_k$  is the  $k$ th branch of the Lambert function. The real branches of the LambertW function can't be expressed in terms of elementary function but some numerical methods exist (Lóczy, 2022; Iacono and Boyd, 2017; Corless et al., 1996).

The LambertW function satisfies the following identities.

$$\text{LambertW}(x)e^{\text{LambertW}(x)} = x, \quad (\text{S40})$$

$$\text{LambertW}_0(x \log(x)) = \log(x) \quad \text{for } \frac{1}{e} \leq x, \quad (\text{S41})$$

$$\text{LambertW}_{-1}(x \log(x)) = \log(x) \quad \text{for } 0 < x \leq \frac{1}{e}, \quad (\text{S42})$$

$$\text{LambertW}_0\left(-\frac{1}{e}\right) = -1. \quad (\text{S43})$$

### S4.3 Secondary lemmas

**Proposition S1.** For any  $z > 0$ , and  $0 \leq \tau < \alpha \leq 1$ , we have

$$\int_{\tau}^{\alpha} s z^s ds = \begin{cases} \frac{\alpha^2 - \tau^2}{2} & z = 1 \\ \frac{z^{\tau} - z^{\alpha} + (\alpha z^{\alpha} - \tau z^{\tau}) \log z}{(\log z)^2} & z \neq 1 \end{cases}$$

**Proof.** For  $z = 1$ , we have  $\int_{\tau}^{\alpha} s ds = \frac{\alpha^2 - \tau^2}{2}$ . For  $z \neq 1$ , using the change of variable  $u = z^s$ ,

$$\begin{aligned} \int_{\tau}^{\alpha} s z^s ds &= \frac{1}{\log^2 z} \int_{z^{\tau}}^{z^{\alpha}} \log u du \\ &= \frac{1}{\log^2 z} (z^{\alpha} \log z^{\alpha} - z^{\alpha} - [z^{\tau} \log z^{\tau} - z^{\tau}]) \\ &= \frac{z^{\tau} - z^{\alpha} + (\alpha z^{\alpha} - \tau z^{\tau}) \log z}{(\log z)^2}. \end{aligned}$$

□

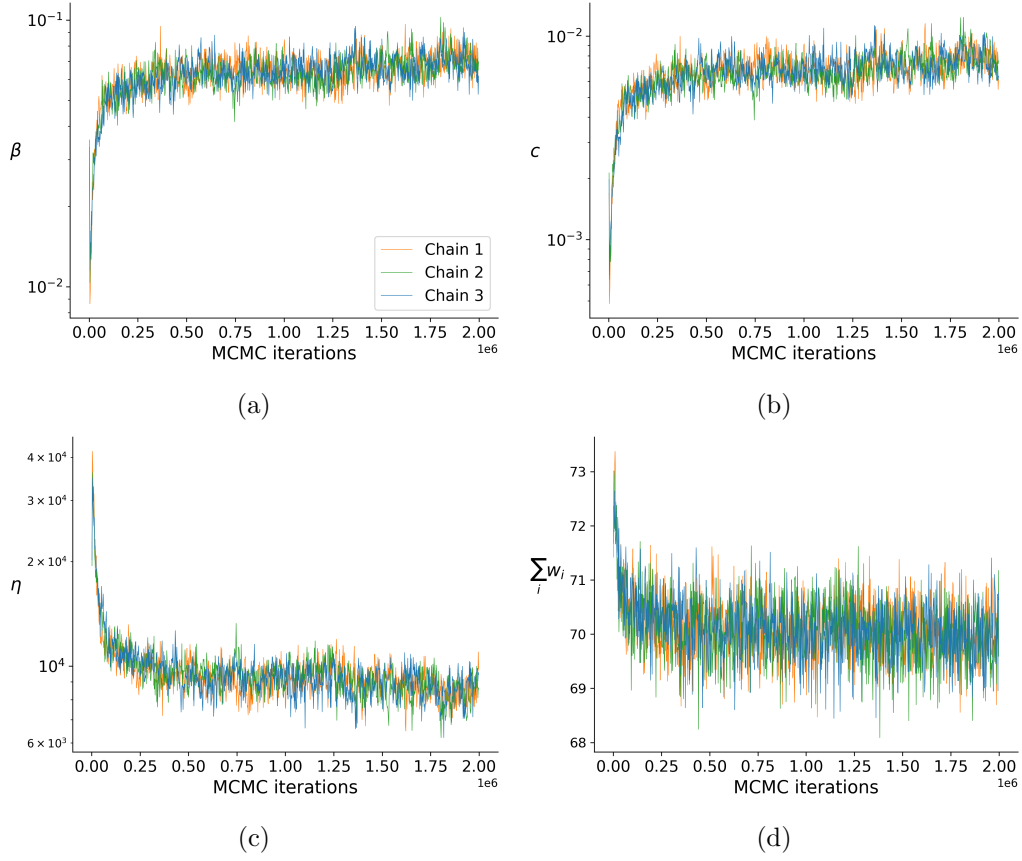


Figure S15: MCMC traceplot of parameters (a)  $\beta$ , (b)  $c$ , (c)  $\eta$  and (d) the total sum of the weights  $w$  for the Douban subgraph

**Proposition S2.** For  $0 \leq \tau < x < \alpha \leq 1$  and  $z > 0$ , let

$$F(x) = \frac{\int_{\tau}^x s z^s ds}{\int_{\tau}^{\alpha} s z^s ds}$$

Then

$$F(x) = \begin{cases} \frac{x^2 - \tau^2}{\alpha^2 - \tau^2} & z = 1, \\ \frac{z^{\tau} - z^x + (x z^x - \tau z^{\tau}) \log z}{z^{\tau} - z^{\alpha} + (\alpha z^{\alpha} - \tau z^{\tau}) \log z} & z \neq 1. \end{cases}$$

and

$$F^{-1}(y) = \begin{cases} \sqrt{(\alpha^2 - \tau^2)y + \tau^2} & z = 1, \\ \frac{\text{LambertW}_0(c(y)/e) + 1}{\log z} & z > 1, \\ \frac{\text{LambertW}_{-1}(c(y)/e) + 1}{\log z} & z < 1. \end{cases}$$

where

$$c(y) = (z^{\tau} - z^{\alpha} + (\alpha z^{\alpha} - \tau z^{\tau}) \log z)y - z^{\tau} + \tau z^{\tau} \log z, \quad (\text{S44})$$

and  $\text{LambertW}_k$  denotes the  $k$ th branch of the Lambert  $W$  function.

**Proof.** The expression for  $F$  follows from [Proposition S1](#). The inverse of  $F$  for  $z = 1$  is straightforward. For  $z \neq 1$ ,  $y = F(x)$  gives

$$\begin{aligned} c(y) &= x z^x \log z - z^x \\ &= e \times (x \log z - 1) e^{x \log z - 1}. \end{aligned}$$

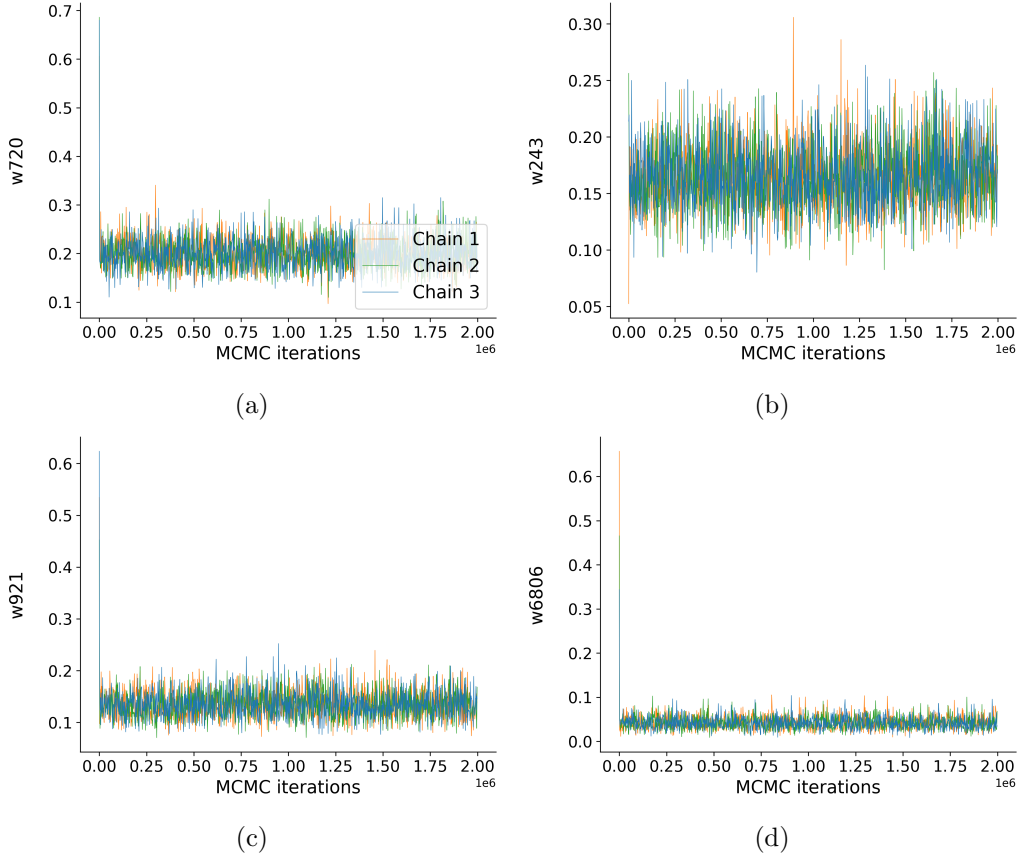


Figure S16: MCMC traceplot of four weights for the Douban subgraph. The degree of the corresponding node are (a) 41, (b) 28, (c) 34 and (d) 9.

where  $c(y) = (z^\tau - z^\alpha + (\alpha z^\alpha - \tau z^\tau) \log z)y - z^\tau + \tau z^\tau \log z$ . Hence we have

$$c(y)/e = f(x \log z - 1),$$

where  $f(w) = we^w$ . We have  $-1 < c(y) < 0$  hence there are two solutions to the equation (in  $x$ )  $c(y)/e = f(x)$ , which can be expressed in terms of the Lambert $W_0$  and Lambert $W_{-1}$  functions. If  $z > 1$ , then  $x \log z - 1 > -1$  for all  $x$ , hence it is expressed in terms of Lambert $W_0$ . Otherwise  $x \log z - 1 < -1$  hence it is expressed in terms of Lambert $W_{-1}$ . We therefore obtain

$$x = \begin{cases} \frac{\text{Lambert}W_0(c(y)/e)+1}{\log z} & z > 1, \\ \frac{\text{Lambert}W_{-1}(c(y)/e)+1}{\log z} & z < 1. \end{cases}$$

□

**Proposition S3.** *The inverse of the function  $\psi_{\text{mSt}}(y; \alpha, 0) = \frac{t^\alpha - 1}{\log t^\alpha}$  is*

$$\psi_{\text{mSt}}^{-1}(y; \alpha, 0) = \begin{cases} \left[ -y \text{Lambert}W_{-1} \left( -\frac{1}{y} e^{-1/y} \right) \right]^{1/\alpha} & \text{for } t > 1, \\ \left[ -y \text{Lambert}W_0 \left( -\frac{1}{y} e^{-1/y} \right) \right]^{1/\alpha} & \text{for } t \in (0, 1), \\ 1 & \text{for } t = 1. \end{cases}$$

where Lambert $W_k$  denotes the  $k$ th branch of the Lambert  $W$  function.

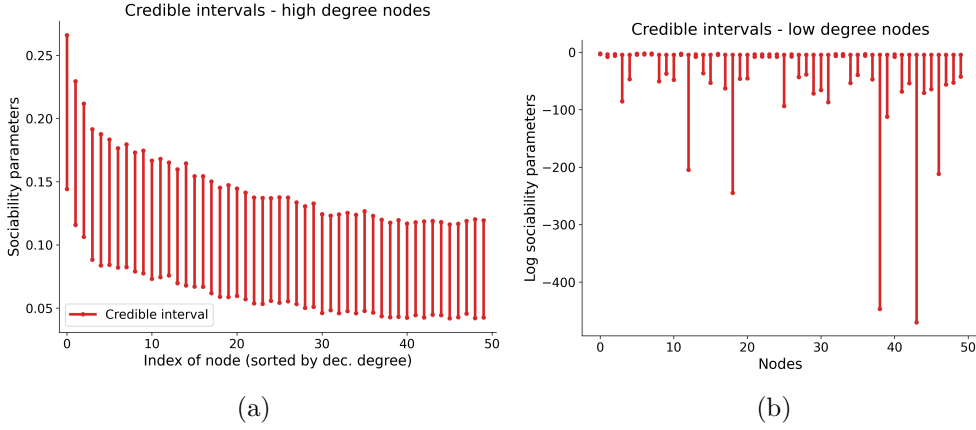


Figure S17: 95% posterior intervals of (a) the sociability parameters  $w_i$  of the 50 nodes with highest degree and (b) the log-sociability parameters  $\log(w_i)$  of the 50 nodes with lowest degree, for the Douban subgraph.

**Proof.** Consider the case  $\tau = 0$ . We are looking for the inverse of  $\psi_{\text{mSt}}(t; \alpha, 0)$ . Let  $y = \psi_{\text{mSt}}(t; \alpha, 0) = \frac{t^\alpha - 1}{\log t^\alpha}$ . This function is monotone increasing and has a well-defined inverse. We have

$$t^\alpha = e^{\frac{t^\alpha - 1}{y}}$$

and

$$-\frac{t^\alpha}{y} e^{-t^\alpha/y} = -\frac{1}{y} e^{-1/y}$$

and so  $-\frac{1}{y} e^{-1/y} = f(-\frac{t^\alpha}{y})$ . Note that  $-\frac{1}{y} e^{-1/y} \in [-1/e, 0)$ , when  $y > 0$ , where the minimum is achieved for  $y = 1$ . The set of solution is given by  $\text{LambertW}_0(x) \in [-1, 0)$  and  $\text{LambertW}_{-1}(x) \in [-1, -\infty)$ .

Note that  $y = \psi_{\text{mSt}}(t; \alpha, 0) < t^\alpha$  for  $t > 1$  as  $\log(1+x) \geq \frac{x}{x+1}$  for  $x > -1$ . So for  $t > 1$  we have

$$t = \left[ -y \text{LambertW}_{-1} \left( -\frac{1}{y} e^{-1/y} \right) \right]^{1/\alpha} = \psi_{\text{mSt}}^{-1}(y; \alpha, 0). \quad (\text{S45})$$

For  $t \in (0, 1)$  we have  $y > t^\alpha$  as  $x > x \ln(x) - 1$  if  $x \in (0, 1)$  so we have

$$t = \left[ -y \text{LambertW}_0 \left( -\frac{1}{y} e^{-1/y} \right) \right]^{1/\alpha} = \psi_{\text{mSt}}^{-1}(y; \alpha, 0). \quad (\text{S46})$$

Finally, for  $t = 1$  we have  $\psi_{\text{mSt}}(1; \alpha, 0) = 1$  so  $\psi_{\text{mSt}}^{-1}(1; \alpha, 0) = 1$ .  $\square$

## S5 Background material on regularly varying functions

Most of the background material in this section originates from the book of [Bingham et al. \(1987\)](#).

### S5.1 Definitions

**Definition S1** (Slowly varying function). *A function  $\ell : (0, \infty) \rightarrow (0, \infty)$  is slowly varying at infinity if for all  $c > 0$ ,*

$$\frac{\ell(cx)}{\ell(x)} \rightarrow 1 \text{ as } x \rightarrow \infty.$$

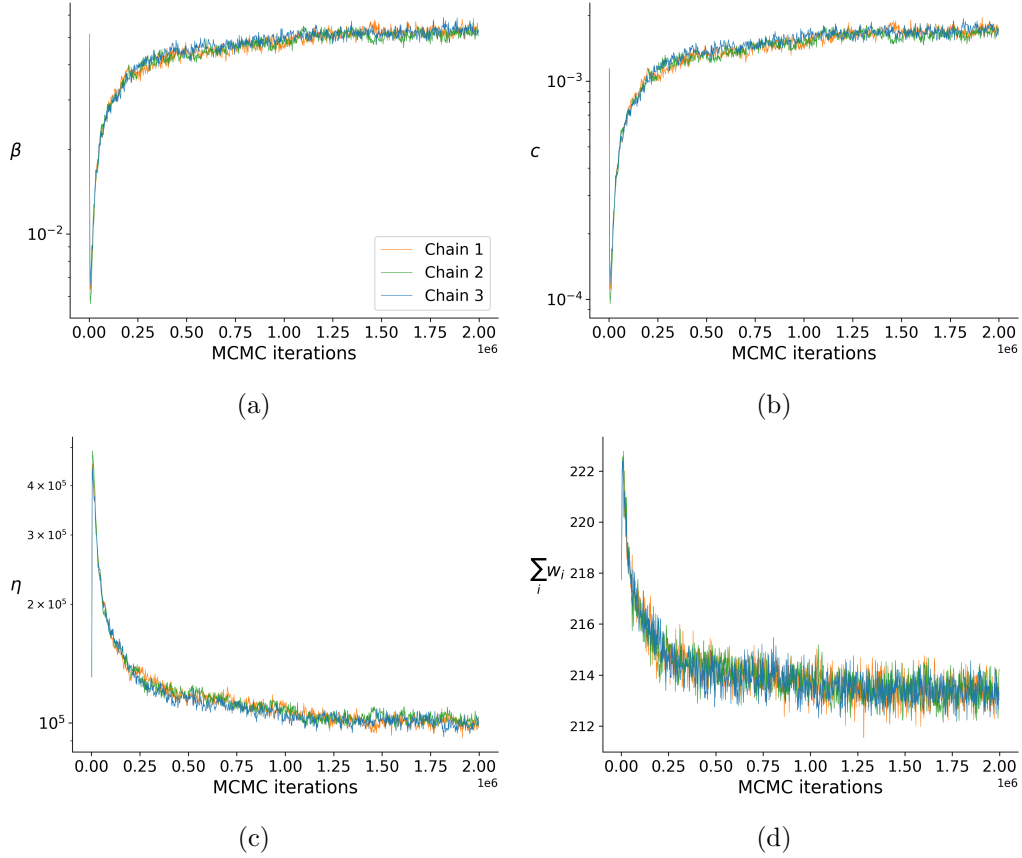


Figure S18: MCMC traceplot of parameters (a)  $\beta$ , (b)  $c$ , (c)  $\eta$  and (d) the total sum of the weights  $w$  for the TwitterCrawl subgraph

Examples of slowly varying functions include  $\log^a$ , for  $a \in \mathbb{R}$ , and functions converging to a constant  $c > 0$ .

**Definition S2** (Regularly varying function). *A function  $\ell : (0, \infty) \rightarrow (0, \infty)$  is regularly varying at infinity with exponent  $\rho \in \mathbb{R}$  if  $f(x) = x^\rho \ell(x)$  for some slowly varying function  $\ell$ . A function  $f$  is regularly varying at 0 if  $f(1/x)$  is regularly varying at infinity that is,  $f(x) = x^{-\rho} \ell(1/x)$  for some  $\rho \in \mathbb{R}$ .*

## S5.2 Karamata theorems

The following propositions and corollaries relate to integrals of regularly varying functions.

**Proposition S3** (Karamata theorem). *See (Bingham et al., 1987, Propositions 1.5.8 and 1.5.10). Let  $U(t) = t^\rho \ell(t)$  for some locally bounded slowly varying function  $\ell$ . Then*

(i) If  $\rho > -1$

$$\int_0^x U(t) dt \sim \frac{1}{1+\rho} x^{\rho+1} \ell(x) \text{ as } x \rightarrow \infty.$$

(ii) If  $\rho < -1$

$$\int_x^\infty U(t) dt \sim -\frac{1}{1+\rho} x^{\rho+1} \ell(x) \text{ as } x \rightarrow \infty.$$

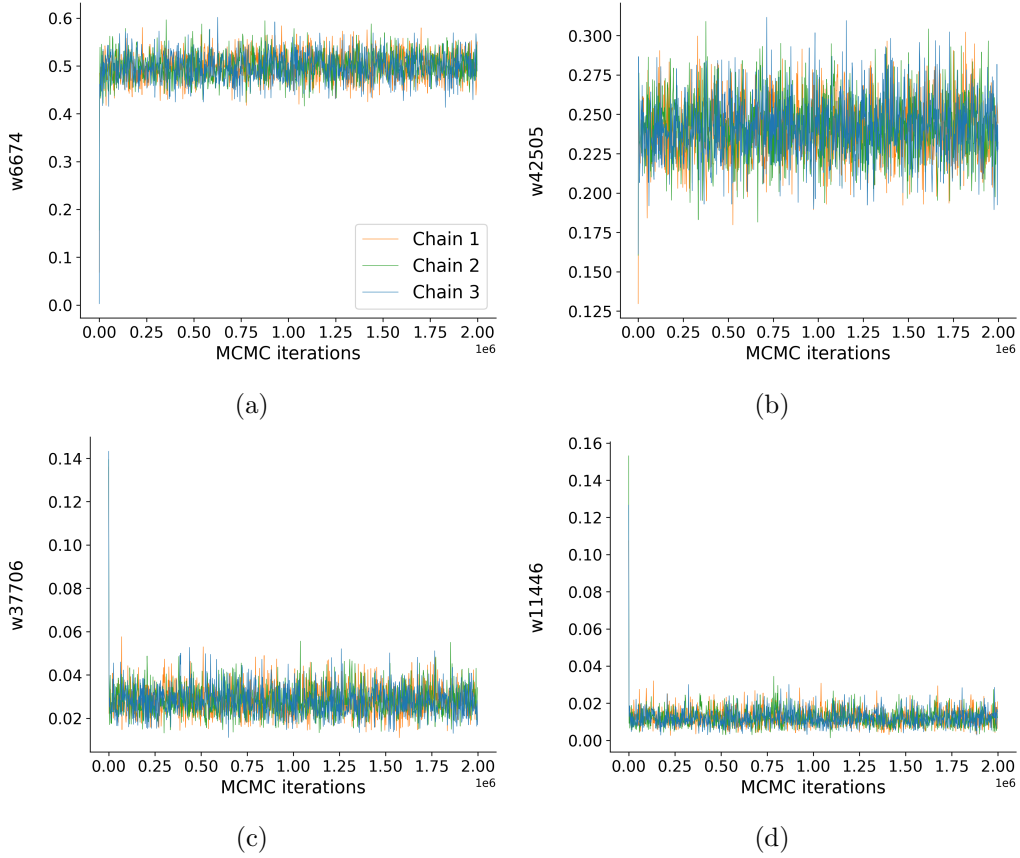


Figure S19: MCMC traceplot of four weights for the TwitterCrawl subgraph. The degree of the corresponding node is (a) 309, (b) 151, (c) 18 and (d) 8.

The following corollaries will be useful.

**Corollary S4.** Let  $U(x) = x^\alpha \ell(1/x)$  for some locally bounded slowly varying function  $\ell$ .

(i) If  $\alpha > -1$

$$\int_0^x U(t) dt \sim \frac{1}{\alpha + 1} x^{1+\alpha} \ell(1/x) \text{ as } x \rightarrow 0.$$

(ii) If  $\alpha < -1$

$$\int_x^\infty U(t) dt \sim -\frac{1}{\alpha + 1} x^{1+\alpha} \ell(1/x) \text{ as } x \rightarrow 0.$$

If  $\ell$  is slowly varying and  $\alpha > -1$  then  $\int_0^x t^\alpha \ell(1/t) dt$  converges and

$$\frac{x^{1+\alpha} \ell(1/x)}{\int_0^x t^\alpha \ell(1/t) dt} \rightarrow \alpha + 1 \text{ as } x \rightarrow 0. \quad (\text{S47})$$

**Proof.** Let  $\rho < -1$ . We have  $\int_x^\infty t^\rho \ell(t) dt = \int_0^{1/x} t^{-\rho-2} \ell(1/t) dt$ . Writing  $\alpha = -\rho - 2 > -1$ , we obtain, using [Proposition S3\(ii\)](#)

$$\frac{x^{-\alpha-1} \ell(x)}{\int_0^{1/x} t^\alpha \ell(1/t) dt} \rightarrow \rho + 1 \text{ as } x \rightarrow \infty, \quad (\text{S48})$$

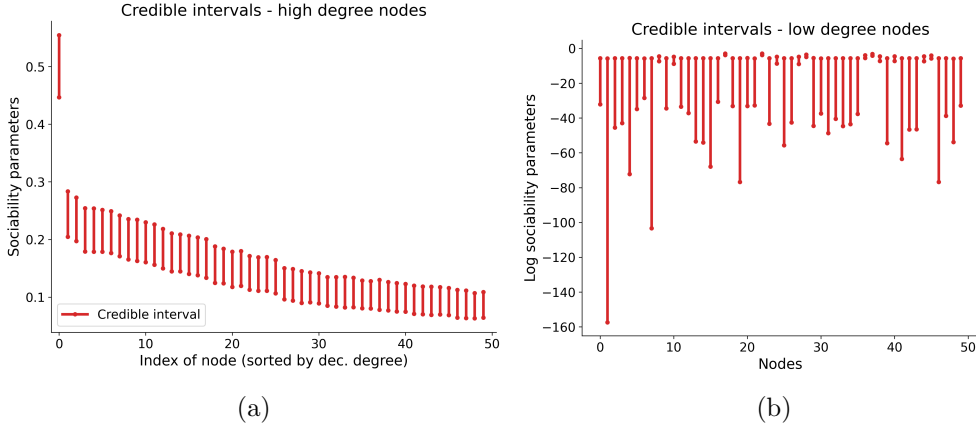


Figure S20: 95% posterior intervals of (a) the sociability parameters  $w_i$  of the 50 nodes with highest degree and (b) the log-sociability parameters  $\log(w_i)$  of the 50 nodes with lowest degree, for the TwitterCrawl subgraph.

or

$$\frac{x^{1+\alpha}\ell(1/x)}{\int_0^x t^\alpha \ell(1/t) dt} \rightarrow \rho + 1 \text{ as } x \rightarrow 0. \quad (\text{S49})$$

The case  $\alpha < -1$  follows similarly, using [Proposition S3\(i\)](#).  $\square$

### S5.3 Asymptotic inverse

If  $f$  is regularly varying then there exists a regularly varying function  $g$  such that  $f(g(x)) \sim g(f(x)) \sim x$  as  $x \rightarrow \infty$ . We say that  $g$  is the asymptotic inverse of  $f$ . It is uniquely determined within asymptotic equivalence, and one version of  $g$  is the generalised inverse of  $f$ .

**Theorem S5.** ([Bingham et al., 1987, Proposition 1.5.13](#)) *If  $\ell$  varies slowly, there exist a slowly varying function  $\ell^\#$ , unique up to asymptotic equivalent, with*

$$\ell(x)\ell^\#(x\ell(x)) \xrightarrow{x \rightarrow \infty} 1 \text{ and } \ell^\#(x)\ell(x\ell^\#(x)) \xrightarrow{x \rightarrow \infty} 1.$$

The function  $\ell^\#$  is the Bruijn conjugate of  $\ell$ ;  $(\ell, \ell^\#)$  is a conjugate pair. Such pairs are of common occurrence, for instance the de Bruijn conjugate of  $\log^a$  is  $\log^{-a}$  for any  $a \in \mathbb{R}$ .

**Proposition S6.** ([Bingham et al., 1987, Proposition 1.5.14](#)) *If  $(\ell, \ell^\#)$  is a conjugate pair,  $A, B, a > 0$  each of the following is also a pair*

$$\begin{aligned} & (\ell(Ax), \ell^\#(Ax)), \\ & (A\ell(x), A^{-1}\ell^\#(x)), \\ & ([\ell(x^a)]^{\frac{1}{a}}, [\ell^\#(x^a)]^{\frac{1}{a}}). \end{aligned}$$

The following is a special case from ([Bingham et al., 1987, Proposition 1.5.15 p.29](#)).

**Proposition S7.** ([Bingham et al., 1987, Proposition 1.5.15](#)) *Let  $a, b > 0$ . Let  $f(x) \sim x^{ab}\ell^a(x^b)$  where  $\ell$  is slowly varying, and let  $g$  be the asymptotic inverse of  $f$ . Then*

$$g(x) \sim x^{\frac{1}{ab}} [\ell^\#]^{\frac{1}{b}}(x^{\frac{1}{a}}) \text{ as } x \rightarrow \infty.$$

## S5.4 Tauberian theorem

The following theorem is a variation of [Bingham et al. \(1987, Theorem 1.7.6 p.46\)](#), where the two limits at 0 and infinity are exchanged. The proof is similar. See also [Feller \(1971, Chapter XIII\)](#).

**Theorem S8** (Tauberian theorem). *Assume  $U(x) \geq 0$ ,  $c \geq 0$ ,  $\rho > -1$ ,  $\widehat{U}(s) = s \int_0^\infty e^{-sx} U(x) dx$  convergent for  $s > 0$ , and  $\ell$  a slowly varying function. Then*

$$U(x) \sim cx^\rho \ell(1/x) / \Gamma(1 + \rho) \text{ as } x \rightarrow 0 \quad (\text{S50})$$

implies

$$\widehat{U}(s) \sim cs^{-\rho} \ell(s) \text{ as } s \rightarrow \infty. \quad (\text{S51})$$

**Proof.** Write  $V(x) = \int_0^x U(y) dy$  (this is finite for any  $x$  as  $\widehat{U}(s)$  is convergent for any  $s$ ), then  $V$  is non-decreasing and by [Corollary S4](#)

$$V(x) \sim \frac{c}{\rho + 1} x^{\rho+1} \ell(1/x) / \Gamma(1 + \rho) \text{ as } x \rightarrow 0. \quad (\text{S52})$$

Then by [Bingham et al. \(1987, Theorem 1.7.1, p.38\)](#), this is equivalent to

$$\widehat{V}(s) = \int_0^\infty e^{-sx} dV(x) \sim cs^{-\rho-1} \ell(s) \text{ as } s \rightarrow \infty. \quad (\text{S53})$$

Finally, note that  $\widehat{V}(s) = \frac{\widehat{U}(s)}{s}$ . Thus the above equation is equivalent to

$$\widehat{U}(s) \sim cs^{-\rho} \ell(s) \text{ as } s \rightarrow \infty. \quad (\text{S54})$$

□

## S6 Size-biased representation of a CRM

**Lemma S1.** *Let  $\rho_0$  be some Lévy intensity, with Laplace exponent  $\psi_0$  and inverse Laplace exponent  $\psi_0^{-1}$ . Let  $\rho_1(w) = \frac{\eta}{c} \rho_0(\frac{w}{c}) e^{-\beta w/c}$ , with Laplace exponent  $\psi_1$ . Then*

$$\begin{aligned} \psi_1(t) &= \eta [\psi_0(ct + \beta) - \psi_0(\beta)] \\ \psi_1^{-1}(t) &= \frac{1}{c} \left[ \psi_0^{-1} \left( \frac{t}{\eta} + \psi_0(\beta) \right) - \beta \right]. \end{aligned}$$

**Proof.**

$$\begin{aligned} \psi_1(t) &= \int_0^\infty (1 - e^{-wt}) \rho_1(w) dw \\ &= \frac{\eta}{c} \int_0^\infty (1 - e^{-wt}) \rho_0\left(\frac{w}{c}\right) e^{-\beta w/c} dw \\ &= \eta \int_0^\infty (1 - e^{-uct}) \rho_0(u) e^{-\beta u} du \\ &= \eta \left[ \int_0^\infty (1 - e^{-u(\beta+ct)}) \rho_0(u) du - \int_0^\infty (1 - e^{-\beta u}) \rho_0(u) du \right] \\ &= \eta [\psi_0(ct + \beta) - \psi_0(\beta)]. \end{aligned}$$

□

The following proposition follows from [Perman et al. \(1992\)](#); see also [Campbell et al. \(2019\)](#); [Lee et al. \(2023\)](#).

**Proposition S2.** (*Perman et al., 1992*) Consider a homogeneous CRM  $G \sim \text{CRM}(\rho, H)$  on  $\Theta$  with Lévy intensity  $\rho(w) = \frac{\eta}{c} \rho_0(\frac{w}{c}) e^{-\beta w/c}$ , where  $\rho_0$  is some base Lévy intensity, with associated Laplace exponent  $\psi_0(t) = \int_0^\infty (1 - e^{-wt}) \rho_0(w) dw$  and tilted moment function  $\kappa_0(m, z) = \int_0^\infty w^m e^{-zw} \rho_0(w) dw$ . The Laplace exponent of the CRM  $G$  takes the form

$$\psi(t) = \eta [\psi_0(ct + \beta) - \psi_0(\beta)].$$

Let  $\psi_0^{-1}$  be the inverse of  $\psi_0$ . Let  $\xi_1, \xi_2, \dots$  be the points of a unit-rate Poisson process on  $(0, \infty)$ . The size-biased representation of the CRM

$$G = \sum_{j \geq 1} W_j \delta_{\theta_j}$$

is given, for  $j \geq 1$ , by  $\theta_j \sim H$  and

$$T_j = \psi_0^{-1} \left( \frac{\xi_j}{\eta} + \psi_0(\beta) \right) - \beta$$

$$W'_j | \{T_j = t\} \sim p(w | t) = \frac{w e^{-(t+\beta)w} \rho_0(w)}{\kappa_0(1, t + \beta)}$$

$$W_j = cW'_j.$$

## Bibliography

- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434*.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*, volume 27. Cambridge university press.
- Campbell, T., Huggins, J. H., How, J. P., and Broderick, T. (2019). Truncated random measures. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 25(2):1256–1288.
- Caron, F. and Fox, E. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79:1295–1366.
- Caron, F., Panero, F., and Rousseau, J. (2023). On sparsity, power-law, and clustering properties of graphex processes. *Advances in Applied Probability*, 55(4):1211–1253.
- Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the LambertW function. *Advances in Computational Mathematics*, 5(1):329–359.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472.
- Iacono, R. and Boyd, J. P. (2017). New approximations to the principal real-valued branch of the Lambert W-function. *Advances in Computational Mathematics*, 43(6):1403–1436.

- Lee, J., Miscouridou, X., and Caron, F. (2023). A unified construction for series representations and finite approximations of completely random measures. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 29(3):2142–2166.
- Lóczy, L. (2022). Guaranteed- and high-precision evaluation of the Lambert W. *Applied Mathematics and Computation*, 433:127406.
- Neal, R. M. (2011). *MCMC Using Hamiltonian Dynamics*, pages 113–162. Chapman and Hall/CRC, New York, 1 edition.
- Perman, M., Pitman, J., and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39.
- Vats, D. and Knudson, C. (2021). Revisiting the Gelman–Rubin Diagnostic. *Statistical Science*, 36(4):518–529.
- Veitch, V. and Roy, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *The Annals of Statistics*, 47(6):3274 – 3299.