

# Truncated simulation and inference in edge-exchangeable networks\*

Xinglong Li and Trevor Campbell

*Department of Statistics  
The University of British Columbia*

**Abstract:** Edge-exchangeable probabilistic network models generate edges as an i.i.d. sequence from a discrete measure, providing a simple means for statistical inference of latent network properties. The measure is often constructed using the self-product of a realization from a Bayesian nonparametric (BNP) discrete prior; but unlike in standard BNP models, the self-product measure prior is not conjugate the likelihood, hindering the development of exact simulation and inference algorithms. Approximation via finite truncation of the discrete measure is a straightforward alternative, but incurs an unknown approximation error. In this paper, we develop methods for forward simulation and posterior inference in random self-product-measure models based on truncation, and provide theoretical guarantees on the quality of the results as a function of the truncation level. The techniques we present are general and extend to the broader class of discrete Bayesian nonparametric models.

**Keywords and phrases:** Truncation, Bayesian nonparametrics, edge-exchangeable, networks, Bayesian inference.

Received May 2020.

## 1. Introduction

Probabilistic generative models have for many years been key tools in the analysis of network data [1, 2]. Recent work in the area [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] has begun to incorporate the use of nonparametric discrete measures, in an effort to address the limitations of traditional models in capturing the sparsity of real large-scale networks [14]. These models construct a discrete random measure  $\Theta$  (often a completely random measure, or CRM [15]) on a space  $\Psi$ , associate each atom of the measure with a vertex in the network, and then use the *self-product* of the measure—i.e., the measure  $\Theta \times \Theta$  on  $\Psi^2$ —to represent the magnitude of interaction between vertices.

While the inclusion of a nonparametric measure enables capturing sparsity, it also makes both generative simulation and posterior inference via Markov chain Monte Carlo (MCMC) [16; 17, Ch. 11, 12] significantly more challenging. In standard Bayesian models with discrete nonparametric measures—such as the Dirichlet process mixture model [18] or beta process latent feature model [19]—this issue is typically addressed by exploiting the conjugacy of the (normalized)

---

\*This work is supported by a National Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Discovery Launch Supplement.

completely random measure prior and the likelihood to marginalize the latent infinite discrete measure [20]. But in the case of nonparametric network models, however, there is no such reprieve; the self-product of a completely random measure is not a completely random measure, and exact marginalization is typically not possible.

Another option is to *truncate* the discrete CRM to have finitely many atoms, and perform simulation and inference based on the truncated CRM. Exact truncation schemes based on auxiliary variables [21, 22, 23] are limited to models where certain tail probabilities can be computed exactly. Fixed truncation [24, 25, 26, 27], on the other hand, is easy to apply to any CRM-based model; but it involves an approximation with potentially unknown error. Although the approximation error of truncated CRM models has been thoroughly studied in past work [27, 28, 29, 30, 31], these results apply only to generative simulation—i.e., not inference—and do not apply to self-product CRMs that commonly appear in network models.

In this work, we provide tractable methods for both generative simulation and posterior inference of discrete Bayesian nonparametric models based on truncation, as well as rigorous theoretical analyses of the error incurred by truncation in both scenarios. In particular, our theory and methods require only the ability to compute bounds on—instead of exact evaluation of—intractable tail probabilities. Our work focuses on the case of self-product-measure-based *edge-exchangeable* network sequences [3, 5, 32, 33], whose edges are simulated i.i.d. conditional on the discrete random product measure  $\Theta \times \Theta$ , but the ideas here apply without much effort to the broader class of discrete Bayesian nonparametric models. As an intermediate step of possible independent interest, we also show that the nonzero rates generated from the *rejection representation* [34] of a Poisson process have the same distribution as the well-known but typically intractable *inverse Lévy* or *Ferguson-Klass* representation [35]. This provides a novel method for simulating the inverse Lévy representation, which has a wide variety of uses in applications of Poisson processes [36, 37, 38].

## 2. Background

### 2.1. Completely random measures and self-products

A *completely random measure* (CRM)  $\Theta$  on  $\Psi$  is a random measure such that for any collection of  $K \in \mathbb{N}$  disjoint measurable sets  $A_1, \dots, A_K \subset \Psi$ , the values  $\Theta(A_1), \dots, \Theta(A_K)$  are independent random variables [15]. In this work, we focus on discrete CRMs taking the form  $\Theta = \sum_k \theta_k \delta_{\psi_k}$ , where  $\delta_x$  is a Dirac measure on  $\Psi$  at location  $x \in \Psi$  (i.e.,  $\delta_x(A) = 1$  if  $x \in A$  and 0 otherwise), and  $(\theta_k, \psi_k)_{k=1}^\infty$  are a sequence of *rates*  $\theta_k$  and *labels*  $\psi_k$  generated from a Poisson process on  $\mathbb{R}_+ \times \Psi$  with mean measure  $\nu(d\theta) \times L(d\psi)$ . Here  $L$  is a diffuse probability measure, and  $\nu$  is a  $\sigma$ -finite measure satisfying  $\nu(\mathbb{R}_+) = \infty$ , which guarantees that the Poisson process has countably infinitely many points almost surely. The space  $\Psi$  and distribution  $L$  will not affect our analysis; thus as a shorthand, we

write  $\text{CRM}(\nu)$  for the distribution of  $\Theta$ :

$$\Theta := \sum_k \theta_k \delta_{\psi_k} \sim \text{CRM}(\nu). \tag{1}$$

One can construct a multidimensional measure  $\Theta^{(d)}$  on  $\Psi^d$ ,  $d \in \mathbb{N}$  from  $\Theta$  defined in Eq. (1) by taking its *self-product*. In particular, we define

$$\Theta^{(d)} := \sum_{i \in \mathbb{N}_{\neq}^d} \vartheta_i \delta_{\zeta_i}, \quad \vartheta_i := \prod_{j=1}^d \theta_{i_j}, \quad \zeta_i := (\psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_d}), \tag{2}$$

where  $i$  is a  $d$ -dimensional multi-index, and  $\mathbb{N}_{\neq}^d$  is the set of such indices with all distinct components. Note that  $\Theta^{(d)}$  is no longer a CRM, as it does not satisfy the independence condition.

### 2.2. Series representations

To simulate a realization  $\Theta \sim \text{CRM}(\nu)$ —e.g., as a first step in the simulation of a self-product measure  $\Theta^{(d)}$ —the rates  $\theta_k$  and labels  $\psi_k$  may be generated in sequence using a *series representation* [39] of the CRM. In particular, we begin by simulating the jumps of a unit-rate homogeneous Poisson process  $(\Gamma_k)_{k=1}^\infty$  on  $\mathbb{R}_+$  in increasing order. For a given distribution  $g$  on  $\mathbb{R}_+$  and nonnegative measurable function  $\tau : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , we set

$$\Theta = \sum_{k=1}^\infty \theta_k \delta_{\psi_k}, \quad \theta_k = \tau(U_k, \Gamma_k), \quad U_k \stackrel{\text{i.i.d.}}{\sim} g, \quad \psi_k \stackrel{\text{i.i.d.}}{\sim} L. \tag{3}$$

Depending on the particular choice of  $\tau$  and  $g$ , one can construct several different series representations of a CRM [28]. For example, the *inverse Lévy representation* [35] has the form

$$\theta_k = \nu^\leftarrow(\Gamma_k), \quad \nu^\leftarrow(x) := \inf \{y : \nu([y, \infty)) \leq x\}. \tag{4}$$

In many cases, computing  $\nu^\leftarrow(x)$  is intractable, making it hard to generate  $\theta_k$  in this manner. Alternatively, we can generate a series of rates from  $\text{CRM}(\nu)$  with the *rejection representation* [34], which has the form

$$\theta_k = T_k \mathbb{1}\left(\frac{d\nu}{d\mu}(T_k) \geq U_k\right), \quad T_k = \mu^\leftarrow(\Gamma_k), \quad U_k \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[0, 1], \tag{5}$$

where  $\mu$  is a measure on  $\mathbb{R}_+$  chosen such that  $\frac{d\nu}{d\mu} \leq 1$  uniformly and  $\mu^\leftarrow(x)$  is easy to calculate in closed-form. While there are many other sequential representations of CRMs [28], the representations in Eqs. (3) to (5) are broadly applicable and play a key role in our theoretical analysis.

### 2.3. Edge-exchangeable graphs

Self-product measures  $\Theta^{(d)}$  of the form Eq. (2) with  $d = 2$  have recently been used as priors in a number of probabilistic network models [3, 4].<sup>1</sup> The focus of the present work are those models that associate each  $\psi_k$  with a vertex, each tuple  $\zeta_i = (\psi_{i_1}, \dots, \psi_{i_d})$  with a (hyper)edge, and then build a growing sequence of networks by sequentially generating edges from  $\Theta^{(d)}$  in rounds  $n = 1, \dots, N$ . There are a number of choices for how to construct such a sequence. For example, in each round  $n$  we may add multiple edges via an *independent likelihood process*  $X_n \sim \text{LP}(h, \Theta^{(d)})$  defined by

$$X_n := \sum_{i \in \mathbb{N}_{\neq}^d} x_{ni} \delta_{\zeta_i}, \quad x_{ni} \stackrel{\text{indep}}{\sim} h(\cdot | \vartheta_i), \quad (6)$$

where  $x_{ni} = k$  denotes that there were  $k$  copies of edge  $\zeta_i$  added at round  $n$ , and  $h(\cdot | \vartheta)$  is a probability distribution on  $\mathbb{N} \cup \{0\}$ . We denote the mean  $\mu(\vartheta) := \sum_{k=0}^{\infty} k \cdot h(k | \vartheta)$  and probability of 0 under  $h$  to be  $\pi(\vartheta) := h(0 | \vartheta)$  for convenience. By the Slivnyak-Mecke theorem [40], if  $h$  satisfies

$$\int_{\mathbb{R}_+^d} \mu \left( \prod_{j=1}^d \theta_j \right) \prod_{j=1}^d \nu(d\theta_j) < \infty, \quad (7)$$

then finitely many edges are added to the graph in each round. We make this assumption throughout this work. Alternatively, if

$$\int_{\mathbb{R}_+} \min(1, \theta) \nu(d\theta) < \infty,$$

then  $\Omega := \Theta^{(d)}(\Psi^d) < \infty$ , and we may add only a single edge per round  $n$  via a *categorical likelihood process*  $X_n \sim \text{Categorical}(\Theta^{(d)})$  defined by

$$X_n := \delta_{\zeta_{I_n}}, \quad I_n \sim \text{Categorical} \left( \left( \frac{\vartheta_i}{\Omega} \right)_{i \in \mathbb{N}_{\neq}^d} \right). \quad (8)$$

This construction has appeared in [4], where  $\Theta$  follows a Dirichlet process, which can be seen as a normalized gamma process [41]. Note that our definition of  $\text{Categorical}(\Theta^{(d)})$  involves simulating from the normalization; we use this definition to avoid introducing new notation for normalized processes.

Using either likelihood process, the edges in the network after  $N$  rounds are

$$\sum_{n=1}^N X_n := \sum_{i \in \mathbb{N}_{\neq}^d} x_i \delta_{\zeta_i}, \quad x_i := \sum_{n=1}^N x_{ni},$$

<sup>1</sup>There are also network models based on self-product measure priors that do not generate edge-exchangeable sequences [8, 11]; it is likely that many of the techniques in the present work would extend to these models, but we leave the study of this to future work.

i.e.,  $x_i \in \mathbb{N} \cup \{0\}$  represents the count of edge  $i$  after  $N$  rounds.

There are three points to note about this formulation. First, since the atom locations  $\zeta_i$  are not used, we can represent the network using only its array of edge counts

$$E_N := (x_i)_{i \in \mathbb{N}_{\neq}^d} \in \mathcal{N}_d, \quad (9)$$

where  $\mathcal{N}_d$  denotes the set of integer arrays indexed by  $\mathbb{N}_{\neq}^d$  with finitely many nonzero entries. Note that  $\mathcal{N}_d$  is a countable set. Second, by construction, the distribution of  $E_N$  is invariant to reorderings of the arrival of edges, and thus the network is *edge-exchangeable* [3, 5, 6, 7]. Finally, note that the network  $E_N$  as formulated in Eq. (9) is in general a directed multigraph with no self-loops (due to the restriction to indices  $i \in \mathbb{N}_{\neq}^d$  rather than  $\mathbb{N}^d$ ). Although the main theoretical results in this work are developed in this setting, we provide an additional result in Section 3.1 to translate to other common network structures (e.g. binary undirected networks).

### 3. Truncated generative simulation

In this section, we consider the tractable generative simulation of network models via truncation, and analyze the approximation error incurred in doing so as a function of  $K \in \mathbb{N}$  (the truncation level) and number of rounds of generation  $N \in \mathbb{N}$ . In particular, to construct a truncated self-product measure, we first split the underlying CRM  $\Theta$  into a truncation and tail component,

$$\Theta = \Theta_K + \Theta_{K+}, \quad \Theta_K = \sum_{k=1}^K \theta_k \delta_{\psi_k}, \quad \Theta_{K+} = \sum_{k=K+1}^{\infty} \theta_k \delta_{\psi_k},$$

and construct the self-product  $\Theta_K^{(d)}$  from the truncation  $\Theta_K$  as in Eq. (2). Fig. 1 provides an illustration of the truncation of  $\Theta$  and  $\Theta^{(2)}$ , showing that  $\Theta^{(2)}$  can be decomposed into a sum of four parts,

$$\Theta^{(2)} = \Theta^2 = (\Theta_K + \Theta_{K+})^2 = \Theta_K^{(2)} + (\Theta_K \times \Theta_{K+} + \Theta_{K+} \times \Theta_K + \Theta_{K+}^2).$$

Thus, while we only discard  $\Theta_{K+}$  in truncating  $\Theta$  to  $\Theta_K$ , we discard three parts in truncating  $\Theta_K^{(2)}$  to  $\Theta_K^{(2)}$ ; and in general, we discard  $2^d - 1$  parts of  $\Theta^{(d)}$  when truncating it to  $\Theta_K^{(d)}$ . We therefore intuitively might expect higher truncation error when approximating  $\Theta_K^{(d)} \approx \Theta^{(d)}$  than when approximating  $\Theta_K \approx \Theta$ ; in Sections 3.2 and 3.3, we will show that this is indeed the case.

Once the measure is truncated, a truncated network—based on  $\Theta_K^{(d)}$ —can be constructed in the same manner as the original network using the independent likelihood process Eq. (6) or categorical likelihood process Eq. (8). We denote  $E_{N,K} = (x_{i,K})_{i \in \mathbb{N}_{\neq}^d} \in \mathcal{N}_d$  to be the corresponding edge set of the truncated network up round  $N$ , where  $x_{i,K} = 0$  for any index  $i \in \mathbb{N}_{\neq}^d$  such that some component  $i_j > K$ . We keep  $E_N$  and  $E_{N,K}$  in the same space in order to compare their distributions in Sections 3.1 to 3.3.

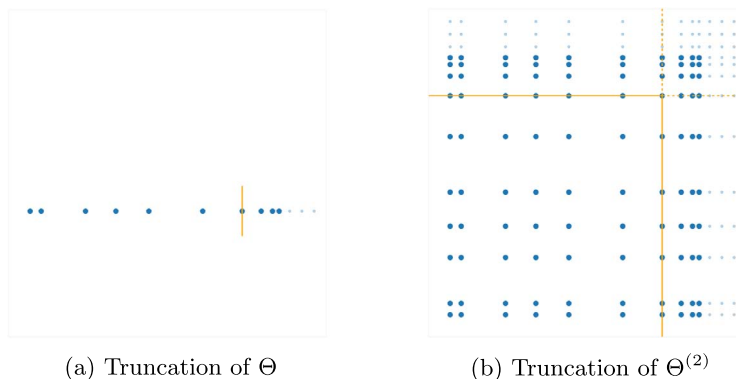


FIG 1. An illustration of the difference between truncation of CRMs ( $d = 1$ ) and self-product CRMs with  $d = 2$ . Intuitively, increasing  $d$  means that a higher proportion of mass is discarded in the truncation process.

### 3.1. Truncation error bound

We formulate the approximation error incurred by truncation as the total variation distance between the marginal network distributions, i.e., of  $E_N$  and  $E_{N,K}$ . The first step in the analysis of truncation error—provided by Lemma 3.1—is to show that this is bounded above by the probability that there are edges in the full network  $E_N$  involving vertices beyond the truncation level  $K$ . To this end, we denote the *maximum vertex index* of  $E_N$  to be

$$I_N := \max_{i \in \mathbb{N}_{\neq}^d} \left( \max_{j \in [d]} i_j \right) \quad \text{s.t.} \quad x_i > 0,$$

and note that by definition,  $I_N \leq K$  if and only if all edges in  $E_N$  fall in the truncated region.

**Lemma 3.1.** *Let  $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$  be a random discrete measure, and  $\Theta_K = \sum_{k=1}^K \theta_k \delta_{\psi_k}$  be its truncation to  $K$  atoms. Let  $\Theta^{(d)}$  be the self-product of  $\Theta$ , and  $\Theta_K^{(d)}$  be the self-product of  $\Theta_K$ . Let  $P_N$  and  $P_{N,K}$  be the marginal distributions of edge sets  $E_N, E_{N,K} \in \mathcal{N}_d$  under either the independent or categorical likelihood process. Then*

$$D_{\text{TV}}(P_N, P_{N,K}) \leq 1 - \mathbb{P}(I_N \leq K).$$

As mentioned in Section 2.3, the network  $E_N$  is in general a directed multi-graph with no self loops. However, Lemma 3.1—and the downstream truncation error bounds presented in Sections 3.2 and 3.3—also apply to any graph  $E'_N = (x'_i)_{i \in \mathbb{N}_{\neq}^d}$  that is a function of the original graph  $E_N = f(E_N)$  such that truncation commutes with the function, i.e.,  $E'_{N,K} = f(E_{N,K})$ . For example, to obtain a truncation error bound for the common setting of undirected binary

graphs, we generate the directed multigraph  $E_N$  as above and construct the undirected binary graph  $E'_N$  via

$$x'_i = \mathbf{1}_{x_i > 0} \cdot \mathbf{1}_{i_1 < i_2 < \dots < i_d}, \quad i \in \mathbb{N}_{\neq}^d. \tag{10}$$

Corollary 3.2 provides the precise statement of the result; note that the bound is identical to that from Lemma 3.1.

**Corollary 3.2.** *Let  $E'_N := (x'_i)_{i \in \mathbb{N}_{\neq}^d} \in \mathcal{N}_d$  be the set of edges of a network with truncation  $E'_{N,K} \in \mathcal{N}_d$ , and denote their marginal distributions  $P'_N$  and  $P'_{N,K}$ . If there exists a measurable function  $f$  such that*

$$E'_N = f(E_N) \quad \text{and} \quad E'_{N,K} = f(E_{N,K}),$$

then

$$D_{\text{TV}}(P'_N, P'_{N,K}) \leq 1 - \mathbb{P}(I_N \leq K).$$

### 3.2. Independent likelihood process

We now specialize Lemma 3.1 to the setting where  $\Theta$  is a CRM generated by a series representation of the form Eq. (3), and the network is generated via the independent likelihood process from Eq. (6). As a first step towards a bound on the truncation error for general hypergraphs with  $d > 1$  in Theorem 3.4, we present a simpler corollary in the case where  $d = 2$ , which is of direct interest in analyzing the truncation error of popular edge-exchangeable networks.

**Corollary 3.3.** *In the setting of Lemma 3.1, suppose  $\Theta$  is a CRM generated from the series representation Eq. (3), edges are generated from the independent likelihood process Eq. (6), and  $d = 2 \leq K$ . Then*

$$\mathbb{P}(I_N \leq K) \geq \exp(-N \cdot B_K),$$

where

$$\begin{aligned} B_K &= B_{K,1} + B_{K,2} + B_{K,3} \\ B_{K,1} &= \mathbb{E} \left[ \int_{\mathbb{R}_+^2} -\log \pi(\tau(U_1, \gamma_1 + \Gamma_K) \tau(U_2, \gamma_2 + \Gamma_K)) \, d\gamma_1 d\gamma_2 \right] \\ B_{K,2} &= 2\mathbb{E} \left[ \int_0^{\Gamma_K} \frac{(K-1)}{\Gamma_K} \int_{\Gamma_K}^{\infty} -\log \pi(\tau(U_1, \gamma_1) \tau(U_2, \gamma_2)) \, d\gamma_1 d\gamma_2 \right] \\ B_{K,3} &= 2\mathbb{E} \left[ \int_{\mathbb{R}_+} -\log \pi(\tau(U_1, \Gamma_K) \tau(U_2, \gamma + \Gamma_K)) \, d\gamma \right]. \end{aligned}$$

The proof of this result (and Theorem 3.4 below) in Appendix B follows by representing the tail of the CRM as a unit-rate Poisson process on  $[\Gamma_K, \infty)$ . Though perhaps complicated at first glance, an intuitive interpretation of the

truncation error terms  $B_{K,i}$  is provided by Fig. 1b.  $B_{K,1}$  corresponds to the truncation error arising from the upper right quadrant, where both vertices participating in the edge were in the discarded tail region.  $B_{K,2}$  is the truncation error arising from the bottom right and upper left quadrants, where one of the two vertices participating in the edge was in the truncation, and the other was in the tail. Finally,  $B_{K,3}$  represents the truncation error arising from edges in which one vertex was at the boundary of tail and truncation, and the other was in the tail.

Theorem 3.4 is the generalization of Corollary 3.3 from  $d = 2$  to the general setting of arbitrary hypergraphs with  $d > 1$ . The bound is analogous to that in Corollary 3.3—with  $B_K$  expressed as a sum of terms, each corresponding to whether vertices were in the tail, boundary, or truncation region—except that there are  $d > 1$  vertices participating in each edge, resulting in more terms in the sum. Note that Theorem 3.4 also guarantees that the bound is not vacuous, and indeed converges to 0 as the truncation level  $K \rightarrow \infty$  as expected.

**Theorem 3.4.** *In the setting of Lemma 3.1, suppose  $\Theta$  is a CRM generated from the series representation Eq. (3), edges are generated from the independent likelihood process Eq. (6), and  $1 < d \leq K$ . Then*

$$\mathbb{P}(I_N \leq K) \geq \exp(-N \cdot B_K),$$

where

$$\begin{aligned} B_K &= B_{K,1} + B_{K,2} + B_{K,3}, \tag{11} \\ B_{K,1} &= \mathbb{E} \left[ \int_{[\Gamma_K, \infty)^d} -\log \pi(\tilde{\theta}) d\gamma \right] \\ B_{K,2} &= \sum_{\ell=1}^{d-1} \binom{d}{\ell} \mathbb{E} \left[ \frac{(K-1)!}{(K-1-\ell)!} \Gamma_K^{-\ell} \int_{[0, \Gamma_K]^\ell \times [\Gamma_K, \infty)^{d-\ell}} -\log \pi(\tilde{\theta}) d\gamma \right] \\ B_{K,3} &= \sum_{\ell=1}^{d-1} \ell \binom{d}{\ell} \mathbb{E} \left[ \frac{(K-1)!}{(K-\ell)!} \Gamma_K^{-(\ell-1)} \int_{[0, \Gamma_K]^\ell \times [\Gamma_K, \infty)^{d-\ell}} -\delta_{\gamma_\ell = \Gamma_K} \log \pi(\tilde{\theta}) d\gamma \right], \end{aligned}$$

$\delta$ . is the Dirac delta,  $d\gamma := \prod_{j=1}^d d\gamma_j$ , and  $\tilde{\theta} := \prod_{j=1}^d \tau(U_j, \gamma_j)$ . Furthermore,  $\lim_{K \rightarrow \infty} B_K = 0$ .

The same geometric intuition from the  $d = 2$ -dimensional case applies to the general hypergraph truncation error in Eq. (11).  $B_{K,1}$  corresponds to the error arising from the edges whose vertices all belong to the tail region  $\Theta_{K+}$ . Each term in the summation in  $B_{K,2}$  corresponds to the error arising from edges that have  $\ell$  out of  $d$  vertices belonging to the truncation  $\Theta_K$ . Each term in the summation in  $B_{K,3}$  corresponds to the error arising from the edges that have  $\ell - 1$  out of  $d$  vertices belonging to the truncation  $\Theta_K$  and have one vertex exactly on the boundary of the truncation. Note that we obtain Corollary 3.3 by taking  $d = 2$  in Eq. (11).



### 3.3. Categorical likelihood process

We may also specialize Lemma 3.1 to the setting where the network is generated via the single-edge-per-step categorical likelihood process in Eq. (8). However, truncation with the categorical likelihood process poses a few key challenges. From a practical angle, certain choices of series representation for generating  $\Theta$  may be problematic. For instance, when using the rejection representation Eq. (5) in the typical case where  $\mu \neq \nu$ , there is a nonzero probability that

$$\sum_{k=1}^K \mathbb{1}[\theta_k > 0] < d,$$

meaning there aren't enough accepted vertices in the truncation to generate a single edge. In this case, the categorical likelihood process—which must generate exactly one edge per step—is ill-defined. An additional theoretical challenge arises from the normalization of the original and truncated networks in Eq. (8), which prevents the use of the usual theoretical tools for analyzing CRMs.

Fortunately, the inverse Lévy representation provides an avenue to address both issues. The rates  $\theta_k$  are all guaranteed to be nonzero—meaning as long as  $K \geq d$ , the categorical likelihood process is well-defined—and are decreasing, which enables our theoretical analysis in Appendix B.1. However, as mentioned earlier, the inverse Lévy representation is well-known to be intractable to use in most practical settings.

Theorem 3.5 provides a solution: we use the rejection representation to simulate the rates  $\theta_k$ , but instead of simulating for iterations  $k = 1, \dots, K$ , we simulate *until we obtain  $K$  nonzero rates*. This is no longer a sample of a truncated rejection representation; but Theorem 3.5 shows that the first  $K$  nonzero rates have the same distribution as simulating  $K$  iterations of the inverse Lévy representation. Therefore, we can tailor the analysis of truncation error for the categorical likelihood process in Theorem 3.6 to the inverse Lévy representation, and simulate its truncation for any  $K$  using the tractable rejection representation in practice.

**Theorem 3.5.** *Let  $\theta_1, \dots, \theta_K$  be the first  $K$  rates from the inverse Lévy representation of a CRM, and let  $\theta'_1, \dots, \theta'_K$  be the first  $K$  nonzero rates from any rejection representation of the same CRM. Then*

$$(\theta_1, \dots, \theta_K) \stackrel{d}{=} (\theta'_1, \dots, \theta'_K).$$

**Theorem 3.6.** *In the setting of Lemma 3.1, suppose  $\Theta$  is a CRM generated from the inverse Lévy representation Eq. (4), edges are generated from the categorical likelihood process Eq. (8), and  $1 < d \leq K$ . Then*

$$\mathbb{P}(I_N \leq K) \geq (1 - B_K)^{Nd} \geq 0,$$

where

$$B_K := \mathbb{E} \left[ \int_{-\infty}^{\infty} Q(\Gamma_K, x) \left( \int_0^1 Q(\Gamma_K u, x) du \right)^{K-d} \left( \frac{d}{dx} e^{\int_0^{\infty} Q(\Gamma_K + \gamma, x) - 1 d\gamma} \right) dx \right],$$

and

$$Q(u, t) = \exp(-\nu^{\leftarrow}(u)e^{-t}) \quad \text{and} \quad \Gamma_K \sim \text{Gam}(K, 1).$$

Furthermore,  $\lim_{K \rightarrow \infty} B_K = 0$ .

### 3.4. Examples

We now apply the results of this section to binary undirected networks ( $d = 2$ ) constructed via Eq. (10) from common edge-exchangeable networks [3]. We derive the convergence rate of truncation error, and provide simulations of the expected number of edges and vertices under the infinite and truncated network. In each simulation we use the rejection representation to simulate  $\Theta$ , and run  $N = 10,000$  steps of network construction.

**Beta-Bernoulli process network** Suppose  $\Theta$  is generated from a beta process, and network  $E_N$  is generated using the independent Bernoulli likelihood process [3]. The beta process  $\text{BP}(\gamma, \lambda, \alpha)$  [42] with discount parameter  $\alpha \in [0, 1)$ , concentration parameter  $\lambda > -\alpha$ , and mass parameter  $\gamma > 0$  is a CRM with rate measure

$$\nu(d\theta) = \gamma \frac{\Gamma(\lambda + 1)}{\Gamma(1 - \alpha)\Gamma(\lambda + \alpha)} \mathbb{1}[\theta \leq 1] \theta^{-1-\alpha} (1 - \theta)^{\lambda + \alpha - 1} d\theta. \quad (12)$$

The Bernoulli likelihood has the form

$$h(x|\theta) = \mathbb{1}[x \leq 1] \theta^x (1 - \theta)^{1-x}. \quad (13)$$

To simulate the process, we use a proposal rate measure  $\mu$  given by

$$\mu(d\theta) = \gamma' \mathbb{1}[\theta \leq 1] \theta^{-1-\alpha} d\theta, \quad \gamma' = \gamma \frac{\Gamma(\lambda + 1)}{\Gamma(1 - \alpha)\Gamma(\lambda + \alpha)}.$$

*Dense network:* When  $\alpha = 0$ , the binary beta-Bernoulli graph is dense and

$$\mu^{\leftarrow}(u) = e^{-u/\gamma'}, \quad \frac{d\nu}{d\mu} = (1 - \theta)^{\lambda-1}.$$

Therefore the rejection representation Eq. (5) of  $\text{BP}(\gamma, \lambda, 0)$  can be written as

$$\theta_k = T_k \mathbb{1}(U_k \leq (1 - T_k)^{\lambda-1}), \quad T_k = e^{-\Gamma_k/\gamma'}.$$

In Appendix C.2, we show that there exists  $K_0 \in \mathbb{N}$  such that

$$\forall K \geq K_0, \quad B_K \leq 12\gamma(1 - e^{-1})^{\lambda-2} \left( \frac{\gamma'}{1 + \gamma'} \right)^K.$$

This implies that the truncation error of the dense binary beta-Bernoulli network converges to 0 geometrically in  $K$ . Fig. 2a corroborates this result in simulation

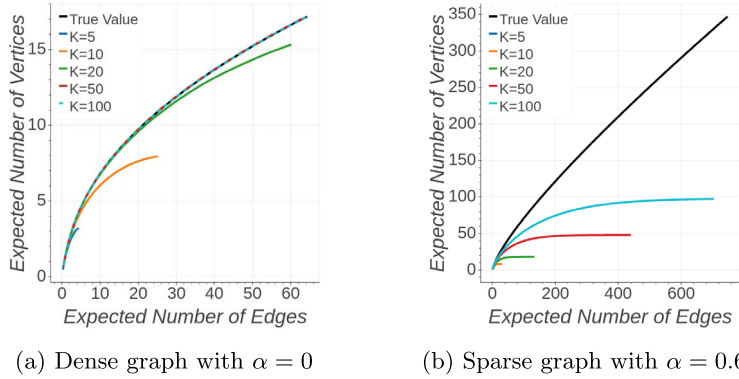


FIG 2. Beta-independent Bernoulli network

with  $\lambda = 2$  and  $\gamma = 1$ ; it can be seen that for the dense beta-Bernoulli graph, truncated graphs with relatively low truncation level—in this case,  $K \approx 50$ —approximate the true network model well.

*Sparse network:* When  $\alpha \in (0, 1)$ , the binary beta-Bernoulli graph is sparse and

$$\mu^{\leftarrow}(u) = \left(1 + \frac{\alpha u}{\gamma'}\right)^{-1/\alpha}, \quad \frac{d\nu}{d\mu} = (1 - \theta)^{\lambda + \alpha - 1}.$$

Therefore the rejection representation Eq. (5) of  $\text{BP}(\gamma, \lambda, \alpha)$  can be written as

$$\theta_k = T_k \mathbb{1}(U_k \leq (1 - T_k)^{\lambda + \alpha - 1}), \quad T_k = (1 + \alpha \Gamma_k / \gamma')^{-1/\alpha}.$$

In Appendix C.2, we show that there exists  $K_0 \in \mathbb{N}$  such that

$$\forall K \geq K_0, \quad B_K \leq 6\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} (K - 1)^{\frac{\alpha-1}{\alpha}}.$$

This bound suggests that the truncation error for the sparse binary beta-Bernoulli network converges to 0 much more slowly than for the dense graph. Fig. 2b corroborates this result in simulation with  $\lambda = 2$ ,  $\gamma = 1$ , and  $\alpha = 0.6$ ; it can be seen that for the sparse beta-Bernoulli graph, truncated graphs behave significantly differently from the true graph for moderate truncation levels. In practice, one should select an appropriate (large) value of  $K$  using our error bounds as guidance, and use sparse data structures to avoid undue memory burden.

**Gamma-independent Poisson network** Next, consider the network with  $\Theta$  generated from a gamma process, and the network  $E_N$  generated using the independent Poisson likelihood process. The gamma process  $\Gamma\text{P}(\gamma, \lambda, \alpha)$  [43] with discount parameter  $\alpha \in [0, 1)$ , scale parameter  $\lambda > 0$  and mass parameter  $\gamma > 0$  has rate measure

$$\nu(d\theta) = \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \theta^{-\alpha-1} e^{-\lambda\theta} d\theta.$$

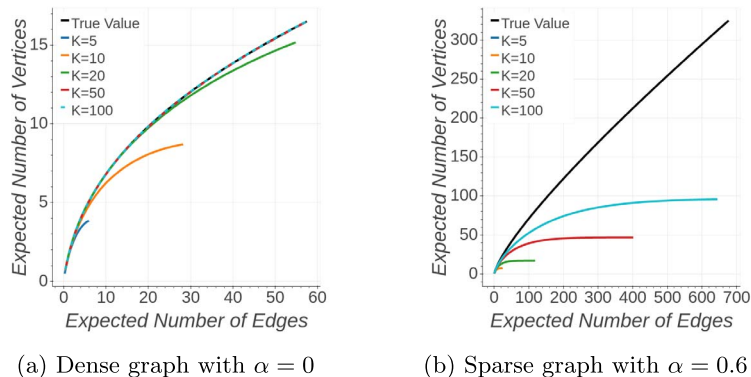


FIG 3. Gamma-independent Poisson graph

The Poisson likelihood has the form

$$h(x|\theta) = \frac{\theta^x}{x!} e^{-\theta}.$$

*Dense network:* When  $\alpha = 0$ , the gamma-Poisson graph is dense, and we choose the proposal measure  $\mu$  to be

$$\mu(d\theta) = \gamma \lambda \theta^{-1} (1 + \lambda \theta)^{-1} d\theta,$$

such that

$$\mu^{\leftarrow}(u) = 1 / \left( \lambda \left( e^{(\gamma \lambda)^{-1} u} - 1 \right) \right), \quad \frac{d\nu}{d\mu} = (1 + \lambda \theta) e^{-\lambda \theta}.$$

Therefore, the rejection representation in Eq. (5) has the form

$$\theta_k = T_k \mathbb{1}(U_k \leq (1 + \lambda T_k) e^{-\lambda T_k}), \quad T_k = \frac{1}{\lambda (e^{(\gamma \lambda)^{-1} T_k} - 1)}.$$

In Appendix C.3, we show that there exists  $K_0 \in \mathbb{N}$  such that

$$\forall K \geq K_0, \quad B_K \leq 6 \frac{\gamma}{\lambda} \left( \frac{\gamma \lambda}{1 + \gamma \lambda} \right)^{K-1}.$$

Again, for the dense network  $B_K$  converges to 0 geometrically, indicating that truncation may provide reasonable approximations to the original network. Fig. 3a corroborates this result in simulation with  $\lambda = 2$  and  $\gamma = 1$ ; for  $K \approx 50$ , no vertices are discarded on average by truncation.

*Sparse network:* When  $\alpha \in (0, 1)$ , the gamma-Poisson graph is sparse, and we choose the proposal measure  $\mu$  to be

$$\mu(d\theta) = \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \theta^{-1-\alpha} d\theta,$$

such that

$$\mu^{\leftarrow}(u) = (\gamma' u^{-1})^{1/d}, \quad \gamma' := \gamma \frac{\lambda^{1-\alpha}}{\alpha \Gamma(1-\alpha)}, \quad \frac{d\nu}{d\mu} = e^{-\lambda\theta}.$$

Therefore the rejection representation in Eq. (5) has the form

$$\theta_k = T_k \mathbb{1}(U_k \leq e^{-\lambda T_k}), \quad T_k = (\gamma' \Gamma_k^{-1})^{1/\alpha}.$$

In Appendix C.3, we show there exists  $K_0 \in \mathbb{N}$  such that

$$\forall K \geq K_0, \quad B_K \leq \frac{12\gamma^2 \lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} \left( \frac{3\gamma'}{K-1} \right)^{\frac{1-\alpha}{\alpha}}.$$

Again, for the sparse network  $B_K$  converges to 0 slowly, suggesting that the truncation error for the sparse binary gamma-Poisson graph converges more slowly than for the dense graph. Fig. 3b corroborates this result in simulation with  $\lambda = 2$ ,  $\gamma = 1$ , and  $\alpha = 0.6$ ; for a moderate range of truncation values  $K \leq 100$ , the truncated graph behaves very differently from the true graph. Therefore in practice, one should follow the above guidance for the beta-Bernoulli network: select a value of  $K$  using our error bounds, and avoid intractable memory requirements by using sparse data structures.

#### 4. Truncated posterior inference

In this section, we develop a tractable approximate posterior inference method for network models via truncation, and analyze its approximation error. In particular, given an observed network  $E_N$ , we want to simulate from the posterior distribution of the CRM rates and the parameters of the CRM rate measure. Since exact expressions of full conditional densities are not available, we truncate the model and run an approximate Markov chain Monte Carlo algorithm. We provide a rigorous theoretical justification for this simple approach by establishing a bound on the total variation distance between the truncated and exact posterior distribution. This in turn provides a method to select the truncation level in a principled manner.

Although this section focuses on self-product-CRM-based network models, the method and theory we develop are both general and extend easily to other CRM-based models, e.g. for clustering [44], feature allocation [45], and trait allocation [6]. In particular, the methodology only requires bounds on tail occupancy probabilities (e.g., in the present context, the probability that  $I_N \leq K$ ) rather than the exact evaluation of these quantities.

##### 4.1. Truncation error bound

We begin by examining the density of the posterior distribution of the  $K^{\text{th}}$  rate from the inverse Lévy representation  $\theta_K$  for some fixed  $K \in \mathbb{N}$ , the unordered

collection of rates  $(\theta_k)_{k=1}^{K-1}$  such that  $\theta_k \geq \theta_K$ , and the parameters  $\sigma \in \mathbb{R}^m$  of the CRM rate measure  $\nu_\sigma$  given the observed set of edges  $E_N$ . We denote  $\nu_\sigma(\theta)$  to be the density of  $\nu_\sigma(d\theta)$  and  $p(\sigma)$  to be the prior density of  $\sigma$ , both with respect to the Lebesgue measure. Given these definitions the posterior density can be expressed as

$$p(\sigma, \theta_{1:K}, X_{1:N}) \propto p(\sigma) \cdot e^{-\nu_\sigma[\theta_K, \infty)} \cdot \prod_{k=1}^K \mathbb{1}[\theta_K \leq \theta_k] \nu_\sigma(\theta_k) \cdot \left( \prod_{n=1}^N \prod_{i \in [K]_{\neq}^d} h(x_{ni} | \prod_{j=1}^d \theta_{i_j}) \right) \cdot p(x_{1:N}^{K+} | \theta_{1:K}, \sigma), \quad (14)$$

where  $[K]_{\neq}^d$  is the subset of  $\mathbb{N}_{\neq}^d$  such that  $\max_{j \in [d]} i_j \leq K$ , and  $x_{1:N}^{K+}$  is shorthand for the set of  $(x_{ni})_{n \in [N], i \in \mathbb{N}_{\neq}^d}$  such that  $i \notin [K]_{\neq}^d$ . All the factors on the first row correspond to the prior of  $\sigma$  and  $(\theta_k)_{k=1}^K$ , and the first factor on the second row corresponds to the likelihood of the portion of the network involving only vertices  $1 \dots K$ ; these are straightforward to evaluate, though  $\nu_\sigma[\theta_K, \infty)$  will occasionally require a 1-dimensional numerical integration. The last factor corresponds to the likelihood of the portion of the network involving vertices beyond  $K$ , and is not generally possible to evaluate exactly.

To handle this term, suppose that  $K$  is large enough that  $x_{1:N}^{K+} = 0$ . Then  $p(x_{1:N}^{K+} | \theta_{1:K}, \sigma) = \mathbb{P}(I_N \leq K | \Gamma_{1:K}, \sigma)$ , i.e., the probability that no portion of the network involves vertices of index  $> K$ . Theorem 4.1 provides upper and lower bounds on this probability akin to those of Theorem 3.4—indeed, Theorem 4.1 is an intermediate step in the proof of Theorem 3.4—that apply conditionally on the values of  $U_{1:K}, \Gamma_{1:K}$  rather than marginally. This theorem also makes the dependence of the bound on the rate measure parameters  $\sigma$  notationally explicit via parametrized series representation components  $\tau_\sigma$  and  $g_\sigma$  from Eq. (3). Finally, though Theorem 4.1 applies to general series representations, we require only the specific instantiation for the inverse Lévy representation in the present context.

**Theorem 4.1.** *The conditional probability  $\mathbb{P}(I_N \leq K | U_{1:K}, \Gamma_{1:K}, \sigma)$  satisfies*

$$1 \geq \mathbb{P}(I_N \leq K | U_{1:K}, \Gamma_{1:K}, \sigma) \geq \exp(-N \cdot B(U_{1:K}, \Gamma_{1:K}, \sigma)),$$

where

$$B(U_{1:K}, \Gamma_{1:K}, \sigma) = \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{\substack{\mathcal{L} \subseteq [K] \\ |\mathcal{L}|=\ell}} B(U_{1:K}, \Gamma_{1:K}, \sigma, \mathcal{L})$$

$$B(U_{1:K}, \Gamma_{1:K}, \sigma, \mathcal{L}) = \int_{[\Gamma_K, \infty)^{d-\ell}} \mathbb{E} \left[ -\log \pi \left( \prod_{j \in \mathcal{L}} \theta_j \prod_{j=1}^{d-\ell} \tau_\sigma(U'_j, \gamma_j) \right) \mid \theta_{1:K} \right] d\gamma,$$

where  $d\gamma = \prod_{j=1}^{d-\ell} d\gamma_j$  and  $U'_j \stackrel{i.i.d.}{\sim} g_\sigma$ .

Since  $U_{1:K}$  is unused in the inverse Lévy representation, in the present context we use the notation  $B(\Gamma_{1:K}, \sigma)$  for brevity. The bound in Theorem 4.1 implies that as long as  $K$  is set large enough such that both  $x_{1:N}^{K+} = 0$  and  $B(\Gamma_{1:K}, \sigma) \leq \epsilon/N$  for some  $\epsilon > 0$  then

$$1 - \epsilon \leq p(x_{1:N}^{K+} | \theta_{1:K}, \sigma) \leq 1.$$

Therefore as  $K$  increases, this term should become  $\approx 1$  and have a negligible effect on the posterior density. We use this intuition to propose a truncated Metropolis–Hastings algorithm that sets  $K > I_N$ , ignores the  $p(x_{1:N}^{K+} | \theta_{1:K}, \sigma)$  term in the acceptance ratio, and fixes  $x_N^{K+}$  to 0. Theorem 4.2 provides a rigorous analysis of the error involved in using the truncated sampler.

**Theorem 4.2.** *Fix  $K > I_N$ . Let  $\Pi$  be the distribution of  $\sigma, \theta_{1:K}$  given  $E_N$ , and let  $\hat{\Pi}$  be the distribution with density proportional to Eq. (14) without the  $p(x_{1:N}^{K+} | \theta_{1:K}, \sigma)$  term and with  $x_N^{K+}$  fixed to 0. If for some  $\eta \in [0, 1)$ ,*

$$\hat{\Pi} \{B(\Gamma_{1:K}, \sigma) \leq \epsilon/N\} \geq 1 - \eta,$$

then

$$D_{\text{TV}}(\hat{\Pi}, \Pi) \leq \frac{3(\epsilon + \eta)}{2} - \epsilon\eta. \quad (15)$$

#### 4.2. Adaptive truncated Metropolis–Hastings

Crucially, as long as one can obtain samples from the *truncated* posterior distribution, one can estimate the bound in Eq. (15) using sample estimates of the tail probability  $\hat{\Pi} \{B(\Gamma_{1:K}, \sigma) \leq \epsilon/N\} \geq 1 - \eta$ . Therefore, one can compute the bound Eq. (15) without needing to evaluate  $p(x_{1:N}^{K+} | \theta_{1:K}, \sigma)$  exactly. This suggests the following adaptive truncation procedure:

1. Pick a value of  $K > I_N$  and desired total variation error  $\xi \in (0, 1)$ .
2. Obtain samples from the truncated posterior.
3. Minimize the bound in Eq. (15) over  $\epsilon \in (0, 1)$ , using the samples to estimate  $\eta = 1 - \hat{\Pi} \{B(\Gamma_{1:K}, \sigma) \leq \epsilon/N\}$  for each value of  $\epsilon$ .
4. If the minimum bound exceeds  $\xi$ , increase  $K$  and return to 2.
5. Otherwise, return the samples.

In this work, we start by initializing  $K$  to  $I_N + 1$ . In order to decide how much to increase  $K$  by in each iteration, we use the sequential representation to extrapolate the total variation bound Eq. (15) to larger values of  $K$  without actually performing MCMC sampling. In particular, for each posterior sample, we use its hyperparameters to generate additional rates from the sequential representation. We then use these extended samples to compute the total variation error guarantee as per step 3 above. We continue to generate additional rates (typically doubling the number each time) until the predicted total variation guarantee is below the desired threshold. Finally, we use linear interpolation between the last two predicted errors to find the next truncation level  $K$  that matches the desired (log) error threshold. This fixes a new value of  $K$ ; at this point, we return to step 2 above and iterate.

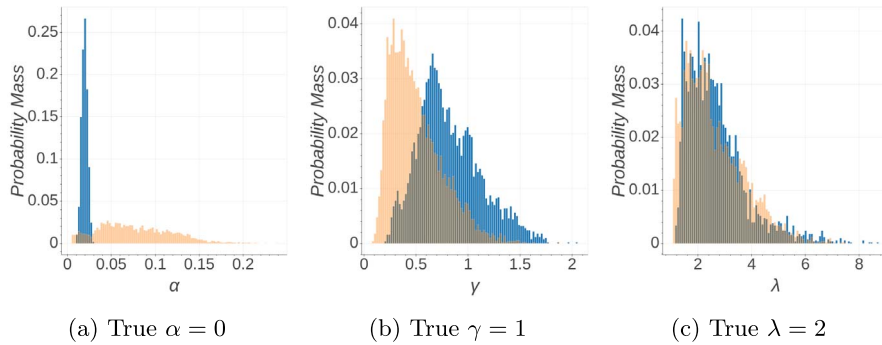


FIG 4. Posterior of  $\alpha$ ,  $\gamma$ ,  $\lambda$  for the dense simulated network. Orange histograms depict the first adaptation iteration, and blue histograms depict the second and final iteration.

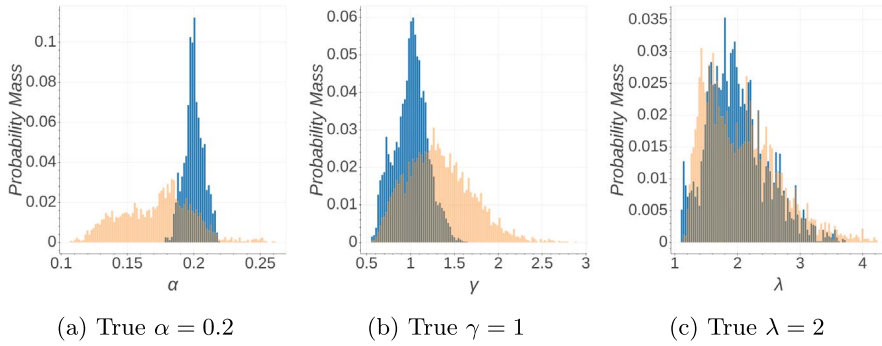


FIG 5. Posterior of  $\alpha$ ,  $\gamma$ ,  $\lambda$  for the sparse simulated network. Orange histograms depict the first adaptation iteration, and blue histograms depict the second and final iteration.

### 4.3. Experiments

In this section, we examine the properties of the proposed adaptive truncated inference algorithm for the beta-independent Bernoulli network model in Eqs. (12) and (13) with discount  $\alpha \in (0, 1)$ , concentration  $\lambda > 1$ , mass  $\gamma > 0$ , unordered collection of rates  $\theta_{1:K-1}$ , and  $K^{\text{th}}$  rate from the sequential representation  $\theta_K$ . In order to simplify inference, we transform each of these parameters to an unconstrained version:

$$\alpha = \frac{\exp(\alpha_u)}{1 + \exp(\alpha_u)}, \quad \lambda = 1 + \exp(\lambda_u), \quad \gamma = \exp(\gamma_u),$$

$$\theta_K = \frac{\exp(\theta_{u,K})}{1 + \exp(\theta_{u,K})}, \quad \theta_k = \frac{\theta_K + \exp(\theta_{u,k})}{1 + \exp(\theta_{u,k})} \quad k = 1, \dots, K - 1.$$

We use a Markov chain Monte Carlo algorithm that includes an exact Gibbs sampling move for  $\gamma$ , and separate Gaussian random-walk Metropolis–Hastings



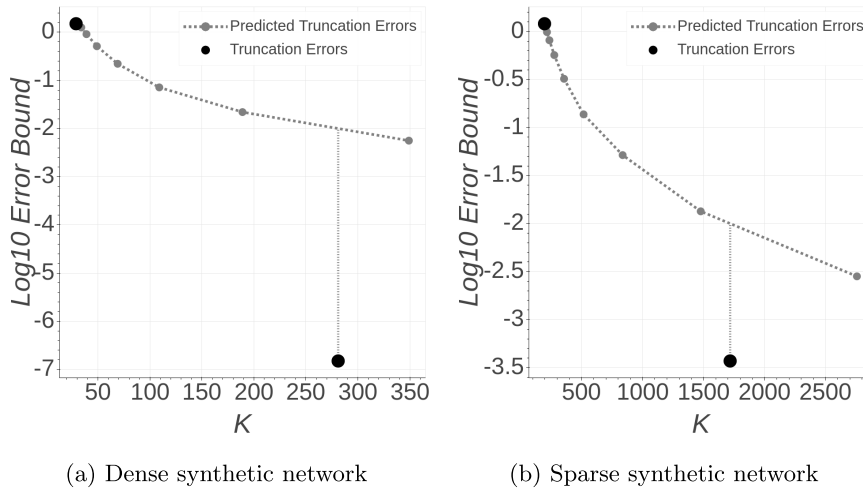


FIG 6. Visualization of the truncation expansion procedure. Black circles denote the truncation level and error in each iteration of the adaptation. The adaptation stops when the truncation error falls below the desired threshold (here  $\log_{10}(0.01) = -2$ ). Grey dashed lines and circles visualize the predicted truncation error using rates generated from the sequential representation. The vertical dotted line shows that the algorithm selects a value of  $K$  that attempts to match the desired  $\log_{10}$  error threshold of  $-2$  using the predictions.

moves for  $\alpha_u$ ,  $\lambda_u$ ,  $\theta_{u,K}$ , all  $\theta_{u,k}$  such that vertex  $k$  has degree 0 (jointly), and each  $\theta_{u,k}$  such that vertex  $k$  has nonzero degree (individually).

**Synthetic data** We first apply the model to synthetic data simulated from the generative model. We simulate a sparse network with parameters  $\lambda = 2$ ,  $\gamma = 1$ ,  $\alpha = 0.2$ , and  $N = 10^5$ , and a dense network with  $\lambda = 2$ ,  $\gamma = 1$ ,  $\alpha = 0$ , and  $N = 10^7$ . In both settings we set the truncation level for data generation to 500, the desired total variation bound from Eq. (15) to 0.01, and initialize the sampler with  $\alpha = 0.4$ ,  $\lambda = 5$ ,  $\gamma = 2$  and  $\theta$  generated from the sequential representation. All Metropolis–Hastings moves have proposal standard deviation 0.1 except the sparse network  $\alpha_u$  move, which has standard deviation 0.03.

Figs. 4 and 5 show histograms of 5,000 marginal posterior samples of the hyperparameters for the dense (true  $\alpha = 0$ ) and sparse (true  $\alpha = 0.2$ ) networks. In both cases, the approximate posterior in the first round of adaptation (orange histogram) does not concentrate on the true hyperparameter values, despite the relatively large number of generative rounds  $N$ . Fig. 6—which displays the truncation error and predictive adaptation procedure—shows why this is the case. In both networks, the first adaptation iteration identifies a large truncation error. After a single round of adaptation, the approximate posterior distributions (blue histograms) in Figs. 4 and 5 concentrate more on the true values as expected, and the truncation errors fall well below the desired threshold ( $\log_{10}(0.01) = -2$ ). It is worth noting that the predictive extrapolation appears to be quite conservative in these examples, and especially in the dense network.

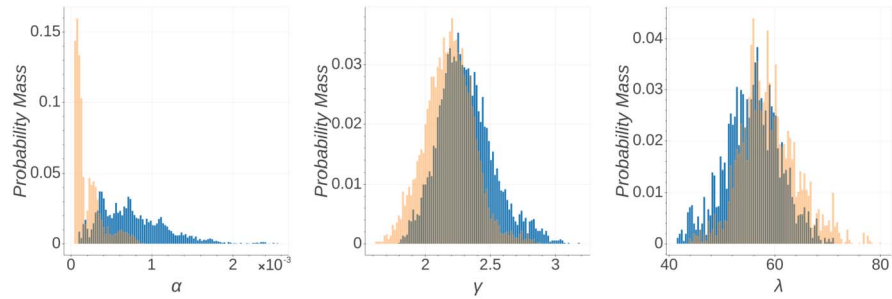
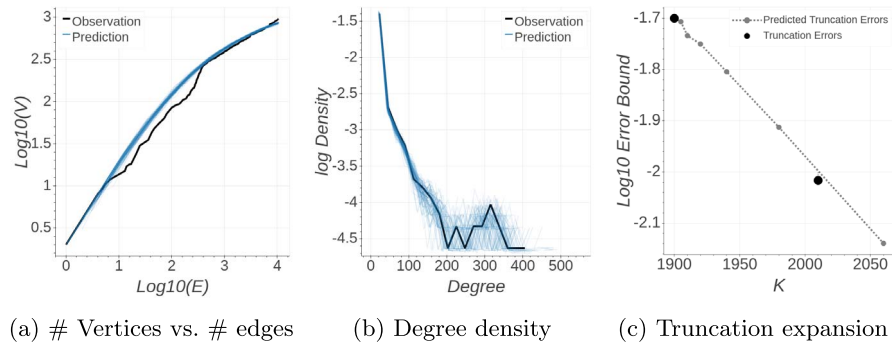
FIG 7. Posterior of  $\alpha$ ,  $\gamma$ ,  $\lambda$  in the Facebook-like network

FIG 8. (8a,8b): The characteristics of the observed Facebook-like network (black) and 100 samples from the posterior predictive distribution (blue). (8c): The truncation adaptation process, with truncation level and error bound in each iteration (black circles) and predictive truncation errors (grey dashed lines).

This happens because the approximate posterior for the dense network (respectively, sparse network) assigns mass to higher values of  $\alpha$  (respectively,  $\gamma$ ) than it should, which results in larger truncation error and thus a larger predicted required truncation level.

**Real network data** Next, we apply the model to a Facebook-like Social Network<sup>2</sup> [46]. The original source network contains a sequence of 61,734 weighted, time-stamped edges, and 1,899 vertices. We preprocess the data by removing the edge weights, binning the edge sequence into 30-minute intervals, and removing the initial transient of network growth, resulting in 1,899 vertices and 10,435 edges over  $N = 6,427$  rounds of generation. We again set a desired total variation error guarantee of 0.01 during inference. All Metropolis–Hastings moves have proposal standard deviation 0.1 except the  $\alpha_u$  and degree-0  $\theta_u$  moves, which have standard deviation 0.04 and 0.03, respectively. We initialize the sampler to  $\alpha = 0.1$ ,  $\gamma = 2$ ,  $\lambda = 20$  and sample rates  $\theta$  from the prior sequential

<sup>2</sup>Available at <https://toreopsahl.com/datasets/>

representation.

Fig. 7 shows the posterior marginal histograms for the hyperparameters  $\alpha, \lambda, \gamma$  in both the first iteration (orange) and the second and final iteration (blue) of truncation adaptation. The posterior distribution suggests that the network is dense (i.e.  $\alpha \approx 0$ ). This conclusion is supported both by the close match of 100 samples from the posterior predictive distribution, shown in Figs. 8a and 8b, and the findings of past work using this data [8]. Further, as in the synthetic examples, the truncation adaptation terminates after two iterations; but in this case the histograms do not change very much between the two. This is essentially because the truncation error in the first iteration is relatively low ( $\approx 0.02$ ), leading to a reasonably accurate truncated posterior and hence accurate predictions of the truncation error at higher truncation levels.

## 5. Conclusion

In this paper, we developed methods for tractable generative simulation and posterior inference in statistical models with discrete nonparametric priors via finite truncation. We demonstrated that these approximate truncation-based approaches are sound via theoretical error analysis. In the process, we also showed that the nonzero rates of the (tractable) rejection representation of a Poisson process are equal in distribution to the rates of the (intractable) inverse Lévy representation. Simulated and real network examples demonstrated that the proposed methods are useful in selecting truncation levels for both forward generation and inference in practice.

## Appendix A: Equivalence between nonzero rates from a rejection representation and the inverse Lévy representation

*Proof of Theorem 3.5.* Denote  $T_{k_1}$  be the first nonzero element that is generated from the rejection representation from Eq. (5) and correspondingly, denote  $\Gamma_{k_1}$  be the jump of the unit-rate homogeneous Poisson process on  $\mathbb{R}_+$  such that  $T_{k_1} = \mu^{\leftarrow}(\Gamma_{k_1})$ , where  $\mu$  is the proposal measure in the rejection representation. Let  $f$  be a bounded continuous function. Then

$$\begin{aligned} \mathbb{E}[f(T_{k_1})] &= \mathbb{E}[f(\mu^{\leftarrow}(\Gamma_{k_1}))] \\ &= \mathbb{E} \left[ \sum_{j=1}^{\infty} f(\mu^{\leftarrow}(\Gamma_j)) \mathbb{1} \left[ \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_j)) \geq U_j \right] \prod_{i=1}^{j-1} \mathbb{1} \left[ \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_i)) < U_i \right] \right] \\ &= \mathbb{E} \left[ \sum_{j=1}^{\infty} f(\mu^{\leftarrow}(\Gamma_j)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_j)) \prod_{i=1}^{j-1} \left( 1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_i)) \right) \right] \\ &= \sum_{j=1}^{\infty} \mathbb{E} \left[ f(\mu^{\leftarrow}(\Gamma_j)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_j)) \mathbb{E} \left[ \prod_{i=1}^{j-1} \left( 1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_i)) \right) \middle| \Gamma_j \right] \right] \end{aligned}$$

Note that given  $\Gamma_j$ ,  $\Gamma_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \Gamma_j)$ , for  $i = 1, \dots, j-1$ , so

$$\mathbb{E} \left[ \prod_{i=1}^{j-1} \left( 1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_i)) \right) \middle| \Gamma_j \right] = \mathbb{E} \left[ 1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(U)) \middle| \Gamma_j \right]^{j-1},$$

where  $U \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \Gamma_j)$ . Using the change of variable  $y = \mu^{\leftarrow}(u)$ , we obtain

$$\mathbb{E} \left[ 1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(U)) \middle| \Gamma_j \right] = 1 - \frac{1}{\Gamma_j} \int_0^{\Gamma_j} \frac{d\nu}{d\mu}(\mu^{\leftarrow}(u)) du = 1 - \frac{1}{\Gamma_j} \int_{\mu^{\leftarrow}(\Gamma_j)}^{\infty} d\nu.$$

Therefore, using the same change of variable trick,

$$\begin{aligned} \mathbb{E}[f(T_{k_1})] &= \sum_{j=1}^{\infty} \mathbb{E} \left[ f(\mu^{\leftarrow}(\Gamma_j)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_j)) \left( 1 - \frac{1}{\Gamma_j} \int_{\mu^{\leftarrow}(\Gamma_j)}^{\infty} d\nu \right)^{j-1} \right] \\ &= \sum_{j=1}^{\infty} \int_0^{\infty} f(\mu^{\leftarrow}(\gamma)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\gamma)) \left( 1 - \frac{1}{\gamma} \int_{\mu^{\leftarrow}(\gamma)}^{\infty} d\nu \right)^{j-1} \frac{\gamma^{j-1}}{(j-1)!} e^{-\gamma} d\gamma \\ &= \sum_{j=1}^{\infty} \int_0^{\infty} f(y) (1 - \mu[y, \infty)^{-1} \nu[y, \infty])^{j-1} \frac{\mu[y, \infty]^{j-1}}{(j-1)!} e^{-\mu[y, \infty]} \nu(dy) \\ &= \int_0^{\infty} f(y) \sum_{j=1}^{\infty} \frac{(\mu[y, \infty] - \nu[y, \infty])^{j-1}}{(j-1)!} e^{-\mu[y, \infty]} \nu(dy) \\ &= \int_0^{\infty} f(y) e^{-\nu[y, \infty]} \nu(dy). \end{aligned}$$

Suppose that  $\theta_1$  is the first rate generated using the inverse Lévy representation. Then

$$\mathbb{E}[f(\theta_1)] = \int_0^{\infty} f(\nu^{\leftarrow}(\gamma)) e^{-\gamma} d\gamma.$$

Making the change of variable  $y = \nu^{\leftarrow}(\gamma)$ , we obtain

$$\mathbb{E}[f(\theta_1)] = \int_0^{\infty} f(y) e^{-\nu[y, \infty]} \nu(dy) = \mathbb{E}[f(T_{k_1})].$$

Therefore, the first nonzero rate  $\theta_{k_1}$  from the rejection representation has the same marginal distribution as the first rate  $\theta_1$  from the inverse Lévy representation.

We now employ an inductive argument. Suppose that we have shown that the marginal distribution of first nonzero  $M$  elements  $\Xi_M := (T_{k_1}, T_{k_2}, \dots, T_{k_M})$  from the rejection representation has the same marginal distribution as the first  $M$  elements  $\Theta_M := (\theta_1, \dots, \theta_M)$  from the inverse Lévy representation. To prove the same for  $M+1$  elements, it suffices to show that the conditional distribution of  $T_{k_{M+1}}$  given  $\Xi_M$  is equal to the conditional distribution of  $\theta_{M+1}$  given  $\Theta_M$  when  $\Xi_M = \Theta_M$ .

Denote  $\Gamma'_j = \sum_{i=1}^j e'_i$ , where  $e'_i \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(1)$ , and  $U'_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$ . Then

$$\begin{aligned} \mathbb{E}[f(T_{k_{M+1}})|\Xi_M] &= \mathbb{E}[f(\mu^{\leftarrow}(\Gamma_{k_{M+1}}))|\Xi_M] \\ &= \mathbb{E}\left[\sum_{j=1}^{\infty} f(\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma'_j))\mathbb{1}[\dots] \middle| \Xi_M\right], \end{aligned}$$

where  $\mathbb{1}[\dots]$  is shorthand for

$$\mathbb{1}[\dots] = \mathbb{1}\left[\frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma'_j)) \geq U'_j\right] \prod_{i=1}^{j-1} \mathbb{1}\left[\frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma_i)) < U'_i\right].$$

Using steps similar to the base case,

$$\mathbb{E}[f(T_{k_{M+1}})|\Xi_M] = \sum_{j=1}^{\infty} \mathbb{E}\left[f(\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma'_j)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma'_j)) \mathbb{E}[\dots]^{j-1} \middle| \Xi_M\right],$$

where

$$\mathbb{E}[\dots] = \mathbb{E}\left[1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + U)) \middle| \Gamma'_j, \Xi_M\right] \quad U \sim \text{Unif}[0, \Gamma'_j].$$

Making the change of variable  $y = \mu^{\leftarrow}(\Gamma_{k_M} + u)$  as before, we obtain

$$\begin{aligned} \mathbb{E}\left[1 - \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + U)) \middle| \Gamma'_j, \Xi_M\right] &= 1 - \int_0^{\Gamma'_j} \frac{1}{\Gamma'_j} \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + u)) du \\ &= 1 - \frac{1}{\Gamma'_j} \int_{\mu^{\leftarrow}(\Gamma_{k_M})}^{\mu^{\leftarrow}(\Gamma_{k_M} + \Gamma'_j)} d\nu. \end{aligned}$$

Making another change of variables  $y = \mu^{\leftarrow}(\Gamma_{k_M} + \gamma)$  in the original integral—and hence  $\gamma = \mu[y, \infty) - \Gamma_{k_M} = \mu[y, \mu^{\leftarrow}(\Gamma_{k_M})]$ —yields

$$\begin{aligned} &\mathbb{E}[f(T_{k_{M+1}})|\Xi_M] \\ &= \sum_{j=1}^{\infty} \int_0^{\infty} f(\mu^{\leftarrow}(\Gamma_{k_M} + \gamma)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_{k_M} + \gamma)) \mathbb{E}[\dots]^{j-1} \frac{\gamma^{j-1}}{(j-1)!} e^{-\gamma} d\gamma \\ &= \sum_{j=1}^{\infty} \int_0^{\mu^{\leftarrow}(\Gamma_{k_M})} f(y) \left(1 - \frac{1}{\gamma} \int_y^{\mu^{\leftarrow}(\Gamma_{k_M})} d\nu\right)^{j-1} \frac{\gamma^{j-1}}{(j-1)!} e^{-\gamma} \nu(dy) \\ &= \sum_{j=1}^{\infty} \int_0^{\mu^{\leftarrow}(\Gamma_{k_M})} f(y) \frac{(\mu[y, \mu^{\leftarrow}(\Gamma_{k_M})] - \nu[y, \mu^{\leftarrow}(\Gamma_{k_M})])^{j-1}}{(j-1)!} e^{-\mu[y, \mu^{\leftarrow}(\Gamma_{k_M})]} \nu(dy) \\ &= \int_0^{\mu^{\leftarrow}(\Gamma_{k_M})} f(y) e^{-\nu[y, \mu^{\leftarrow}(\Gamma_{k_M})]} \nu(dy) = \int_0^{T_{k_M}} f(y) e^{-\nu[y, T_{k_M}]} \nu(dy). \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbb{E}[f(\theta_{M+1})|\Theta_M] &= \mathbb{E}[f(\nu^{\leftarrow}(\Gamma_M + \Gamma'_1))|\Theta_M] = \int_0^\infty f(\nu^{\leftarrow}(\Gamma_M + \gamma))e^{-\gamma}d\gamma \\ &= \int_0^{\nu^{\leftarrow}(\Gamma_M)} f(y)e^{-\nu[y, \nu^{\leftarrow}(\Gamma_M)]}\nu(dy) = \int_0^{\theta_M} f(y)e^{-\nu[y, \theta_M]}\nu(dy). \end{aligned}$$

Thus the distribution of the  $(M + 1)^{\text{th}}$  nonzero rate in the rejection representation  $T_{k_{M+1}}$  given  $\Xi_M$  is equal to the distribution of the  $(M + 1)^{\text{th}}$  rate from the inverse Lévy representation  $\theta_{M+1}$  given  $\Theta_M$  when  $\Xi_M = \Theta_M$ .  $\square$

**Appendix B: Truncation error bounds for self-product measures**

*Proof of Lemma 3.1.* Denote  $\tilde{\Theta} = \{\Theta^{(d)}, \Theta_K^{(d)}\}$ . Denote the marginal probability mass function (PMF) of  $E_N \in \mathcal{N}_d$  and  $E_{N,K} \in \mathcal{N}_d$  as  $P_N$  and  $P_{N,K}$ , and denote their PMFs given  $\tilde{\Theta}$  as  $f(x|\tilde{\Theta})$  and  $f_K(x|\tilde{\Theta})$  respectively.

$$\begin{aligned} D_{\text{TV}}(P_N, P_{N,K}) &= \frac{1}{2} \sum_{x \in \mathcal{N}_d} |P_N(x) - P_{N,K}(x)| \\ &= \frac{1}{2} \sum_{x \in \mathcal{N}_d} |\mathbb{E}[f(x|\tilde{\Theta})] - \mathbb{E}[f_K(x|\tilde{\Theta})]| \\ &\leq \frac{1}{2} \mathbb{P}(I_N \leq K) \sum_{x \in \mathcal{N}_d} |\mathbb{E}[f(x|\tilde{\Theta})|I_N \leq K] - \mathbb{E}[f_K(x|\tilde{\Theta})|I_N \leq K]| \\ &\quad + \frac{1}{2} \mathbb{P}(I_N > K) \sum_{x \in \mathcal{N}_d} |\mathbb{E}[f(x|\tilde{\Theta})|I_N > K] - \mathbb{E}[f_K(x|\tilde{\Theta})|I_N > K]| \end{aligned}$$

Conditioned on  $I_N \leq K$ ,  $f(x|\tilde{\Theta}) = f_K(x|\tilde{\Theta})$  under both the independent and categorical likelihood. So

$$\mathbb{E}[f(x|\tilde{\Theta})|I_N \leq K] - \mathbb{E}[f_K(x|\tilde{\Theta})|I_N \leq K] = 0.$$

By Fubini’s Theorem,

$$\begin{aligned} D_{\text{TV}}(P_N, P_{N,K}) &\leq \frac{1}{2} \mathbb{P}(I_N > K) \sum_{x \in \mathcal{N}_d} (\mathbb{E}[f(x|\tilde{\Theta})|I_N > K] + \mathbb{E}[f_K(x|\tilde{\Theta})|I_N > K]) \\ &= \frac{1}{2} \mathbb{P}(I_N > K) \left( \mathbb{E} \left[ \sum_{x \in \mathcal{N}_d} f(x|\tilde{\Theta}) \middle| I_N > K \right] + \mathbb{E} \left[ \sum_{x \in \mathcal{N}_d} f_K(x|\tilde{\Theta}) \middle| I_N > K \right] \right) \\ &= \mathbb{P}(I_N > K) = 1 - \mathbb{P}(I_N \leq K). \end{aligned} \quad \square$$

*Proof of Theorems 3.4 and 4.1.* Denote the set of indices

$$\mathcal{I}_{\ell,K} := \{i \in \mathbb{N}_{\neq}^d : 1 \leq i_1, \dots, i_\ell \leq K, K+1 \leq i_{\ell+1}, \dots, i_d < \infty\}$$

such that  $i \in \mathcal{I}_{\ell,K}$  indicates that the first  $\ell$  elements of  $i$  belong to the truncation, and the remaining  $d - \ell$  elements belong to the tail. By Jensen's inequality,

$$\begin{aligned} & \mathbb{P}(I_N \leq K \mid U_{1:K}, \Gamma_{1:K}) \\ &= \mathbb{E} \left[ \exp \left\{ N \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{i \in \mathcal{I}_{\ell,K}} \log \pi(\vartheta_i) \right\} \mid U_{1:K}, \Gamma_{1:K} \right] \\ &\geq \exp \left\{ N \sum_{\ell=0}^{d-1} \binom{d}{\ell} \mathbb{E} \left[ \sum_{i \in \mathcal{I}_{\ell,K}} \log \pi(\vartheta_i) \mid U_{1:K}, \Gamma_{1:K} \right] \right\}. \end{aligned} \quad (16)$$

This equation arises by noting that  $I_N \leq K$  if and only if for all  $i$  involving an index  $i_j > K$ , the count of edge  $i$  is 0 after  $N$  rounds; the factor  $\binom{d}{\ell}$  accounts for the fact that  $\vartheta_i = \prod_{j=1}^d \theta_{i_j}$  is independent of the ordering of the  $i_j$ .

Consider a single term  $\sum_{i \in \mathcal{I}_{\ell,K}} \log \pi(\vartheta_i)$  in the above sum. Since we are conditioning on  $U_{1:K}, \Gamma_{1:K}$ , we have that  $\theta_{i_1}, \dots, \theta_{i_\ell}$  are fixed in the expectation, and the remaining steps  $\Gamma_{K+1}, \Gamma_{K+2}, \dots$  are the ordered jumps of a unit rate homogeneous Poisson process on  $[\Gamma_K, \infty)$ . By the marking property of the Poisson process [47], conditioned on  $\Gamma_K$ , we have that  $(U_i, \Gamma_i)_{i=K+1}^\infty$  is a Poisson process on  $\mathbb{R}_+ \times [\Gamma_K, \infty)$  with rate measure  $g(du)d\gamma$ . Thus we apply the Slivnyak-Mecke theorem [40] to the remaining  $d - \ell$  indices to obtain

$$\begin{aligned} & \mathbb{E} \left[ \sum_{i \in \mathcal{I}_{\ell,K}} \log \pi(\vartheta_i) \mid U_{1:K}, \Gamma_{1:K} \right] \\ &= \mathbb{E} \left[ \sum_{i \in \mathcal{I}_{\ell,K}} \log \pi \left( \prod_{j=1}^{\ell} \theta_{i_j} \prod_{j=\ell+1}^d \tau(U_{i_j}, \Gamma_{i_j}) \right) \mid U_{1:K}, \Gamma_{1:K} \right] \\ &= \sum_{i_1 \neq \dots \neq i_\ell \leq K} \mathbb{E} \left[ \int_{\mathbb{R}_+^{d-\ell}} \log \pi \left( \prod_{j=1}^{\ell} \theta_{i_j} \prod_{j=\ell+1}^d \tau(U_{i_j}, \Gamma_K + \gamma_j) \right) \prod_{j=\ell+1}^d d\gamma_j \mid U_{1:K}, \Gamma_{1:K} \right] \\ &= \sum_{\substack{\mathcal{L} \subseteq [K] \\ |\mathcal{L}|=\ell}} \int_{[\Gamma_K, \infty)^{d-\ell}} \mathbb{E} \left[ \log \pi \left( \prod_{j \in \mathcal{L}} \theta_j \cdot \prod_{j=1}^{d-\ell} \tau(U'_j, \gamma_j) \right) \mid \theta_{1:K} \right] d\gamma_{1:d-\ell}, \end{aligned}$$

where  $U'_j \stackrel{\text{i.i.d.}}{\sim} g$ . Substitution of this expression into Eq. (16) yields the result of Theorem 4.1. Next, we consider the bound on the marginal probability  $\mathbb{P}(I_N \leq K)$ . By Jensen's inequality applied to Eq. (16) and following the previous derivation, we find that

$$\mathbb{P}(I_N \leq K) \geq \exp \left\{ N \sum_{\ell=0}^{d-1} \binom{d}{\ell} \mathbb{E} \left[ \sum_{i \in \mathcal{I}_{\ell,K}} \log \pi(\vartheta_i) \right] \right\},$$

and

$$\begin{aligned} & \mathbb{E} \left[ \sum_{i \in \mathcal{I}_{\ell, K}} \log \pi(\vartheta_i) \right] \\ &= \sum_{i_1 \neq \dots \neq i_\ell \leq K} \mathbb{E} \left[ \int_{[\Gamma_K, \infty)^{d-\ell}} \log \pi \left( \prod_{j=1}^{\ell} \theta_{i_j} \prod_{j=1}^{d-\ell} \tau(U'_j, \gamma_j) \right) d\gamma_{1:d-\ell} \right]. \end{aligned}$$

Using the fact that conditioned on  $\Gamma_K$ ,  $\Gamma_{1:K-1}$  are uniformly distributed on  $[0, \Gamma_K]$ , and that at most one  $i_j$  can be equal to  $K$ ,

$$\begin{aligned} & \sum_{i_1 \neq \dots \neq i_\ell \leq K} \mathbb{E} \left[ \int_{[\Gamma_K, \infty)^{d-\ell}} \log \pi \left( \prod_{j=1}^{\ell} \theta_{i_j} \prod_{j=1}^{d-\ell} \tau(U'_j, \gamma_j) \right) d\gamma_{1:d-\ell} \right] \\ &= \frac{(K-1)!}{(K-1-\ell)!} \mathbb{E} \left[ \Gamma_K^{-\ell} \int_{[0, \Gamma_K]^\ell \times [\Gamma_K, \infty)^{d-\ell}} \log \pi \left( \prod_{j=1}^d \tau(U_j, \gamma_j) \right) d\gamma_{1:d} \right] \\ &+ \ell \frac{(K-1)!}{(K-\ell)!} \mathbb{E} \left[ \Gamma_K^{-(\ell-1)} \int_{[0, \Gamma_K]^\ell \times [\Gamma_K, \infty)^{d-\ell}} \delta_{\gamma_1 = \Gamma_K} \log \pi \left( \prod_{j=1}^d \tau(U_j, \gamma_j) \right) d\gamma_{1:d} \right], \end{aligned}$$

where the first and second terms arise from portions of the sum where all indices satisfy  $i_j \neq K$  and one index satisfies  $i_j = K$ , respectively.

To complete the result, we study the asymptotics of the marginal probability bound. It follows from Eq. (7) that  $I_N < \infty$  almost surely. Therefore

$$\lim_{K \rightarrow \infty} \mathbb{E} \left[ \exp \left\{ N \sum_{\ell=0}^{d-1} \binom{d}{\ell} \sum_{i \in \mathcal{I}_{\ell, K}} \log \pi(\vartheta_i) \right\} \right] = \lim_{K \rightarrow \infty} \mathbb{P}(I_N \leq K) = 1.$$

It then follows from [28, Lemma B.1] and continuous mapping theorem that

$$\sum_{\ell=0}^{d-1} \left[ \binom{d}{\ell} \sum_{i \in \mathcal{I}_{\ell, K}} \log \pi(\vartheta_i) \right] \xrightarrow{p} 0 \quad \text{as } K \rightarrow \infty.$$

Since this sequence is monotonically increasing in  $K$ , we have that

$$B_K = \mathbb{E} \left[ \sum_{\ell=0}^{d-1} \left[ \binom{d}{\ell} \sum_{i \in \mathcal{I}_{\ell, K}} -\log \pi(\vartheta_i) \right] \right] \rightarrow 0 \quad \text{as } K \rightarrow \infty. \quad \square$$

**B.1. Proof of Theorem 3.6**

We first state an useful results which states that if one perturbs the probabilities of a countable discrete distribution by i.i.d. Gumbel(0, 1) random variables, the arg max of the resulting set is a sample from that distribution.



**Lemma B.1.** [28, Lemma 5.2] Let  $(p_j)_{j=1}^\infty$  be a countable collection of positive numbers such that  $\sum_j p_j < \infty$  and let  $\bar{p}_j = \frac{p_j}{\sum_k p_k}$ . If  $(W_j)_{j=1}^\infty$  are i.i.d. Gumbel(0,1) random variables, then  $\arg \max_{j \in \mathbb{N}} W_j + \log p_j$  exists, is unique a.s., and has distribution

$$\arg \max_{j \in \mathbb{N}} W_j + \log p_j \sim \text{Categorical}((\bar{p}_j)_{j=1}^\infty).$$

*Proof of Theorem 3.6.* Since the  $N$  edges from the categorical likelihood process are i.i.d. categorical random variables, by Jensen’s inequality,

$$\mathbb{P}(I_N \leq K) = \mathbb{E}[\mathbb{P}(I_N \leq K | \Theta)] = \mathbb{E}[\mathbb{P}(I_1 \leq K | \Theta)^N] \geq \mathbb{E}[\mathbb{P}(I_1 \leq K | \Theta)]^N.$$

Next, since  $\vartheta_i = \prod_{j=1}^d \theta_{i_j}$ , we can simulate a categorical variable with probabilities proportional to  $\vartheta_i$ ,  $i \in \mathbb{N}_d^\neq$  by sampling  $d$  indices  $(J_1, \dots, J_d)$  independently from a categorical distribution with probabilities proportional to  $(\theta_1, \theta_2, \dots)$  and discarding samples where  $J_j = J_k$  for some  $1 \leq j, k \leq d$ . Denote  $\theta'_k = \theta_k / \sum_k \theta_k$  to be the normalized rates,  $P_{J,K} := \sum_{j=J}^K \theta'_j$ , and the event  $\mathcal{Q} := \{J_j \neq J_k, \forall 1 \leq j, k \leq d\}$ . Then

$$\mathbb{P}(I_1 \leq K | \Theta) = \mathbb{P}(1 \leq J_j \leq K, 1 \leq j \leq d | \mathcal{Q}, \Theta).$$

Since the normalized rates  $\theta'_k$  are generated from the inverse Lévy representation, they are monotone decreasing. Therefore

$$\begin{aligned} \mathbb{P}(1 \leq J_j \leq K, 1 \leq j \leq d | \mathcal{Q}, \Theta) &\geq \frac{P_{1,K}}{1-0} \cdot \frac{P_{2,K}}{1-P_{1,1}} \cdots \frac{P_{d,K}}{1-P_{1,d-1}} \\ &\geq \left( \frac{P_{d,K}}{1-P_{1,d-1}} \right)^d. \end{aligned}$$

By Jensen’s inequality,

$$\mathbb{E}[\mathbb{P}(1 \leq J_j \leq K, 1 \leq j \leq d | \mathcal{Q}, \Theta)] \geq \mathbb{E} \left[ \frac{P_{d,K}}{1-P_{1,d-1}} \right]^d.$$

Note that for a categorical random variable  $J$  with class probabilities given by  $\theta'_j / (1 - P_{1,d-1})$ ,  $j \geq d$ , the quantity  $P_{d,K} / (1 - P_{1,d-1})$  is the probability that  $d \leq J \leq K$ . So by the infinite Gumbel-max sampling lemma,

$$\mathbb{E} \left[ \frac{P_{d,K}}{1-P_{1,d-1}} \right] = \mathbb{P} \left( d \leq \arg \max_{j \geq d} (\log \theta_j + W_j) \leq K \right), \quad W_j \stackrel{\text{i.i.d.}}{\sim} \text{Gumbel}(0, 1),$$

where we can replace  $\theta'_j$  with the unnormalized  $\theta_j$  because the normalization does not affect the arg max. Denoting

$$M_K := \max_{d \leq k \leq K} \log \nu^{\leftarrow}(\Gamma_k) + W_k, \quad M_{K+} := \sup_{k > K} \log \nu^{\leftarrow}(\Gamma_k) + W_k,$$

we have that

$$\mathbb{P}(I_N \leq K) \geq (1 - \mathbb{E}[\mathbb{P}(M_K < M_{K+} | \Gamma_K)])^{N \cdot d},$$

and so the remainder of the proof focuses on the conditional expectation. Conditioned on  $\Gamma_K$ ,

$$M_K \stackrel{d}{=} \max \left\{ \log \nu^{\leftarrow}(\Gamma_K) + W_K, \max_{d \leq k \leq K} \log \nu^{\leftarrow}(u_k) + W_k \right\}.$$

The cumulative distribution function and the probability density function of the Gumbel distribution  $\text{Gumbel}(0, 1)$  is

$$F(x) = e^{-e^{-x}}, \quad f(x) = e^{-(x+e^{-x})}.$$

So

$$\mathbb{P}(\log \nu^{\leftarrow}(\Gamma_K) + W_K \leq x \mid \Gamma_K) = e^{-\nu^{\leftarrow}(\Gamma_K)e^{-x}},$$

and

$$\mathbb{P}(\log \nu^{\leftarrow}(u_k) + W_k \leq x \mid \Gamma_K) = \int_0^1 e^{-\nu^{\leftarrow}(\Gamma_K u)e^{-x}} du.$$

Therefore,

$$\mathbb{P}(M_K \leq x \mid \Gamma_K) = \left( \int_0^1 e^{-\nu^{\leftarrow}(\Gamma_K u)e^{-x}} du \right)^{K-d} e^{-\nu^{\leftarrow}(\Gamma_K)e^{-x}}.$$

Denote  $Q(u, t) = e^{-\nu^{\leftarrow}(u)e^{-t}}$ , and

$$\mathbb{P}(M_K \leq x \mid \Gamma_K) = \left( \int_0^1 Q(\Gamma_K u, x) du \right)^{K-d} Q(\Gamma_K, x).$$

Conditioned on  $\Gamma_K$ , the tail  $\Gamma_{K+1}, \Gamma_{K+2}, \dots$  is a unit rate homogeneous Poisson process on  $[\Gamma_K, \infty)$  that is independent of  $\Gamma_1, \dots, \Gamma_{K-1}$ . So conditioned on  $\Gamma_K$ ,

$$M_{K+} \stackrel{d}{=} \sup_{k \geq 1} \log \nu^{\leftarrow}(\Gamma_K + \Gamma'_k) + W_k,$$

where  $\Gamma'_k$  is unit rate of homogeneous Poisson process on  $\mathbb{R}_+$ . Since  $\Gamma'_k$  is a Poisson process,  $\log \nu^{\leftarrow}(\Gamma_K + \Gamma'_k) + W_k$  is also a Poisson process with the rate measure

$$\left( \int_0^\infty e^{-(t - \log \nu^{\leftarrow}(\Gamma_K + \gamma)) - e^{-(t - \log \nu^{\leftarrow}(\Gamma_K + \gamma))}} d\gamma \right) dt.$$

$\mathbb{P}(M_{K+} \leq x \mid \Gamma_K)$  is the probability that no atom of the Poisson process is greater than  $x$ . For a Poisson process with rate measure  $\mu(dt)$ , this probability is  $e^{-\int_x^\infty \mu(dt)}$ ,

$$\mathbb{P}(M_{K+} \leq x \mid \Gamma_K) = e^{-\int_x^\infty \left( \int_0^\infty e^{-(t - \log \nu^{\leftarrow}(\Gamma_K + \gamma)) - e^{-(t - \log \nu^{\leftarrow}(\Gamma_K + \gamma))}} d\gamma \right) dt}$$

$$\begin{aligned}
 &= e^{\int_0^\infty (e^{-\nu^\leftarrow(\Gamma_K+\gamma)e^{-x}} - 1) d\gamma} \\
 &= e^{\int_0^\infty Q(\Gamma_K+\gamma, x) - 1 d\gamma},
 \end{aligned}$$

where the second equation comes from the fact that the inner integrand is the probability density function of a Gumbel distribution. Therefore,

$$\begin{aligned}
 &\mathbb{P}(M_K < M_{K+} \mid \Gamma_K) \\
 &= \int_{-\infty}^\infty \mathbb{P}(M_K \leq x \mid \Gamma_K) \frac{d}{dx} \mathbb{P}(M_{K+} \leq x \mid \Gamma_K) dx \\
 &= \int_{-\infty}^\infty Q(\Gamma_K, x) \left( \int_0^1 Q(\Gamma_K u, x) du \right)^{K-d} \left( \frac{d}{dx} e^{\int_0^\infty Q(\Gamma_K+\gamma, x) - 1 d\gamma} \right) dx.
 \end{aligned}$$

For the categorical variable  $J$  with class probabilities given by  $\theta'_j / (1 - P_{1,d-1})$ ,  $j \geq d$ , it holds that  $\mathbb{P}(d \leq J \leq K) \uparrow 1$  as  $K \rightarrow \infty$ . By the monotone convergence theorem

$$B_K = \mathbb{P} \left( \arg \max_{j \geq d} (\log \theta_j + W_j) > K \right) = 1 - \mathbb{E} \left[ \frac{P_{d,K}}{1 - P_{1,d}} \right] \rightarrow 0. \quad \square$$

**Appendix C: Error bounds for edge-exchangeable networks**

**C.1. Rejection representation**

We first derive the specific form of  $B_K$  in Corollary 3.3 for the rejection representation from Eq. (5), where

$$\tau(U, \Gamma) = \mu^\leftarrow(\Gamma) \mathbb{1} \left[ \frac{d\nu}{d\mu}(\mu^\leftarrow(\Gamma)) \geq U \right],$$

and  $U \sim \text{Unif}[0, 1]$ . So in Corollary 3.3,

$$\begin{aligned}
 &B_{K,1} \\
 &= \mathbb{E} \left[ \int_{\mathbb{R}_+^4} -\log \pi(\tau(u_1, \gamma_1 + \Gamma_K) \tau(u_2, \gamma_2 + \Gamma_K)) d\gamma_1 d\gamma_2 g(u_1) du_1 g(u_2) du_2 \right] \\
 &= \mathbb{E} \left[ \int_{\Gamma_K}^\infty \int_{\Gamma_K}^\infty \int_0^1 \int_0^1 -\log \pi(\tau(u_1, \gamma_1) \tau(u_2, \gamma_2)) du_1 du_2 d\gamma_1 d\gamma_2 \right].
 \end{aligned}$$

Since  $\pi(0) = 1$ ,  $\log \pi(0) = 0$ , we can take the indicator in  $\tau$  out of the function  $\log \pi(\tau(u_1, \gamma_1) \tau(u_2, \gamma_2))$  to obtain

$$\begin{aligned}
 &\int_0^1 \int_0^1 \log \pi(\tau(u_1, \gamma_1) \tau(u_2, \gamma_2)) du_1 du_2 \\
 &= \int_0^1 \int_0^1 \log \pi(\mu^\leftarrow(\gamma_1) \mu^\leftarrow(\gamma_2)) \mathbb{1} \left[ \frac{d\nu}{d\mu}(\mu^\leftarrow(\gamma_1)) \geq u_1 \right] \mathbb{1} \left[ \frac{d\nu}{d\mu}(\mu^\leftarrow(\gamma_2)) \geq u_2 \right] du_1 du_2
 \end{aligned}$$

$$= \log \pi(\mu^{\leftarrow}(\gamma_1)\mu^{\leftarrow}(\gamma_2)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\gamma_1)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\gamma_2)).$$

Transforming variables via  $x_1 = \mu^{\leftarrow}(\gamma_1)$  and  $x_2 = \mu^{\leftarrow}(\gamma_2)$ , and noting that  $\mu^{\leftarrow}(\Gamma_K) \geq x \iff \Gamma_K \leq \mu[x, \infty)$ ,

$$\begin{aligned} B_{K,1} &= \mathbb{E} \left[ \int_0^{\mu^{\leftarrow}(\Gamma_K)} \int_0^{\mu^{\leftarrow}(\Gamma_K)} -\log \pi(x_1 x_2) \frac{d\nu}{d\mu}(x_1) \frac{d\nu}{d\mu}(x_2) \mu(dx_1) \mu(dx_2) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+^2} -\log \pi(x_1 x_2) \mathbb{1}[x_1 \leq \mu^{\leftarrow}(\Gamma_K)] \mathbb{1}[x_2 \leq \mu^{\leftarrow}(\Gamma_K)] \nu(dx_1) \nu(dx_2) \right] \\ &= \int_{\mathbb{R}_+^2} -\log \pi(x_1 x_2) \mathbb{E}[\mathbb{1}[\Gamma_K \leq \mu[\max\{x_1, x_2\}, \infty)]] \nu(dx_1) \nu(dx_2) \\ &= \int_{\mathbb{R}_+^2} -\log \pi(x_1 x_2) F_K(\mu[\max\{x_1, x_2\}, \infty)) \nu(dx_1) \nu(dx_2), \end{aligned} \quad (17)$$

where  $F_K$  is the cumulative distribution function of  $\Gamma_K$ . Using the same variable transformation again,

$$\begin{aligned} &\frac{1}{2(K-1)} B_{K,2} \\ &= \mathbb{E} \left[ \int_0^{\Gamma_K} \int_{\mathbb{R}_+^3} -\frac{1}{\Gamma_K} \log \pi(\tau(u_1, \gamma_1) \tau(u_2, \gamma_2 + \Gamma_K)) g(u_1) du_1 g(u_2) du_2 d\gamma_1 d\gamma_2 \right] \\ &= \mathbb{E} \left[ \int_{\mu^{\leftarrow}(\Gamma_K)}^{\infty} \int_0^{\mu^{\leftarrow}(\Gamma_K)} -\frac{1}{\Gamma_K} \log \pi(x_1 x_2) \frac{d\nu}{d\mu}(x_1) \frac{d\nu}{d\mu}(x_2) \mu(dx_1) \mu(dx_2) \right] \\ &= \int_{\mathbb{R}_+^2} -\mathbb{E} \left[ \frac{1}{\Gamma_K} \mathbb{1}[\mu[x_1, \infty) \leq \Gamma_K \leq \mu[x_2, \infty)] \right] \log \pi(x_1 x_2) \nu(dx_1) \nu(dx_2). \end{aligned}$$

$$\begin{aligned} B_{K,3} &= 2\mathbb{E} \left[ \int_{\mathbb{R}_+^3} -\log \pi(\tau(u_1, \Gamma_K) \tau(u_2, \gamma_2 + \Gamma_K)) g(u_1) du_1 g(u_2) du_2 d\gamma_2 \right] \\ &= 2\mathbb{E} \left[ \int_{\Gamma_K}^{\infty} -\log \pi(\mu^{\leftarrow}(\Gamma_K) \mu^{\leftarrow}(\gamma_2)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\gamma_2)) d\gamma_2 \right] \\ &= 2\mathbb{E} \left[ \int_0^{\mu^{\leftarrow}(\Gamma_K)} -\log \pi(\mu^{\leftarrow}(\Gamma_K) x) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \frac{d\nu}{d\mu}(x) \mu(dx) \right] \\ &= 2\mathbb{E} \left[ \int_{\mathbb{R}_+} -\log \pi(\mu^{\leftarrow}(\Gamma_K) x) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \mathbb{1}[\Gamma_K \leq \mu[x, \infty)] \nu(dx) \right]. \end{aligned}$$

### C.2. Beta-independent Bernoulli process network

**Dense network** When  $\alpha = 0$ ,

$$\nu(d\theta) = \gamma \lambda \theta^{-1} (1 - \theta)^{\lambda-1} d\theta, \quad \mu(d\theta) = \gamma \lambda \theta^{-1} d\theta,$$

and

$$\frac{d\nu}{d\mu} = (1 - \theta)^{\lambda-1}, \quad \mu([x, 1]) = -\gamma' \log x, \quad \mu^{\leftarrow}(u) = e^{-u/\gamma'}.$$

Substituting  $\nu(dx), \mu(dx)$  and  $\pi(\theta)$  into Eq. (17) and noting that the integrand is symmetric around the line  $x_1 = x_2$ ,

$$B_{K,1} = 2 \int_0^1 \int_{x_1}^1 -\log(1 - x_1x_2) F_K(\mu[x_2, 1]) \gamma \lambda x_2^{-1} (1 - x_2)^{\lambda-1} dx_2 \nu(dx_1).$$

Next, note that  $0 \leq F_K(\mu[x_2, 1]) \leq F_K(\mu[x_1, 1])$  when  $x_2 \geq x_1$  and

$$0 \geq \log(1 - x_1x_2) \geq -\left(\frac{1}{1 - x_1x_2} - 1\right) = -\left(\frac{x_1x_2}{1 - x_1x_2}\right) \geq -\left(\frac{x_1x_2}{1 - x_1}\right).$$

So

$$\begin{aligned} B_{K,1} &\leq 2\gamma\lambda \int_0^1 F_K(\mu([x_1, 1])) \int_{x_1}^1 \frac{x_1x_2}{1 - x_1} x_2^{-1} (1 - x_2)^{\lambda-1} dx_2 \nu(dx_1) \\ &= 2\gamma^2\lambda \int_0^1 (1 - x)^{2(\lambda-1)} F_K(\mu([x, 1])) dx. \end{aligned}$$

For any  $a \in (0, 1)$ , dividing the integral into two parts and bounding each part separately,

$$B_{K,1} \leq 2\gamma^2\lambda \left[ \int_0^a (1 - x)^{2(\lambda-1)} dx + \int_a^1 F_K(-\gamma' \log x) (1 - x)^{2(\lambda-1)} dx \right].$$

Assume for the moment that  $\lambda \neq 0.5$ . Use the fact that  $F_K(t) \leq (3t/K)^K$  and note also that when  $x \in [a, 1]$ ,  $-\log x \leq [-\log a/(1 - a)](1 - x)$ ,

$$\begin{aligned} &\frac{1}{2\gamma^2\lambda} B_{K,1} \\ &\leq \frac{1 - (1 - a)^{2\lambda-1}}{2\lambda - 1} + \int_a^1 \left( \frac{3\gamma' [-\log a/(1 - a)](1 - x)}{K} \right)^K (1 - x)^{2\lambda-2} dx \\ &= \frac{1 - (1 - a)^{2\lambda-1}}{2\lambda - 1} + \left( -\frac{3\gamma' \log a}{K} \right)^K \frac{1}{K + 2\lambda - 1} (1 - a)^{2\lambda-1}. \end{aligned}$$

It can be seen that there exists a constant value  $c > 1$  such that  $3\gamma' \log c = 1/c$ . Setting  $a = c^{-K}$  and using the first order Taylor's expansion to approximate the first term, it can be seen that

$$\frac{1 - (1 - a)^{2\lambda-1}}{2\lambda - 1} \sim c^{-K}, \quad \left( -\frac{3\gamma' \log a}{K} \right)^K \frac{1}{K + 2\lambda - 1} (1 - a)^{2\lambda-1} \sim \frac{1}{K} c^{-K}.$$

Therefore, there exists  $K_0$  such that when  $K > K_0$ ,

$$B_{K,1} \leq 4\gamma^2\lambda c^{-K}.$$

If  $\lambda = 0.5$ ,

$$\begin{aligned} B_{K,1} &\leq 2\gamma^2\lambda \int_0^1 (1-x)^{-1} F_K(\mu([x, 1])) dx \\ &\leq 2\gamma^2\lambda \left[ -\log(1-a) + \left(\frac{3\gamma'}{K}\right)^K \frac{1}{K} (-\log a)^K \right], \end{aligned}$$

which can be bounded similarly by choosing the same  $c$  and setting  $a = c^{-K}$ . Therefore we can find a constant  $K_0$  and for  $K > K_0$ ,

$$B_{K,1} \leq 4\gamma^2\lambda c^{-K}.$$

Next, the term  $B_{K,2}$  is bounded via

$$\begin{aligned} &\frac{1}{2}B_{K,2} \\ &= \int_0^1 \int_{x_1}^1 -\log(1-x_1x_2) [F_{K-1}(\mu([x_1, 1])) - F_{K-1}(\mu([x_2, 1]))] \nu(dx_2)\nu(dx_1) \\ &\leq \int_0^1 \int_{x_1}^1 -\log(1-x_1x_2) F_{K-1}(\mu[x_1, 1]) \nu(dx_2)\nu(dx_1). \end{aligned}$$

This has exactly the same form as  $B_{K,1}$ , except that the CDF of  $\Gamma_K$  is replaced with that of  $\Gamma_{K-1}$ . Therefore, it can be shown that for large  $K$ ,

$$B_{K,2} \leq 4\gamma^2\lambda c^{-(K-1)}.$$

Finally,  $B_{K,3}$  may be expressed as

$$\begin{aligned} &\frac{1}{2}B_{K,3} \\ &= \mathbb{E} \left[ \int_0^1 -\log(1 - \mu^{\leftarrow}(\Gamma_K)x) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \mathbf{1}[x \leq \mu^{\leftarrow}(\Gamma_K)] \right] \nu(dx) \\ &= \gamma\lambda \mathbb{E} \left[ \int_0^{\mu^{\leftarrow}(\Gamma_K)} -\log(1 - \mu^{\leftarrow}(\Gamma_K)x) (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda-1} x^{-1} (1-x)^{\lambda-1} dx \right]. \end{aligned}$$

Since  $\log(1 - \mu^{\leftarrow}(\Gamma_K)x) \geq -\mu^{\leftarrow}(\Gamma_K)x / (1 - \mu^{\leftarrow}(\Gamma_K))$ ,

$$\begin{aligned} B_{K,3} &\leq 2\gamma\lambda \mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda-2} \mu^{\leftarrow}(\Gamma_K) \int_0^{\mu^{\leftarrow}(\Gamma_K)} (1-x)^{\lambda-1} dx \right] \\ &\geq 2\gamma \mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda-2} \mu^{\leftarrow}(\Gamma_K) \right] \\ &= 2\gamma \int_0^\infty (1 - e^{-x/\gamma'})^{\lambda-2} e^{-x/\gamma'} \frac{x^{K-1}}{\Gamma(K)} e^{-x} dx. \end{aligned}$$

We split the analysis of this term into two cases. In the first case, assuming  $\lambda \geq 2$ , we have that

$$B_{K,3} \leq 2\gamma \int_0^\infty \frac{x^{K-1}}{\Gamma(K)} e^{-x(1+\gamma')/\gamma'} dx = 2\gamma \left( \frac{\gamma'}{1+\gamma'} \right)^K.$$

On the other hand, if  $\lambda < 2$ , we bound the integral over  $[0, \gamma']$  and over  $[\gamma', \infty)$  separately. Since  $1 - e^{-x} \geq x^2$  for  $x \in [0, 1]$ ,

$$\begin{aligned} & \frac{1}{2\gamma} B_{K,3} \\ & \leq \int_0^{\gamma'} \left( \frac{x}{\gamma'} \right)^{2(\lambda-2)} \frac{x^{K-1}}{\Gamma(K)} e^{-x(1+\gamma')/\gamma'} dx + (1 - e^{-1})^{\lambda-2} \int_{\gamma'}^\infty \frac{x^{K-1}}{\Gamma(K)} e^{-x(1+\gamma')/\gamma'} dx \\ & \leq \gamma'^{2(2-\lambda)} \int_0^{\gamma'} \frac{x^{K-1+2(\lambda-2)}}{\Gamma(K)} dx + (1 - e^{-1})^{\lambda-2} \int_0^\infty \frac{x^{K-1}}{\Gamma(K)} e^{-x(1+\gamma')/\gamma'} dx \\ & = \frac{1}{\Gamma(K)} \frac{\gamma'^K}{K + 2(\lambda - 2)} + (1 - e^{-1})^{\lambda-2} \left( \frac{\gamma'}{1 + \gamma'} \right)^K. \end{aligned}$$

As  $K \rightarrow \infty$ , the second term will dominate the first term, so when  $\lambda - 2 < 0$ , the following inequality holds for large  $K$ ,

$$B_{K,3} \leq 4\gamma(1 - e^{-1})^{\lambda-2} \left( \frac{\gamma'}{1 + \gamma'} \right)^K.$$

$B_K = B_{K,1} + B_{K,2} + B_{K,3}$  and as  $K \rightarrow \infty$ ,  $B_{K,3}$  will dominate  $B_{K,1}$  and  $B_{K,2}$ . So there exists a  $K_0 \in \mathbb{N}$  such that for  $K > K_0$ ,

$$B_K \leq 12\gamma(1 - e^{-1})^{\lambda-2} \left( \frac{\gamma'}{1 + \gamma'} \right)^K \rightarrow 0.$$

**Sparse network** When  $\alpha > 0$ ,

$$\nu(d\theta) = \gamma'\theta^{-1-\alpha}(1 - \theta)^{\lambda+\alpha-1}d\theta, \quad \mu(d\theta) = \gamma'\theta^{-1-\alpha}d\theta,$$

and

$$\frac{d\nu}{d\mu} = (1 - \theta)^{\lambda+\alpha-1}, \quad \mu([x, 1]) = \gamma'\alpha^{-1}(x^{-\alpha} - 1), \quad \mu^\leftarrow(u) = \left( 1 + \frac{\alpha u}{\gamma'} \right)^{-1/\alpha}.$$

Similar to the case when  $\alpha = 0$ ,

$$B_{K,1} = 2 \int_0^1 \int_{x_1}^1 -\log(1 - x_1x_2) F_K(\mu([x_2, 1])) \gamma' x_2^{-1-\alpha} (1 - x_2)^{\lambda+\alpha-1} dx_2 \nu(dx_1).$$

Since  $\log(1 - x_1x_2) \geq -x_1x_2/(1 - x_1)$ ,

$$B_{K,1} \leq 2\gamma'^2 \int_0^1 F_K(\gamma'\alpha^{-1}(x^{-\alpha} - 1)) x^{-2\alpha} (1 - x)^{2[(\lambda+\alpha)-1]} dx$$

$$\begin{aligned}
&\leq 2\gamma'^2 \int_0^1 \frac{(\gamma'\alpha^{-1}x^{-\alpha}(1-x^\alpha))^K}{\Gamma(K+1)} x^{-2\alpha}(1-x)^{2[(\lambda+\alpha)-1]} dx \\
&\leq 2\gamma'^2 \int_0^1 \frac{(\gamma'\alpha^{-1}x^{-\alpha})^K}{\Gamma(K+1)} x^{-2\alpha}(1-x)^{2[(\lambda+\alpha)-1]} dx \\
&\leq 2\alpha^2 \int_0^1 \frac{(\gamma'\alpha^{-1}x^{-\alpha})^{K+2}}{\Gamma(K+1)} dx.
\end{aligned}$$

Denoting  $t = \gamma'\alpha^{-1}x^{-\alpha}$ , then

$$\begin{aligned}
B_{K,1} &\leq 2\alpha(\gamma'\alpha^{-1})^{1/\alpha} \int_{\gamma'\alpha^{-1}}^{\infty} \frac{t^{K+1-1/\alpha}}{\Gamma(K+1)} dt \\
&\leq 2\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} \frac{\Gamma(K+2-1/\alpha)}{\Gamma(K+1)} \int_{\gamma'\alpha^{-1}}^{\infty} \frac{t^{K+1-1/\alpha}}{\Gamma(K+2-1/\alpha)} e^{-t} dt \\
&\leq 2\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} \frac{\Gamma(K+2-1/\alpha)}{\Gamma(K+1)}.
\end{aligned}$$

By Stirling's formula,

$$B_{K,1} \leq 2\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} K^{\frac{\alpha-1}{\alpha}}.$$

Now we consider the error bound for  $B_{K,2}$ . As we have shown in the example when  $\alpha = 0$ , here we can obtain that

$$B_{K,2} \leq 2\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} (K-1)^{\frac{\alpha-1}{\alpha}}.$$

Similar to  $B_{K,1}$  and  $B_{K,2}$ ,

$$\begin{aligned}
&B_{K,3} \\
&= 2\mathbb{E} \left[ \int_0^1 -\log(1 - \mu^{\leftarrow}(\Gamma_K)x) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \mathbf{1}(\Gamma_K \leq \mu^{\leftarrow}[x, 1]) \nu(dx) \right] \\
&\leq 2\gamma'\mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda+\alpha-1} \int_0^{\mu^{\leftarrow}(\Gamma_K)} \frac{\mu^{\leftarrow}(\Gamma_K)x}{1-x} x^{-1-\alpha}(1-x)^{\lambda+\alpha-1} dx \right] \\
&= 2\gamma'\mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda+\alpha-1} \mu^{\leftarrow}(\Gamma_K) \int_0^{\mu^{\leftarrow}(\Gamma_K)} x^{-\alpha}(1-x)^{\lambda+\alpha-2} dx \right].
\end{aligned}$$

We again split our analysis into two cases. First, suppose that  $\lambda + \alpha - 2 \geq 0$ . Then

$$\begin{aligned}
&B_{K,3} \\
&\leq \frac{2\gamma'}{1-\alpha} \mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda+\alpha-1} \mu^{\leftarrow}(\Gamma_K)^{2-\alpha} \right] \\
&= \frac{2\gamma'}{1-\alpha} \int_0^\infty \left[ 1 - \left( 1 + \frac{\alpha x}{\gamma'} \right)^{-1/\alpha} \right]^{\lambda+\alpha-1} \left( 1 + \frac{\alpha x}{\gamma'} \right)^{-(2-\alpha)/\alpha} \frac{x^{K-1}}{\Gamma(K)} e^{-x} dx.
\end{aligned}$$



Since  $\alpha < 1$ ,

$$\left(1 + \frac{\alpha x}{\gamma'}\right)^{-(2-\alpha)/\alpha} \leq \left(\frac{\alpha x}{\gamma'}\right)^{-(2-\alpha)/\alpha},$$

and so

$$\begin{aligned} B_{K,3} &\leq \frac{2\gamma'}{1-\alpha} \int_0^\infty \left(\frac{\alpha}{\gamma'}\right)^{-(2-\alpha)/\alpha} \frac{x^{K-1-(2-\alpha)/\alpha}}{\Gamma(K)} e^{-x} dx \\ &\leq \frac{2\gamma'}{1-\alpha} \left(\frac{\alpha}{\gamma'}\right)^{-(2-\alpha)/\alpha} \frac{\Gamma(K-(2-\alpha)/\alpha)}{\Gamma(K)} \\ &\sim \frac{2\gamma'}{1-\alpha} \left(\frac{\alpha}{\gamma'}\right)^{-(2-\alpha)/\alpha} K^{-(2-\alpha)/\alpha}, \end{aligned}$$

where the last equation is obtained from Stirling's formula.

On the other hand, if  $\lambda + \alpha - 2 < 0$ ,

$$\begin{aligned} &B_{K,3} \\ &\leq 2\gamma' \mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{\lambda+\alpha-1} \mu^{\leftarrow}(\Gamma_K) \int_0^{\mu^{\leftarrow}(\Gamma_K)} x^{-\alpha} (1-x)^{\lambda+\alpha-2} dx \right] \\ &\leq \frac{2\gamma'}{1-\alpha} \mathbb{E} \left[ (1 - \mu^{\leftarrow}(\Gamma_K))^{2(\lambda+\alpha)-3} \mu^{\leftarrow}(\Gamma_K)^{2-\alpha} \right] \\ &= \frac{2\gamma'}{1-\alpha} \int_0^\infty \left[ 1 - \left(1 + \frac{\alpha x}{\gamma'}\right)^{-1/\alpha} \right]^{2(\lambda+\alpha)-3} \left(1 + \frac{\alpha x}{\gamma'}\right)^{-(2-\alpha)/\alpha} \frac{x^{K-1}}{\Gamma(K)} e^{-x} dx. \end{aligned}$$

If  $2(\lambda + \alpha) - 3 \geq 0$ , we get the same result as in the case when  $\lambda + \alpha - 2 \geq 0$ . When  $2(\lambda + \alpha) - 3 \leq 0$ , note that we can find an  $x_0$  such that when  $x \in [0, x_0]$ ,

$$1 - \left(1 + \frac{\alpha x}{\gamma'}\right)^{-1/\alpha} \geq x^2.$$

So

$$\begin{aligned} &B_{K,3} \\ &\leq \frac{2\gamma'}{1-\alpha} \left\{ \int_0^{x_0} x^{4(\lambda+\alpha)-6} \left(1 + \frac{\alpha x}{\gamma'}\right)^{-(2-\alpha)/\alpha} \frac{x^{K-1}}{\Gamma(K)} e^{-x} dx \right. \\ &\quad \left. + \left[ 1 - \left(1 + \frac{\alpha x_0}{\gamma'}\right)^{-1/\alpha} \right]^{2(\lambda+\alpha)-3} \int_{x_0}^\infty \left(1 + \frac{\alpha x}{\gamma'}\right)^{-(2-\alpha)/\alpha} \frac{x^{K-1}}{\Gamma(K)} e^{-x} dx \right\} \\ &\leq \frac{\gamma'}{1-\alpha} \left(\frac{\alpha}{\gamma'}\right)^{-(2-\alpha)/\alpha} \left\{ \int_0^{x_0} \frac{x^{K-1+4(\lambda+\alpha)-6-(2-\alpha)/\alpha}}{\Gamma(K)} e^{-x} dx \right. \\ &\quad \left. + \left[ 1 - \left(1 + \frac{\alpha x_0}{\gamma'}\right)^{-1/\alpha} \right]^{2(\lambda+\alpha)-3} \frac{\Gamma(K-(2-\alpha)/\alpha)}{\Gamma(K)} \right\}. \end{aligned}$$

Because here we assume that  $2(\lambda + \alpha) - 3 < 0$ , the second term will dominate. So in this case we obtain that for large  $K$ ,

$$B_{K,3} \leq \frac{4\gamma'}{1-\alpha} \left(\frac{\alpha}{\gamma'}\right)^{-(2-\alpha)/\alpha} \left[1 - \left(1 + \frac{\alpha x_0}{\gamma'}\right)^{-1/\alpha}\right]^{2(\lambda+\alpha)-3} K^{-(2-\alpha)/\alpha}.$$

Asymptotically,  $B_{K,2}$  will dominate  $B_{K,1}$  and  $B_{K,3}$ , so there exists  $K_0 \in \mathbb{N}$  such that when  $K > K_0$ ,

$$B_K \leq 6\alpha(\gamma'\alpha^{-1})^{1/\alpha} e^{\gamma'\alpha^{-1}} (K-1)^{\frac{\alpha-1}{\alpha}} \rightarrow 0.$$

### C.3. Gamma-independent Poisson network

**Dense network** When  $\alpha = 0$ ,

$$\nu(d\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}d\theta, \quad \mu(d\theta) = \gamma\lambda\theta^{-1}(1+\lambda\theta)^{-1}d\theta.$$

In this case,

$$\frac{d\nu}{d\mu} = (1+\lambda\theta)e^{-\lambda\theta}, \quad \mu[x, \infty) = \gamma\lambda \log(1 + (\lambda x)^{-1}), \quad \mu^{\leftarrow}(u) = \frac{1}{\lambda(e^{(\gamma\lambda)^{-1}u} - 1)}.$$

For Poisson distribution,  $\pi(\theta) = e^{-\theta}$ , so

$$B_{K,1} = \int_{\mathbb{R}_+^2} x_1 x_2 F_K(\mu[\max\{x_1, x_2\}, \infty)) \nu(dx_1) \nu(dx_2).$$

Note that the integrand is symmetric around the line  $x_1 = x_2$ , so we only need to compute the integral above the line  $x_1 = x_2$ . In this region,  $e^{-\lambda x_2} \leq e^{-\lambda x_1}$  and  $0 \leq F_K(\mu[x_2, \infty)) \leq F_K(\mu[x_1, \infty)) \leq 1$ . So

$$\begin{aligned} B_{K,1} &= 2 \int_{\mathbb{R}_+} x_1 \int_{x_1}^{\infty} x_2 F_K(\mu[x_2, \infty)) \lambda \gamma x_2^{-1} e^{-\lambda x_2} dx_2 \nu(dx_1) \\ &\leq 2 \int_{\mathbb{R}_+} x_1 F_K(\mu[x_1, \infty)) \gamma e^{-\lambda x_1} \nu(dx_1) \\ &= 2\gamma^2 \lambda \int_{\mathbb{R}_+} F_K(\mu[x, \infty)) e^{-2\lambda x} dx. \end{aligned}$$

For any  $a > 0$ , we divide the integral into two parts and bound each part separately. We denote  $b = \log(1 + (\lambda a)^{-1})$  and use the fact that  $\int_0^a e^{-2\lambda x} dx \leq a$  and  $F_K(t) \leq t^K/K! \leq (3t/K)^K$ . So

$$\begin{aligned} B_{K,1} &\leq 2\gamma^2 \lambda \left[ \int_0^a e^{-2\lambda x} dx + F_K(\mu[a, \infty)) \int_a^{\infty} e^{-2\lambda x} dx \right] \\ &\leq 2\gamma^2 \lambda \left[ a + \frac{1}{2\lambda} F_K(\gamma\lambda \log(1 + (\lambda a)^{-1})) \right] \end{aligned}$$

$$\leq 2\gamma^2\lambda \left[ (\lambda(e^b - 1))^{-1} + \frac{1}{\lambda} \left( \frac{3\gamma\lambda b}{K} \right)^K \right].$$

Setting two terms equal and use the fact that  $(e^b - 1)^{-1} \approx e^{-b}$  when  $b$  is large, we get  $b = KW_0((3\gamma\lambda)^{-1})$  and  $W_0$  is defined by

$$W_0(y) = x \iff xe^x = y.$$

Therefore,

$$B_{K,1} \leq \frac{4\gamma^2}{e^{KW_0((3\gamma\lambda)^{-1})} - 1} \sim 4\gamma^2 e^{-KW_0((3\gamma\lambda)^{-1})}.$$

Similarly,

$$\begin{aligned} & \frac{1}{2(K-1)} B_{K,2} \\ &= \int_{\mathbb{R}_+^2} x_1 x_2 \mathbb{E} \left[ \frac{1}{\Gamma_K} \mathbf{1}(\mu[x_2, \infty) \leq \Gamma_K \leq \mu[x_1, \infty)) \right] \nu(dx_1) \nu(dx_2). \end{aligned}$$

Note that

$$\begin{aligned} & \mathbb{E} \left[ \frac{1}{\Gamma_K} \mathbf{1}(\mu[x_2, \infty) \leq \Gamma_K \leq \mu[x_1, \infty)) \right] \\ &= \frac{1}{K-1} [F_{K-1}(\mu[x_1, \infty)) - F_{K-1}(\mu[x_2, \infty))]. \end{aligned}$$

Keeping only the positive part,

$$\begin{aligned} B_{K,2} &\leq 2 \int_{\mathbb{R}_+} x_1 F_{K-1}(\mu[x_1, \infty)) \int_{x_1}^{\infty} \gamma \lambda e^{-\lambda x_2} dx_2 \nu(dx_1) \\ &= 2\gamma^2 \lambda \int_{\mathbb{R}_+} F_{K-1}(\mu[x, \infty)) e^{-2\lambda x} dx. \end{aligned}$$

This has the same form as  $B_{K,1}$ , so

$$B_{K,2} \leq \frac{4\gamma^2}{e^{(K-1)W_0((3\gamma\lambda)^{-1})} - 1} \sim 4\gamma^2 e^{-(K-1)W_0((3\gamma\lambda)^{-1})}.$$

Next,

$$\begin{aligned} & B_{K,3} \\ &= 2\mathbb{E} \left[ \int_{\mathbb{R}_+} x \mu^{\leftarrow}(\Gamma_K) \frac{d\nu}{d\mu}(\mu^{\leftarrow}(\Gamma_K)) \mathbf{1}[\Gamma_K \leq \mu[x, \infty)] \right] \nu(dx) \\ &= 2\mathbb{E} \left[ \int_0^{\mu^{\leftarrow}(\Gamma_K)} x \mu^{\leftarrow}(\Gamma_K) (1 + \lambda \mu^{\leftarrow}(\Gamma_K)) e^{-\lambda \mu^{\leftarrow}(\Gamma_K)} \gamma \lambda x^{-1} e^{-\lambda x} dx \right] \end{aligned}$$

$$\begin{aligned} &= 2\gamma \mathbb{E} \left[ \mu^{\leftarrow}(\Gamma_K) (1 + \lambda \mu^{\leftarrow}(\Gamma_K)) e^{-\lambda \mu^{\leftarrow}(\Gamma_K)} \left( 1 - e^{-\lambda \mu^{\leftarrow}(\Gamma_K)} \right) \right] \\ &\leq 2\gamma \lambda \mathbb{E} \left[ \mu^{\leftarrow}(\Gamma_K)^2 (1 + \lambda \mu^{\leftarrow}(\Gamma_K)) e^{-\lambda \mu^{\leftarrow}(\Gamma_K)} \right]. \end{aligned}$$

Note that  $(1 + x)e^{-x} \leq 1$ , so

$$\begin{aligned} B_{K,3} &\leq 2\lambda \gamma \mathbb{E} \left[ \mu^{\leftarrow}(\Gamma_K)^2 \right] \\ &\leq 2 \frac{\gamma}{\lambda} \mathbb{E} \left[ e^{-(\gamma\lambda)^{-1} \Gamma_K} \right] = 2 \frac{\gamma}{\lambda} \left( \frac{\gamma\lambda}{1 + \gamma\lambda} \right)^{K-1}. \end{aligned}$$

Since  $B_{K,3}$  will dominate  $B_{K,1}$  and  $B_{K,2}$  asymptotically, there exists  $K_0 \in \mathbb{N}$  such that for  $K > K_0$ ,

$$B_K \leq 3B_{K,3} \leq 6 \frac{\gamma}{\lambda} \left( \frac{\gamma\lambda}{1 + \gamma\lambda} \right)^{K-1}.$$

**Sparse network** When  $\alpha > 0$ ,

$$\nu(d\theta) = \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \theta^{-\alpha-1} e^{-\lambda\theta} d\theta, \quad \mu(d\theta) = \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \theta^{-\alpha-1} d\theta,$$

and

$$\frac{d\nu}{d\mu} = e^{-\lambda\theta}, \quad \mu[x, \infty) = \gamma' x^{-\alpha}, \quad \mu^{\leftarrow}(u) = (\gamma' u^{-1})^{1/\alpha}, \quad \gamma' = \gamma \frac{\lambda^{1-\alpha}}{\alpha \Gamma(1-\alpha)}.$$

Similar to the example when  $\alpha = 0$ ,

$$\begin{aligned} B_{K,1} &= \int_{\mathbb{R}_+^2} x_1 x_2 F_K(\mu[\max\{x_1, x_2\}, \infty)) \nu(dx_1) \nu(dx_2) \\ &= 2 \int_{\mathbb{R}_+} x_1 \int_{x_1}^{\infty} F_K(\mu[x_2, \infty)) \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x_2^{-\alpha} e^{-\lambda x_2} dx_2 \nu(dx_1) \\ &\leq 2\gamma \int_{\mathbb{R}_+} x_1 F_K(\mu[x_1, \infty)) \int_{x_1}^{\infty} \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x_2^{-\alpha} e^{-\lambda x_2} dx_2 \nu(dx_1). \end{aligned}$$

Note that the integrand with respect to  $x_2$  is the density function of the gamma distribution with shape  $\alpha$  and rate  $\lambda$ , so the integral is less than 1. We partition the outer integral into two parts and bound them separately,

$$\begin{aligned} B_{K,1} &\leq 2\gamma \int_{\mathbb{R}_+} F_K(\mu[x, \infty)) \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x^{-\alpha} e^{-\lambda x} dx \\ &\leq 2\gamma^2 \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \left[ \int_0^a x^{-\alpha} dx + F_K(\mu[a, \infty)) \int_a^{\infty} x^{-\alpha} e^{\lambda x} dx \right] \\ &\leq 2\gamma^2 \frac{1}{\Gamma(1-\alpha)} \left[ \frac{\lambda^{1-\alpha}}{1-\alpha} a^{1-\alpha} + \Gamma(1-\alpha) \left( \frac{3\gamma' a^{-\alpha}}{K} \right)^K \right]. \end{aligned}$$

By setting the two terms in the brackets equal, we get

$$a = \left[ \frac{(1-\alpha)\Gamma(1-\alpha)}{\lambda^{1-\alpha}} \right]^{\frac{1}{(K-1)\alpha+1}} \left( \frac{3\gamma'}{K} \right)^{\frac{K}{(K-1)\alpha+1}} \sim \left( \frac{3\gamma'}{K} \right)^{\frac{1}{\alpha}}.$$

So

$$B_{K,1} \leq \frac{4\gamma^2\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} \left( \frac{3\gamma'}{K} \right)^{\frac{1-\alpha}{\alpha}}.$$

Similar to the last example where  $\alpha = 0$ , here

$$\begin{aligned} B_{K,2} &= 2(K-1) \int_{\mathbb{R}_+^2} x_1 x_2 \mathbb{E} \left[ \frac{1}{\Gamma_K} \mathbf{1}(\mu[x_2, \infty) \leq \Gamma_K \leq \mu[x_1, \infty)) \right] \nu(dx_1)\nu(dx_2) \\ &\leq 2 \int_{\mathbb{R}_+} x_1 F_{K-1}(\mu[x_1, \infty)) \int_{x_1}^{\infty} x_2 \nu(dx_2)\nu(dx_1) \\ &\leq 2\gamma \int_{\mathbb{R}_+} x_1 F_{K-1}(\mu[x_1, \infty)) \nu(dx_1). \end{aligned}$$

This has the same form as  $B_{K,1}$ , and therefore

$$B_{K,2} \leq \frac{4\gamma^2\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} \left( \frac{3\gamma'}{K-1} \right)^{\frac{1-\alpha}{\alpha}}.$$

Finally, since both  $e^{-\lambda x} < 1$  and  $e^{-\lambda\mu^{\leftarrow}(\Gamma_K)} < 1$ ,

$$\begin{aligned} B_{K,3} &= 2\mathbb{E} \left[ \int_{\mathbb{R}_+} \mu^{\leftarrow}(\Gamma_K) x e^{-\lambda\mu^{\leftarrow}(\Gamma_K)} \mathbf{1}[x \leq \mu^{\leftarrow}(\Gamma_K)] \gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} x^{-1-\alpha} e^{-\lambda x} dx \right] \\ &\leq 2\gamma \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha)} \mathbb{E} \left[ \mu^{\leftarrow}(\Gamma_K) \int_0^{\mu^{\leftarrow}(\Gamma_K)} x^{-\alpha} dx \right] \\ &\leq 2\gamma \frac{\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} \mathbb{E} [\mu^{\leftarrow}(\Gamma_K)^{2-\alpha}] \\ &= 2\gamma \frac{\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} (\gamma')^{\frac{2-\alpha}{\alpha}} \frac{\Gamma(K - \frac{2-\alpha}{\alpha})}{\Gamma(K)}. \end{aligned}$$

By Stirling's formula,

$$\frac{\Gamma(K - (2-\alpha)/\alpha)}{\Gamma(K)} \sim \frac{\sqrt{2\pi(K - \frac{2-\alpha}{\alpha})} \left( \frac{K - (2-\alpha)/\alpha}{e} \right)^{K - (2-\alpha)/\alpha}}{\sqrt{2\pi K} \left( \frac{K}{e} \right)^K} \sim K^{-\frac{2-\alpha}{\alpha}}.$$

So

$$B_{K,3} \leq 2\gamma \frac{\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} (\gamma')^{\frac{2-\alpha}{\alpha}} K^{-\frac{2-\alpha}{\alpha}}.$$

$B_{K,2}$  dominates  $B_{K,1}$  and  $B_{K,3}$  asymptotically, so there exists  $K_0 \in \mathbb{N}$  such that for  $K > K_0$ ,

$$B_K \leq \frac{12\gamma^2\lambda^{1-\alpha}}{(1-\alpha)\Gamma(1-\alpha)} \left( \frac{3\gamma'}{K-1} \right)^{\frac{1-\alpha}{\alpha}}.$$

**Appendix D: Truncated inference**

*Proof of Theorem 4.2.* Fix  $K \in \mathbb{N}$  and  $\epsilon > 0$ , and define the subset of state space  $A = \left\{ X_{N+1}^{K+} = 0, B(\Gamma_{1:K}, \sigma) \leq \frac{\epsilon}{N+1} \right\}$ . By assumption,

$$\hat{\Pi}(A) \geq 1 - \eta.$$

Further, by applying the bound from Theorem 4.1, we know that for states in  $A$ ,

$$1 - \epsilon \leq p(X_{1:N+1}^{K+} | \theta_{1:K}, \sigma) \leq 1.$$

Suppose  $p$  is the RHS of Eq. (14) that is proportional to the density of  $\Pi$ , and  $\hat{p}$  is the RHS of Eq. (14) removing the term  $p(X_{1:N}^{K+} | \theta_{1:K}, \sigma)$ , which is proportional to the density of  $\hat{\Pi}$ :

$$p \leq \hat{p}, \quad \text{and within } A, \quad p \geq (1 - \epsilon)\hat{p}.$$

Define the normalization constants  $Z, \hat{Z}$  such that  $\int p/Z = \int \hat{p}/\hat{Z} = 1$ . Then the above bounds yield

$$\hat{Z} = \int \hat{p} \geq \int p = \int_A p + \int_{A^c} p \geq \int_A (1 - \epsilon)\hat{p} = (1 - \epsilon) \frac{\int_A \hat{p}}{\int \hat{p}} \int \hat{p} \geq (1 - \epsilon)(1 - \eta)\hat{Z}.$$

Therefore,  $(1 - \epsilon)(1 - \eta)\hat{Z} \leq Z \leq \hat{Z}$ , and hence

$$\Pi(A) = \int_A p/Z \geq (1 - \epsilon) \int_A \hat{p}/Z \geq (1 - \epsilon) \int_A \hat{p}/\hat{Z} \geq (1 - \epsilon)(1 - \eta).$$

The above results yield the total variation bound via

$$\begin{aligned} \frac{1}{2} \int \left| \frac{p}{Z} - \frac{\hat{p}}{\hat{Z}} \right| &= \frac{1}{2} \int_A \left| \frac{p}{Z} - \frac{\hat{p}}{\hat{Z}} \right| + \frac{1}{2} \int_{A^c} \left| \frac{p}{Z} - \frac{\hat{p}}{\hat{Z}} \right| \\ &\leq \frac{1}{2} \int_A \left| \frac{p}{Z} - \frac{p}{\hat{Z}} \right| + \frac{1}{2} \int_A \left| \frac{p}{\hat{Z}} - \frac{\hat{p}}{\hat{Z}} \right| + \frac{1}{2} \left( \Pi(A^c) + \hat{\Pi}(A^c) \right) \\ &= \frac{1}{2} \left| \frac{1}{Z} - \frac{1}{\hat{Z}} \right| \int_A p + \frac{1}{2} \frac{1}{\hat{Z}} \int_A (\hat{p} - p) + \frac{1}{2} \left( \Pi(A^c) + \hat{\Pi}(A^c) \right) \\ &\leq \frac{1}{2} (1 - (1 - \epsilon)(1 - \eta)) + \frac{1}{2} \epsilon + \frac{1}{2} (1 - (1 - \eta)(1 - \epsilon) + \eta) \\ &= \frac{3(\epsilon + \eta)}{2} - \epsilon\eta. \end{aligned}$$

□

## References

- [1] Paul Erdős and Alfréd Rényi. On random graphs I. *Publicationes Mathematicae*, 6:290–297, 1959. [MR0120167](#)
- [2] Paul Holland, Kathryn Laskey, and Samuel Leinhardt. Stochastic block models: first steps. *Social Networks*, 5:109–137, 1983. [MR0718088](#)
- [3] Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems*, pages 4249–4257, 2016.
- [4] Sinead Williamson. Nonparametric network models for link prediction. *The Journal of Machine Learning Research*, 17(1):7102–7121, 2016. [MR3580355](#)
- [5] Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, 113(523):1311–1326, 2018. [MR3862359](#)
- [6] Trevor Campbell, Diana Cai, and Tamara Broderick. Exchangeable trait allocations. *Electronic Journal of Statistics*, 12(2):2290–2322, 2018. [MR3832093](#)
- [7] Svante Janson. On edge exchangeable random graphs. *Journal of Statistical Physics*, 173(3-4):448–484, 2018. [MR3876897](#)
- [8] François Caron and Emily Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366, 2017. [MR3731666](#)
- [9] Adrien Todeschini, Xenia Miscouridou, and François Caron. Exchangeable random measures for sparse and modular graphs with overlapping communities. *arXiv:1602.02114*, 2016. [MR4084173](#)
- [10] Tue Herlau, Mikkel Schmidt, and Morten Mørup. Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems*, pages 4260–4268, 2016.
- [11] Victor Veitch and Daniel Roy. The class of random graphs arising from exchangeable random measures. *arXiv:1512.03099*, 2015. [MR3781748](#)
- [12] Christian Borgs, Jennifer Chayes, Henry Cohn, and Nina Holden. Sparse exchangeable graphs and their limits via graphon processes. *The Journal of Machine Learning Research*, 18(1):7740–7810, 2017. [MR3827098](#)
- [13] François Caron and Judith Rousseau. On sparsity and power-law properties of graphs based on exchangeable point processes. *arXiv:1708.03120*, 2017.
- [14] Peter Orbanz and Daniel Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2014.
- [15] John Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967. [MR0210185](#)
- [16] Christian Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004. [MR2080278](#)
- [17] Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. *Bayesian data analysis*. CRC Press, 3rd edition, 2013. [MR3235677](#)
- [18] Michael Escobar and Mike West. Bayesian density estimation and inference

- using mixtures. *Journal of the American Statistical Association*, 90:577–588, 1995. [MR1340510](#)
- [19] Thomas Griffiths and Zoubin Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems*, 2005.
- [20] Tamara Broderick, Ashia Wilson, and Michael Jordan. Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli*, 24(4B):3181–3221, 2018. [MR3788171](#)
- [21] Yee Whye Teh, Dilan Görür, and Zoubin Ghahramani. Stick-breaking construction for the Indian buffet process. In *Artificial Intelligence and Statistics*, 2007.
- [22] Maria Kalli, Jim Griffin, and Stephen Walker. Slice sampling mixture models. *Statistics and Computing*, 21:93–105, 2011. [MR2746606](#)
- [23] Peiyuan Zhu, Alexandre Bouchard-Côté, and Trevor Campbell. Slice sampling for general completely random measures. In *Uncertainty in Artificial Intelligence*, 2020.
- [24] David Blei and Michael Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143, 2006. [MR2227367](#)
- [25] David Blei and John Lafferty. A correlated topic model of science. *The Annals of Applied Statistics*, 1(1):17–35, 2007. [MR2393839](#)
- [26] Chong Wang, John Paisley, and David Blei. Online variational inference for the hierarchical Dirichlet process. In *International Conference on Artificial Intelligence and Statistics*, 2011.
- [27] Finale Doshi, Kurt Miller, Jurgen Van Gael, and Yee Whye Teh. Variational inference for the Indian buffet process. In *Artificial Intelligence and Statistics*, pages 137–144, 2009.
- [28] Trevor Campbell, Jonathan Huggins, Jonathan How, and Tamara Broderick. Truncated random measures. *Bernoulli*, 25(2):1256–1288, 2019. [MR3920372](#)
- [29] Hemant Ishwaran and Lancelot James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001. [MR1952729](#)
- [30] Hemant Ishwaran and Mahmoud Zarepour. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002. [MR1926065](#)
- [31] Anirban Roychowdhury and Brian Kulis. Gamma Processes, Stick-Breaking, and Variational Inference. In *International Conference on Artificial Intelligence and Statistics*, 2015.
- [32] Diana Cai and Tamara Broderick. Completely random measures for modeling power laws in sparse graphs. In *NIPS 2015 Workshop on Networks in the Social and Informational Sciences*, 2015.
- [33] Harry Crane and Walter Dempsey. A framework for statistical network modeling. *arXiv:1509.08185*, 2015.
- [34] Jan Rosiński. Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Springer, 2001. [MR1833707](#)



- [35] Thomas Ferguson and Michael Klass. A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics*, 43(5):1634–1643, 1972. [MR0373022](#)
- [36] Robert Wolpert and Katja Ickstadt. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85(2):251–267, 1998. [MR1649114](#)
- [37] Yee Whye Teh and Dilan Gorur. Indian buffet processes with power-law behavior. In *Advances in Neural Information Processing Systems*, pages 1838–1846, 2009.
- [38] Yee Whye Teh and Michael Jordan. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*, 1:158–207, 2010. [MR2730663](#)
- [39] Jan Rosiński. On series representations of infinitely divisible random vectors. *The Annals of Probability*, pages 405–430, 1990. [MR1043955](#)
- [40] Günter Last and Mathew Penrose. *Lectures on the Poisson process*, volume 7. Cambridge University Press, 2017. [MR3791470](#)
- [41] Thomas Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973. [MR0350949](#)
- [42] Nils Lid Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294, 1990. [MR1062708](#)
- [43] Anders Brix. Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31(4):929–953, 1999. [MR1747450](#)
- [44] Charles Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2:1152–1174, 1974. [MR0365969](#)
- [45] Thomas Griffiths and Zoubin Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems*, 2006.
- [46] Pietro Panzarasa, Tore Opsahl, and Kathleen M Carley. Patterns and dynamics of users’ behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.
- [47] John Kingman. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, 1992. [MR1207584](#)