Statistics and Computing (2021) 31:63
https://doi.org/10.1007/s11222-021-10037-3

Nonnegative Bayesian nonparametric factor models with completely random measures

Fadhel Ayed[1] · **François Caron**[1]

Received: 8 July 2019 / Accepted: 26 July 2021 / Published online: 24 August 2021
© The Author(s) 2021

Abstract
We present a Bayesian nonparametric Poisson factorization model for modeling dense network data with an unknown and potentially growing number of overlapping communities. The construction is based on completely random measures and allows the number of communities to either increase with the number of nodes at a specified logarithmic or polynomial rate, or be bounded. We develop asymptotics for the number and size of the communities of the network and derive a Markov chain Monte Carlo algorithm for targeting the exact posterior distribution for this model. The usefulness of the approach is illustrated on various real networks.

Keywords Poisson factorization · Generalized gamma process · Community detection

1 Introduction

Nonnegative matrix factorization (NMF) methods (Paatero and Tapper 1994; Lee and Seung 2001) aim to find a latent representation of a positive $n \times m$ matrix A as a sum of K nonnegative factors. For integer-valued data, Poisson factorization models (Dunson and Herring 2005) offer a flexible probabilistic framework for nonnegative matrix factorization and have found wide applicability in signal processing (Virtanen et al. 2008; Cemgil 2009) or recommender systems (Ma et al. 2011; Gopalan et al. 2015). In this paper, we focus on the application to network analysis, where $m = n$ and the $n \times n$ count matrix A, the adjacency matrix, represents the number of directed or undirected interactions between n individuals; the latent factors may be interpreted as latent and potentially overlapping communities (Ball et al. 2011), such as sport team members or other social activities circles. We also consider binary data where the matrix represents the existence or absence of a directed or undirected link between individuals. The estimated latent factors can be used for the prediction of missing links/interactions, or for interpretation of the uncovered latent community structure.

Poisson factorization approaches require the user to set the number K of latent factors, which is typically assumed to be independent of the sample size n. To address this problem, Zhou et al. (2012), Gopalan et al. (2014) and Zhou (2015) proposed Bayesian nonparametric approaches that allow the number of latent factors to be estimated from the data, and to grow unboundedly with the size n of the matrix. In particular, Gopalan et al. (2014) and Zhou (2015) considered a Poisson factorization model

$$A_{ij} \sim \text{Poisson}\left(\sum_{k=1}^{+\infty} r_k v_{ik} v_{jk}\right), \quad 1 \le i, j \le n \tag{1}$$

where the positive weights $(r_k)_{k\ge1}$ represent the importance of community k, and $v_{ik} > 0$ represents the level of affiliation of individual i to community k. Gopalan et al. (2014) and Zhou (2015), extending work from Titsias (2008), assume that the weights (r_k) are the jumps of a gamma process, ensuring the sum in Eq. (1) is almost surely finite. Using properties of Poisson random variables, the model (1) can be equivalently represented as

$$A_{ij} = \sum_{k=1}^{+\infty} Z_{ijk} \tag{2}$$

$$Z_{ijk} \sim \text{Poisson}(r_k v_{ik} v_{jk}), \quad k = 1, 2, \ldots \tag{3}$$

✉ Fadhel Ayed
fadhel.ayed@gmail.com

François Caron
caron@stats.ox.ac.uk

[1] Department of Statistics, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK

for $1 \leq i, j \leq n$. The latent count variables Z_{ijk} may be interpreted as the number of latent interactions between two individuals i and j via community k, the overall number A_{ij} of interactions being the sum of those community interactions. For example, two members of the same company who also play sport together may meet five times at the company, and twice at the sport center, resulting in seven interactions overall. The overall number

$$K_n = \sum_{k=1}^{+\infty} \mathbb{1}_{\sum_{1\leq i,j\leq n} Z_{ijk}>0} \tag{4}$$

of communities k that generated at least one interaction between the n individuals is termed the number of *active communities*. For the gamma process Poisson factor model (Zhou 2015), the number of active communities K_n grows logarithmically with the number n of individuals. The logarithmic growth assumption may be too restrictive. For example, the number of active communities may actually be unknown but bounded above; alternatively, it may increase at a rate faster or slower than logarithmic.

In this paper, we consider generalizations of the gamma process Poisson factorization model, using completely random measures (CRM) (Kingman 1967). CRMs offer a flexible and tractable modeling framework (Lijoi and Prünster 2010). The proposed models fit in the class of multivariate generalized Indian Buffet process priors recently developed by James (2017) and are also related to compound completely random measures (Griffin and Leisen 2017). We consider that (r_k) are the points of Poisson point process with mean measure ρ. Depending on the properties of this measure, the number of active communities K_n is either (i) bounded, with a random upper bound, (ii) unbounded and grows subpolynomially (e.g., $\log n$ or $\log\log n$) or (iii) unbounded and grows as $n^{2\sigma}$, for some $\sigma \in (0, 1)$. For the implementation, we focus in particular on the generalized gamma process (Brix 1999) where a single parameter flexibly controls all three behaviors.

The article is organized as follows. In Sect. 2, we describe the statistical model for count and binary matrices. The asymptotic properties of the model are derived in Sect. 3. In particular, we relate the asymptotic growth of the number of active features to the regular variation properties of the measure ρ. In Sect. 4, we derive a Markov chain Monte Carlo algorithm for posterior inference that does not require any approximation to the original model. In Sect. 5, we consider applications of our approach to overlapping community detection and link detection in networks, considering real network data with up to tens of thousands of nodes.

2 Statistical model for count and binary data

2.1 General construction

We present here the model for directed count or binary observations, but the model can be straightforwardly adapted to undirected interactions. Let $(r_k)_{k=1,2\ldots,}$ be the points of a Poisson point process with σ-finite mean measure ρ on $(0, +\infty)$, and assume that v_{ik}, $i = 1, \ldots, n$, $k \geq 1$, are independent and identically distributed from some probability distribution F on $\mathbb{R}_+ = [0, +\infty)$. The variable v_{ik} can be interpreted as the level of affiliation of an individual i to community k, and r_k to the importance of that community.

For count data (A_{ij}), where A_{ij} denotes the number of directed interactions from node i to node j, we consider the Poisson factorization model

$$A_{ij} \mid (r_k, v_{ik}) \sim \text{Poisson}\left(\sum_{k=1}^{+\infty} r_k v_{ik} v_{jk}\right), \ 1 \leq i, j \leq n. \tag{5}$$

Denoting $\Lambda_{ij} = \sum_{k=1}^{+\infty} r_k v_{ik} v_{jk}$ the Poisson rate for A_{ij}, the $n \times n$ rate matrix $\Lambda^{(n)} = (\Lambda_{ij})_{1\leq i,j\leq n}$ admits the following factorization as an infinite sum of rank-1 matrices

$$\Lambda^{(n)} = \sum_{k=1}^{+\infty} r_k v_{1:n,k} v_{1:n,k}^{\intercal}$$

where $v_{1:n,k} = (v_{1k}, \ldots, v_{nk})^{\intercal}$. For the model to be well specified, the sum in the right-hand side of Eq. (5) needs to be almost surely finite. A necessary and sufficient condition is

$$\iint (1 - e^{-rv^2})\rho(dr)F(dv) < +\infty \quad \text{and}$$
$$\iint (1 - e^{-rv_1v_2})\rho(dr)F(dv_1)F(dv_2) < +\infty. \tag{6}$$

A sufficient set of conditions[1], which we will assume to hold in the rest of this article, is that ρ is a Lévy measure and F has finite second moment, that is

$$\int_0^{+\infty} (1 - e^{-r})\rho(dr) < +\infty \quad \text{and} \tag{A1}$$

$$\int_0^{+\infty} v^2 F(dv) < +\infty. \tag{A2}$$

[1] The sufficientness follows from the bound (34) given in Appendix.

In this case, denoting δ_v the Dirac measure at vector v, the community affiliations and weights for n nodes can be conveniently represented by a completely random measure

$$G = \sum_{k \geq 1} r_k \delta_{v_{1:n,k}} \tag{7}$$

on $\mathbb{R}_+^n$ with mean measure $\rho(dr) F^{\otimes^n}(dv_1, \ldots, dv_n)$ where $F^{\otimes^n}$ denotes the nth product measure of F; see Kingman (1967) and Lijoi and Prünster (2010) for background on CRMs and their applications. If the Lévy measure is finite, that is, if

$$\int_0^{+\infty} \rho(dr) < +\infty$$

then the number of points (r_k), and therefore the number of communities, is almost surely finite. Otherwise, when $\int \rho(dr) = +\infty$, the number of communities is infinite.

When we have binary observation (Y_{ij}), we treat the count matrix (A_{ij}) as a latent variable, and consider that $Y_{ij} = \mathbb{1}_{A_{ij}>0}$ as in (Caron and Fox 2017; Zhou 2015), where $\mathbb{1}$ is the indicator function. Integrating out (A_{ij}), this leads to the following model for binary observations

$$Y_{ij} \mid (r_k, v_{ik}) \sim \text{Ber}\left(1 - \exp\left[\sum_{k=1}^{+\infty} r_k v_{ik} v_{jk}\right]\right), \quad 1 \leq i, j \leq n. \tag{8}$$

2.2 Specific model

In the inference and experimental part, we use the following choice for the ρ and F. The Lévy measure ρ is taken to be that of a generalized gamma process (GGP, see Hougaard (1986), Brix (1999), James (2002), Pitman (2003))

$$\rho(dr) = \frac{\kappa}{\Gamma(1-\sigma_0)} r^{-1-\sigma} e^{-\tau r} dr \tag{9}$$

where $\sigma_0 \in (-\infty, 1)$, $\kappa > 0$ and $\tau > 0$. When $\sigma_0 = 0$, we obtain a gamma process, and the model corresponds to that of Zhou (2015). When $\sigma_0 < 0$, the Lévy measure is finite, while when $\sigma_0 \geq 0$, the Lévy measure is infinite.

Concerning the affiliations, we will assume that F is a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$. That is, the probability density function (pdf) f is given by

$$f(v) = \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v}$$

where Γ denotes the usual gamma function. The hyperparameters $(\kappa, \sigma_0, \tau, \alpha, \beta)$ and $(\kappa' = \kappa/\beta^{2\sigma_0}, \sigma_0, \tau' = \tau\beta^2, \alpha, 1)$ induce the same distribution for the latent factors (Λ_{ij}). In order to guarantee the identifiability of the hyperparameters, we therefore set $\beta = 1$.

2.3 Related work

Several network models building on latent factors have been proposed in the last years and have proven to be very useful tools (Hoff et al. 2002; Airoldi et al. 2008; Hoff 2009; Durante and Dunson 2014). In general, these models differ from Poisson factor models since they use a different likelihood for the connections. However, they share a similar approach: every node i is embedded in $\mathbb{R}_+^K$ (where K is the number of latent factors or communities), resulting in a latent representation X_i quantifying the affiliation of node i to each latent factor. Then, the probability of an edge (i, j) is function of the similarity between X_i and X_j.

The model introduced in this section can be seen from different perspectives that nicely connect it to the existing literature. First, the model can be seen as obtained from a functional of a CRM. Recall the definition of the CRM G in Eq. (7). Define the $n \times n$ matrix $\Lambda^{(n)}$ as the following functional of G

$$\Lambda^{(n)} = \int_{(0,+\infty)^n} h(u) G(du) = \sum_{k \geq 1} r_k v_{1:n,k} v_{1:n,k}^\intercal$$

where $h(u) = uu^\intercal$. Alternatively, this can be interpreted in the framework of compound completely random measures (Griffin and Leisen 2017). For each $1 \leq i, j \leq n$, denote $G_{ij} = \sum_{k \geq 1} r_k v_{ik} v_{jk} \delta_{\zeta_k}$ where ζ_k are some community locations in some domain Θ, iid from some distribution H, irrelevant here. Then, $(G_{ij})_{1 \leq i,j \leq n}$ are compound CRMs on Θ and $\Lambda_{ij} = G_{ij}(\Theta)$. In the same vein, the model can also be interpreted as an instance of the class of Generalized Indian Buffet Processes introduced by James (2017), where the Bernoulli likelihood (of the classical IBP) is replaced by any likelihood as long as the observation can take the value 0 with strictly positive probability. More precisely, if we denote $Z_k^{(n)}$ the $n \times n$ matrix with entries Z_{ijk}, then the matrix-valued process $\sum_{k \geq 1} Z_k^{(n)} \delta_{\zeta_k}$ is a draw from a generalized multivariate Indian buffet process.

Finally, as mentioned in the introduction, the model admits as a special case the Poisson factorization based on the gamma process of Zhou (2015).

3 Asymptotic Properties

In this section, we study the asymptotic properties of the proposed class of models, and in particular the growth rate of the number of active communities as the sample size n grows, and the asymptotic proportion of communities of a given size.

For a given sequence $(r_k)_{k\geq 1}$ and $(v_{ik})_{i\geq 1,k\geq 1}$, denote $A_{ij}^{(n)}$ and $Z_{ijk}^{(n)}$ where $n \geq 1$, $1 \leq i, j \leq n$, $k \geq 1$, respectively, the number of directed interactions and the number of community directed interactions distributed from Eqs. (2) and (3). We consider two different asymptotic settings

- Unconstrained setting. This setting is more general, and we only assume that $A_{ij}^{(n)}$ and $Z_{ijk}^{(n)}$ are marginally sampled from Eqs. (2) and (3).
- Constrained setting. For any $1 \leq m \leq n$, and $1 \leq i, j \leq m$, $A_{ij}^{(n)} = A_{ij}^{(m)}$. In this setting, we suppose that the connections between the already observed nodes remain unchanged. It is equivalent to assuming that there is an infinite but fixed graph and $A^{(n)}$ represents the connections between the n first nodes of that graph.

All the results of this section, otherwise stated, hold for the unconstrained setting. We indicate when a stronger result holds in the constrained setting. All proofs are given in Appendix A.

3.1 General model

Let $d_k^{(n)}$ be the degree of the community/feature k, corresponding to the number of interactions amongst n individuals due to community k, and defined as

$$d_k^{(n)} = \sum_{1\leq i,j\leq n} Z_{ijk}^{(n)}. \tag{10}$$

A community is active if $d_k^{(n)} \geq 1$. The number of active communities is therefore defined as

$$K_n = \sum_{k=1}^{+\infty} \mathbb{1}_{d_k^{(n)}\geq 1} \tag{11}$$

Denote $K_{n,j}$ the number of communities with degree $j \geq 1$

$$K_{n,j} = \sum_{k=1}^{+\infty} \mathbb{1}_{d_k^{(n)}=j}$$

Note that under the constrained setting, $d_k^{(n)}$, K_n and $\sum_{\ell\geq j} K_{n,\ell}$ are all almost surely non-decreasing with the sample size n, whereas this is not necessarily the case for the unconstrained setting.

Proposition 1 *Under Assumptions (A1) and (A2), the number of active communities K_n is a Poisson random variable with mean*

$$\Psi(n) = \int \left(1 - e^{-r(\sum_{i=1}^n v_i)^2}\right) \left[\prod_{i=1}^n F(dv_i)\right] \rho(dr) < +\infty. \tag{12}$$

The number $K_{n,j}$ of communities with degree j is also Poisson distributed, with mean

$$\Psi_j(n) = \frac{1}{j!} \int r^j \left(\sum_{i=1}^n v_i\right)^{2j} e^{-r(\sum_{i=1}^n v_i)^2} \left[\prod_{i=1}^n F(dv_i)\right] \rho(dr). \tag{13}$$

Finally, for $j \geq 1$, $\sum_{k\geq j} K_{n,k}$, the number of communities with degree at least j is also Poisson distributed with mean $\sum_{k\geq j} \Psi_k(n)$.

In the rest of the section, we relate the asymptotic behavior of quantities of interest to the properties of the mean measure ρ. Let consider the tail Lévy intensity defined as

$$\forall x > 0,\ \overline{\rho}(x) = \int_x^{+\infty} \rho(dr).$$

We assume that $\overline{\rho}$ is a regularly varying function at 0, that is

$$\overline{\rho}(x) \asymp x^{-\sigma}\ell(1/x) \text{ as } x \to 0 \tag{A4}$$

where $\sigma \in [0, 1)$ and ℓ is a slowly varying function verifying $\lim_{t\to+\infty} \ell(at)/\ell(t) = 1$ for all $a > 0$. Besides, we write $a(x) \asymp b(x)$ if $\lim a(x)/b(x) = 1$. Examples of slowly varying functions include functions converging to a constant, $\log^a t$ for any t, $\log\log t$, etc. Note that the CRM is finite activity if and only if $\sigma = 0$ and $\ell(t) \to C < +\infty$.

Now, let us consider the asymptotic behavior of the number of active communities K_n.

Proposition 2 *Let K_n be the number of active communities. Then for $0 \leq \sigma < 1$,*

$$\mathbb{E}[K_n] \asymp \Gamma(1-\sigma) m_f^{2\sigma} n^{2\sigma} \ell(n^2) \tag{14}$$

as n tends to infinity, where $m_f = \int vF(dv)$. Additionally, for $0 < \sigma < 1$,

$$K_n \asymp \mathbb{E}[K_n] \ \ a.s. \tag{15}$$

If we further assume that the sequence $(K_n)_{n\geq 1}$ is almost surely non-decreasing (as in the constrained setting), then (15) holds for $\sigma = 0$ and $\ell(t) \to +\infty$ as well. In the finite activity case, that is $\sigma = 0$ and $\ell(t) \to \overline{\rho}(0) = \int_0^{+\infty} \rho(dr) < \infty$, we have

$$K_n \to K_\infty$$

as n tends to infinity, where K_∞ is a Poisson random variable with mean $\overline{\rho}(0)$. The above convergence holds in distribution for the unconstrained setting and almost surely for the constrained setting.

Proposition 3 *Let $K_{n,j}$ be the number of communities of degree j. Then for $0 < \sigma < 1$ and any $j \geq 1$,*

$$K_{n,j} \asymp \frac{\sigma \Gamma(j-\sigma)}{j!} m_f^{2\sigma} n^{2\sigma} \ell(n^2) \quad a.s. \tag{16}$$

as n tends to infinity. Therefore,

$$\frac{K_{n,j}}{K_n} \to \frac{\sigma \Gamma(j-\sigma)}{\Gamma(1-\sigma) j!} \quad a.s. \tag{17}$$

as n tends to infinity. This corresponds to a power-law behavior as

$$\frac{\sigma \Gamma(j-\sigma)}{\Gamma(1-\sigma) j!} \asymp \frac{\sigma}{j^{\sigma+1}}$$

for large j. If we further assume that for all $k \geq 1$, $\left(\sum_{j \geq k} K_{n,j}\right)_{n \geq 1}$ is non-decreasing (constrained setting), then (17) holds also for $\sigma = 0$ and $\ell(t) \to +\infty$.

It has been observed empirically that in many networks, the distribution of the sizes of the communities displays a power-law behavior $f_S(s) \sim s^{-1-\sigma}$ where f_S is the distribution of community sizes and $\sigma > 0$ [(see for example (Stegehuis et al. 2016; Radicchi et al. 2004; Clauset et al. 2004; Arenas et al. 2004)]. As stated in Proposition 17, this property cannot be captured in the framework of Zhou (2015) for example where $\sigma = 0$ is constant. These empirical observations seem to indicate that models with flexible σ are needed.

Finally, let $c^{(n)}(k, k')$ denote the cosine between the corresponding affiliation vectors

$$c^{(n)}(k, k') = \frac{\sum_{i=1}^n v_{ik} v_{ik'}}{\sqrt{\sum_i v_{ik}^2} \sqrt{\sum_i v_{ik'}^2}}.$$

This coefficient gives a measure of the overlap between two communities k and k'. By the law of large numbers, for any $k \neq k'$,

$$c^{(n)}(k, k') \asymp \frac{(\int v F(dv))^2}{\int v^2 F(dv)} \quad a.s. \text{ as } n \to +\infty.$$

3.2 Specific case of the GGP

In the case of the GGP, we have

$$\overline{\rho}(x) = \frac{\kappa \tau^{\sigma_0} \Gamma(-\sigma_0, \tau x)}{\Gamma(1-\sigma_0)} \asymp \begin{cases} -\frac{\kappa \tau^{\sigma_0}}{\sigma_0} & \text{if } \sigma_0 < 0 \\ \kappa \log(1/x) & \text{if } \sigma_0 = 0 \\ \frac{\kappa x^{-\sigma_0}}{\sigma_0 \Gamma(1-\sigma_0)} & \text{if } \sigma_0 > 0 \end{cases}$$

as x tends to 0, where $\Gamma(a, x)$ is the incomplete gamma function. Note that $\overline{\rho}(x)$ is of the form $x^{-\sigma} \ell(1/x)$ where $\sigma = \max(0, \sigma_0)$ and

$$\ell(t) = \begin{cases} -\frac{\kappa \tau^{\sigma_0}}{\sigma_0} & \text{if } \sigma_0 < 0 \\ \kappa \log(t) & \text{if } \sigma_0 = 0 \\ \frac{\kappa}{\sigma_0 \Gamma(1-\sigma_0)} & \text{if } \sigma_0 > 0 \end{cases}$$

is a slowly varying function at infinity. The results of the previous subsection therefore apply. For simplicity, we state the results for the constrained setting. We have, almost surely as $n \to +\infty$

$$K_n \asymp \begin{cases} K_\infty & \text{if } \sigma_0 < 0 \\ 2\kappa \log(n) & \text{if } \sigma_0 = 0 \\ \kappa \alpha^{2\sigma_0} n^{2\sigma_0} / \sigma_0 & \text{if } \sigma_0 > 0 \end{cases}$$

where $K_\infty \sim \text{Poisson}(-\kappa \tau^{\sigma_0} / \sigma_0)$. Additionally, for $\sigma \geq 0$,

$$\frac{K_{n,j}}{K_n} \to \frac{\sigma_0 \Gamma(j-\sigma_0)}{\Gamma(1-\sigma_0) j!}$$

almost surely as $n \to +\infty$. Finally,

$$c^{(n)}(k, k') \to \frac{\alpha}{\alpha+1}.$$

Therefore, σ_0 governs the asymptotic behavior of the number of active communities. K_n is bounded with a random upper bound ($\sigma_0 < 0$), increases logarithmically ($\sigma_0 = 0$) or polynomially ($\sigma_0 > 0$). In the polynomial case, σ_0 also controls the power-law exponent of the proportion of communities of a given size. The parameter κ is an overall linear scaling parameter. Finally, the parameter α governs the amount of overlapping between two communities.

4 Simulation, posterior characterization and inference

In this section, we describe the marginal distribution and conditional characterization of the model. Building on these, we derive an exact sampler for simulating from the model, and a Markov chain Monte Carlo algorithm to approximate the posterior distribution. Importantly, the sampler targets the distribution of interest and does not require any truncation

or approximation. For simplicity of exposition, we assume that ρ and F are absolutely continuous with respect to the Lebesgue measure, with $\rho(dr) = \rho(r)dr$ and $F(dx) = f(x)dx$.

4.1 Marginal distribution and simulation

For a fixed n, recall that K_n denotes the number of active communities. Let $((\widetilde{r}_1, \widetilde{v}_{1:n,1}), \ldots, (\widetilde{r}_{K_n}, \widetilde{v}_{1:n,K_n}))$ be the subsequence of $(r_k, v_{1:n,k})$ such that community k is active, meaning that $\sum_{1\le i,j\le n} Z_{ijk} \ge 1$, arranged in random order. Let $\widetilde{Z}_{ijk}$ be the number of community interactions corresponding to the active community $(\widetilde{r}_k, \widetilde{v}_{1:n,k})$. Note that

$$A_{ij} = \sum_{k=1}^{K_n} \widetilde{Z}_{ijk}. \tag{18}$$

Let $\widetilde{Z}_k = (\widetilde{Z}_{ijk})_{1\le i,j\le n}$. Using Proposition 5.2 of James (2017), we obtain the following lemma.

Lemma 1 (Marginal distribution) *The joint distribution of* $(K_n, (\widetilde{r}_{1:K_n}, \widetilde{v}_{1:n,1:K_n}), (\widetilde{Z}_k)_{k=1,\ldots,K_n})$ *is given by*

$$K_n \sim \text{Poisson}(\Psi(n)) \tag{19}$$

where $\Psi(n)$ *is defined in Eq.(12), and*

$$p((\widetilde{r}_{1:K_n}, \widetilde{v}_{1:n,1:K_n})|K_n) = \prod_{k=1}^{K_n} p(\widetilde{r}_k, \widetilde{v}_{1:n,k}|K_n)$$

where

$$p(\widetilde{r}_k, \widetilde{v}_{1:n,k}|K_n) \propto (1 - e^{-\widetilde{r}_k(\sum_{i=1}^n \widetilde{v}_{ik})^2})\rho(\widetilde{r}_k)\prod_{i=1}^n f(\widetilde{v}_{ik}). \tag{20}$$

Finally, for each $k = 1, \ldots, K_n$,

$$\widetilde{Z}_k|(\widetilde{r}_{1:K_n}, \widetilde{v}_{1:n,1:K_n}) \sim \text{tPoisson}(\widetilde{r}_k\widetilde{v}_{1:n,k}\widetilde{v}_{1:n,k}^\intercal) \tag{21}$$

where tPoisson(Λ) *denotes the distribution of a integer-valued matrix with Poisson entries with mean values* Λ_{ij}, *conditionally on the sum of the entries being strictly positive. This has probability mass function*

$$p(A) = \begin{cases} (1 - e^{-\sum_{ij}\Lambda_{ij}})^{-1}\prod_{1\le i,j\le n}\frac{\Lambda_{ij}^{A_{ij}}e^{-\Lambda_{ij}}}{A_{ij}!} & \text{if } \sum_{ij} A_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

The model has an infinite number of parameters, but Lemma 1 allows us to derive an algorithm to exactly sample from it, by successively simulating K_n, $(\widetilde{r}_{1:K_n}, \widetilde{v}_{1:n,1:K_n})$, $(\widetilde{Z}_k)_{k=1,\ldots,K_n}$ and A using Eqs. (19), (20), (21) and (18).

Sampling from the conditional distribution (21) can be done efficiently by first sampling the number of multi-edges $\sum_{i,j}\widetilde{Z}_{i,j,k}$ from a truncated Poisson with mean $\widetilde{r}_k(\sum_i \widetilde{v}_{i,k})^2$, then sampling iid the end nodes of the edges proportionally to the affiliation vector. Simulating from the conditional distribution (20) can be more challenging since it requires sampling a $n+1$-dimensional vector. However, if we suppose that the affiliations are Gamma distributed, the problem reduces to sampling $(\widetilde{r}_k, \sum_i \widetilde{v}_{i,k})$, which is a two-dimensional vector, and independently sample the normalized affiliations from a Dirichlet distribution. Indeed, if the affiliations are Gamma distributed, we consider the following change of variable.

$$\widetilde{\varsigma}_k = \sum_{i=1}^n \widetilde{v}_{ik}, \quad k = 1, \ldots, K_n \tag{22}$$

$$\widetilde{\varphi}_{ik} = \frac{\widetilde{v}_{ik}}{\widetilde{\varsigma}_k}, \quad k = 1, \ldots, K_n;\ i = 1, \ldots, n \tag{23}$$

This gives the following algorithm for exact simulation from the model.

1. Sample K_n from Eq. (19)
2. For $k = 1, \ldots, K_n$
   (a) Sample $(\widetilde{\varphi}_{1k}, \ldots, \widetilde{\varphi}_{nk}) \sim \text{Dirichlet}(\alpha, \ldots, \alpha)$
   (b) Sample $\widetilde{\varsigma}_k$ from
   $$p(\widetilde{\varsigma}) \propto \psi(\widetilde{\varsigma}^2)\,\text{Gamma}(\widetilde{\varsigma}; n\alpha, \beta) \tag{24}$$
   (c) Sample $\widetilde{r}_k|\widetilde{\varsigma}_k$ from
   $$p(\widetilde{r} \mid \widetilde{\varsigma}) \propto (1 - e^{-\widetilde{r}\widetilde{\varsigma}^2})\rho(\widetilde{r}) \tag{25}$$
   (d) Sample $\widetilde{Z}_k^{(n)}$ from Eq. (21)
3. For $1 \le i, j \le n$, set $A_{ij} = \sum_{k=1}^{K_n}\widetilde{Z}_{ijk}$

where $\psi(t) = \int_0^{+\infty}(1 - e^{-wt})\rho(dw)$ is the Laplace exponent, Dirichlet$(\alpha, \ldots, \alpha)$ denotes the standard Dirichlet distribution and Gamma$(x; a, b)$ denotes the probability density function of a Gamma random variable with parameters a and b, evaluated at x. In the case of the GGP, the Laplace exponent is

$$\psi(t) = \frac{\kappa}{\sigma}((t+\tau)^\sigma - \tau^\sigma). \tag{26}$$

One can sample from Eqs. (24) and (25) using rejection.

4.2 Posterior characterization

Using Proposition 5.1 in (James 2017), one can characterize the conditional distribution of the CRM G given the latent community counts $\widetilde{Z}_{ijk}$.

Lemma 2 *Conditionally on* $(\widetilde{Z}_k^{(n)})_{k=1,\ldots,K_n}$, *the CRM G has the same distribution as*

$$G' + \sum_{k=1}^{K_n} \tilde{r}_k \delta_{\tilde{v}_{1:n,k}}$$

where G' *is an inhomogeneous CRM on* $\mathbb{R}_+^n$ *with mean intensity*

$$e^{-r(\sum_{i=1}^{n} v_i)^2} \rho(r) \prod_{i=1}^{n} f(v_i)$$

and $(\tilde{r}_k, \tilde{v}_{1:n,k})_{k=1,\ldots,K_n}$ *are independent of* G' *and iid with density*

$$p(\tilde{r}_k, \tilde{v}_{1:n,k} | \widetilde{Z}_k^{(n)}) = e^{-\tilde{r}_k (\sum_i \tilde{v}_{ik})^2} \tilde{r}_k^{\tilde{d}_k} \rho(\tilde{r}_k) \prod_{i=1}^{n} \tilde{v}_{ik}^{\tilde{m}_{ik}} f(\tilde{v}_{ik}) \tag{27}$$

where $\widetilde{m}_{ik} = \sum_j \widetilde{Z}_{ijk} + \widetilde{Z}_{jik}$ *and* $\widetilde{d}_k = \sum_{i,j} \widetilde{Z}_{ijk}$. .

In the case where f is a gamma pdf, we can use the same reparameterization as in the previous subsection with $(\widetilde{\varsigma}_k, \widetilde{\varphi}_{1:n,k})$ in place of $\widetilde{v}_{1:n,k}$. This leads to the following conditional distributions.

$$\widetilde{\phi}_{1:n,k} | \widetilde{Z}_k^{(n)} \sim \text{Dirichlet}(\alpha + \widetilde{m}_{1k}, \ldots, \alpha + \widetilde{m}_{nk})$$
$$p(\widetilde{\varsigma}_k | \widetilde{Z}_k^{(n)}) \propto \varkappa(\widetilde{d}_k, \widetilde{\varsigma}^2)\, \text{Gamma}(\widetilde{\varsigma}; n\alpha + 2\widetilde{d}_k, \beta)$$
$$p(\widetilde{r}_k | \widetilde{\varsigma}_k, \widetilde{Z}_k^{(n)}) \propto e^{-\widetilde{r}_k \widetilde{\varsigma}_k^2} \widetilde{r}_k^{\widetilde{d}_k} \rho(\widetilde{r}_k)$$

where $\varkappa(m, t) = \int_0^{+\infty} r^m e^{-rt} \rho(r) dr$. In the GGP case, we have

$$\varkappa(m, t) = \kappa \frac{\Gamma(m - \sigma)}{\Gamma(1 - \sigma)} (t + \tau)^{\sigma - m}$$

and

$$\widetilde{r}_k | \widetilde{\varsigma}_k, \widetilde{Z}_k^{(n)} \sim \text{Gamma}(\widetilde{d}_k - \sigma, \widetilde{\varsigma}_k^2 + \tau).$$

4.3 Slice sampler for posterior inference

We recall that θ denote the set of hyperparameters of the mean measure ρ and pdf f. To simplify the presentation, here we suppose that we observe the complete adjacency matrix A, which means that we observe a directed and weighted graph with no missing (hidden) edge. The objective is to obtain samples distributed from the conditional distribution

$$p(K_n, (\widetilde{r}_k, \widetilde{v}_{1:n,k})_{k=1,\ldots,K_n}, \theta \mid A).$$

In the Appendix, we show how to do inference when we only observe a partial graph (with missing edges to predict) that can be directed or undirected, weighted or binary. In order to leverage the Poisson factorization construction, we augment the model with the latent community counts $\widetilde{Z}_k$. Additionally, to deal with the unknown number of active communities K_n, we use auxiliary slice variables, similarly to other Gibbs sampler for Bayesian nonparametric models (Walker 2007; Kalli et al. 2011; Favaro and Teh 2013). For each directed pair (i, j) such that $A_{ij} \geq 1$ consider the scalar latent variable

$$s_{ij} | (\widetilde{r}_k, \widetilde{Z}_{ijk})_{k=1,\ldots,K_n} \sim \text{Unif}\left(0, \min_{\{k | \widetilde{Z}_{ijk} \geq 1\}} \widetilde{r}_k\right) \tag{28}$$

and denote $s = \min_{ij} s_{ij}$. Note that by definition, $\widetilde{r}_k \geq s$ for all $k = 1, \ldots, K_n$. In the following, we will consider the communities k, active or inactive, which r is higher than s. We use the notation $\overline{\cdot}$ to denote the corresponding variables. More precisely, let

$$\overline{G} = \sum_k r_k \delta_{v_{1:n,k}} \mathbb{1}_{r_k \geq s} := \sum_{k=1}^{\overline{K}_n} \overline{r}_k \delta_{\overline{v}_{1:n,k}}$$

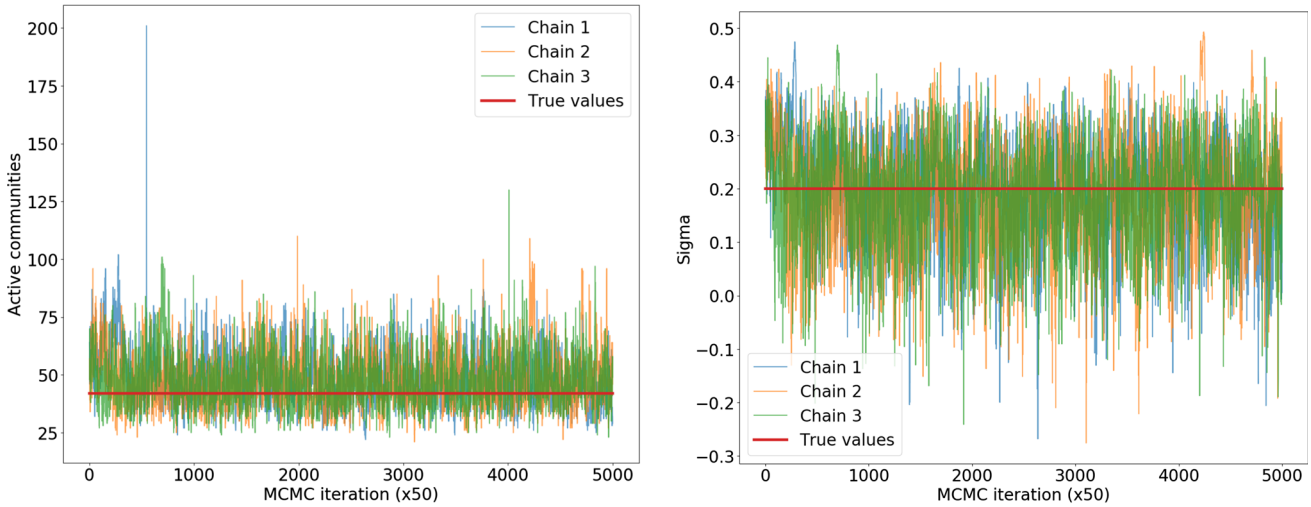
be the CRM corresponding to the set of active or inactive communities with weight $r_k \geq s$, of (almost surely finite) cardinality $\overline{K}_n \geq K_n$. Denote $\overline{Z}_{ijk} \geq 0$ the associated community interactions, and $\overline{Z}_k = (\overline{Z}_{ijk})$. The data augmented slice sampler draws samples asymptotically distributed from

$$p((\overline{Z}_k)_{k=1,\ldots,\overline{K}_n}, \overline{G}, \theta, s \mid A).$$

The main steps of the algorithm are as follows.

1. For each directed pair (i, j) such that $A_{ij} \geq 1$, Update $(\overline{Z}_{ijk})_{k=1,\ldots,\overline{K}_n}$ given the rest of the variables,
2. Update the hyperparameters θ given the rest of the variables,
3. Update $(\overline{G}, s)$ given the rest of the variables.

The details of each step are given in Appendix B. Each iteration of the Gibbs sampler has a time complexity scaling in $\overline{K}_n S$ where S is the number of nonzero entries of the matrix. Therefore, the algorithm takes advantage of the sparsity of the networks. Additionally, each entry of the sparse graph can be dealt with independently, making the algorithm straightforwardly parallelizable.



(a) Trace plot of K_n **(b)** Trace plot of σ

Fig. 1 Trace plots of **a** the number of active communities K_n and **b** σ, on a synthetic example

5 Experiments

We implement the algorithm described in the previous section with the GGP-Gamma scores model. We assign Gamma priors on the hyperparameters κ, τ, α with parameters $(0.1, 0.1)$. We fix $\beta = 1$. We allow up to a linear growth of the number of communities, corresponding to $\sigma < 0.5$, for small datasets and use a Gamma prior with parameter $(0.1, 0.1)$ on $1-2\sigma$. For larger datasets, we restrict $\sigma < 0.25$, meaning that the number of communities cannot grow at a faster rate than $\sqrt{n}$. This is obtained by using a Gamma prior with parameter $(0.1, 0.1)$ on $1 - 4\sigma$.

The model allows overlapping communities but, for visualization purposes for example, it is useful to obtain an associated partition of the nodes. For each iteration, one can cluster the nodes by assigning each node to the community where it is most active. That is, at iteration t of the MCMC algorithm, define for $i = 1, \ldots, n$

$$c_i^{(t)} = \text{argmax}_k \{\sqrt{r_k^{(t)}} v_{ik}^{(t)}\} \tag{29}$$

the cluster membership of node i. We then compute an approximate Bayesian point estimate $\widehat{c} = (\widehat{c}_1, \ldots, \widehat{c}_n)$ of the partition of the nodes, using Binder's loss function (Lau and Green 2007).

5.1 Synthetic datasets

Data generated from the GGP-gamma model. We first run the algorithm on a synthetic dataset simulated from our model, to check that the algorithm can recover the true parameters. We sample a directed and unweighted graph from the GGP-gamma model with size $n = 800$ and $\sigma = 0.2, \kappa = 1, \tau = 0.15, \alpha = 0.05, \beta = 0.2$. The number of edges of the obtained graph is 20198, and the true number of active communities is 42. We run three chains in parallel with 500, 000 iterations, with 250, 000 iterations for burn-in. We show in Fig. 1 trace plots of the number of active communities K_n and parameter σ showing the MCMC algorithm can recover these parameters.

Data generated from a Poisson factor model with a fixed number of communities. Following Miller and Harrison (2013), we know that the Dirichlet process is inconsistent for estimating the number of clusters if the number of clusters is fixed and does not increase with n. Similarly, we conjecture that when $\sigma = 0$ is fixed, corresponding to the setting of Zhou (2015), the model will fail recovering the right number of communities, whereas if σ is free, we can expect it to concentrate on negative values. The model should then recover the correct number of communities if the data is generated from a Poisson factor model. A proper posterior consistency analysis is beyond the scope of this article. However, we design a simple numerical experiment to support this conjecture. We take $K = 5$ true communities, the affiliations $v_{i,k}$ are iid Gamma(0.1, 1) and the five communities importance $(r_1, \ldots, r_5)$ are iid Gamma(2, 0.2). Networks of increasing sizes $n = 500, 1000, 1500, 3000$ are then generated from the Poisson factor model (3).

We estimate the number of active communities K_n under the model with $\sigma = 0$ and with σ unknown. In Table 1, we report the posterior mean and variance of the recovered number of communities. We can see that the gamma process model ($\sigma = 0$) does not seem to concentrate around the right value. On the contrary, when σ is free, we see that the posterior of K_n seems to concentrate around $K = 5$. For

Table 1 Concentration of the posterior of K_n when increasing n on synthetic dataset with a number of communities $K = 5$ fixed (independent of n)

Statistic	$n = 500$	$n = 1000$	$n = 1500$	$n = 3000$
$\sigma = 0$	6.55 ($\pm$1.36)	6.11 ($\pm$1.16)	6.17 ($\pm$1.26)	6.59 ($\pm$1.20)
σ free	6.14 ($\pm$1.18)	5.40 ($\pm$0.85)	5.21 ($\pm$0.57)	5.12 ($\pm$0.43)
$\mathbb{P}(\sigma < 0)$	0.69	0.93	0.97	0.99

We report the mean and the standard deviation (between parenthesis) of the posterior of K_n for two models: first line $\sigma = 0$ is constant, recovering the Dirichlet model second line σ is free and inferred from the data, in the third line we report the posterior probability of $\sigma < 0$ when σ is considered unknown

$n = 500$, the posterior on σ ranges from negative to positive values, which explains the higher variance. However, from $n = 1000$ onward, the posterior of σ concentrates on negative values, which translates in a significant decrease of variance for the posterior of K_n.

Data generated from the Stochastic Block Model (SBM). From our experimentation, it seems that Poisson factorization models do not capture well the generating process of the SBM. Both when $\sigma = 0$ and σ free, the model creates many very small communities. However, the model is able to recover with high precision the true communities once we cluster the nodes using (29).

We generate a synthetic dataset from a SBM with $n = 600$ with three communities of size 200 each. The probability of an edge between nodes of a same community is $p_{in} = 0.1$; the probability of an edge between nodes from different communities is $p_{out} = 0.01$. The resulting dataset is hence undirected and unweighted with 7239 edges.

The GGP-Gamma model does not capture well the SBM generating process and the posterior mean of the number of communities is $K_n = 38.6$, with σ positive. We obtain 3 main communities and the rest are very small (composed of a few edges each). As explained previously, we cluster the nodes by assigning each one of them to the community to which it has the highest scaled affiliation. We then get an average of 8.35 communities, 3 of which are on average composed of 195.5 nodes each. We plot the distribution of the sizes of the small communities in Figure 2. In Table 2, we report the contingency table of the posterior clusters.

We also use the model with $\sigma = 0$ on this dataset and find very similar results. After clustering the nodes, we find that the average number of communities is 7.66, which slightly less than previously, but the small communities are slightly larger on average, giving at the end an average size for each large cluster of 195.5, which is exactly the same. We do not report here the contingency matrix as it is very similar to the one we obtain with σ free.

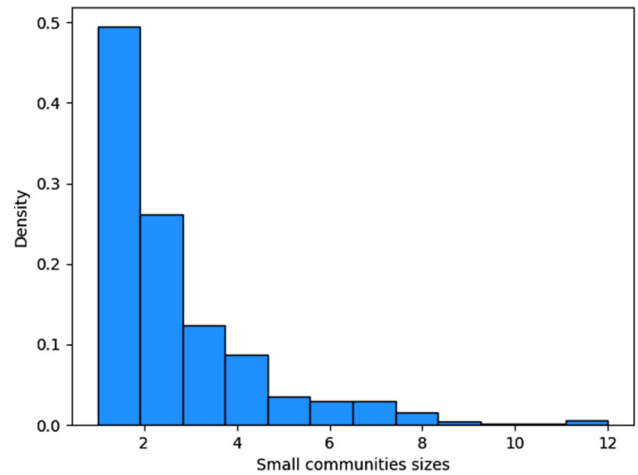


Fig. 2 Posterior density of the sizes of the small communities for the dataset generated from a Stochastic Block Model

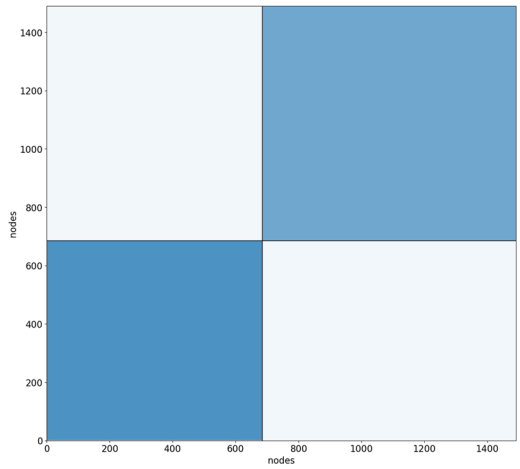
Table 2 Contingency table of the posterior communities in per cent of the true communities size

	Estimated communities			
	$\widehat{C}_1$	$\widehat{C}_2$	$\widehat{C}_3$	Others
C_1	**97.67**	0.09	0.09	2.15
C_2	0.01	**97.94**	0.0	1.95
C_3	0.14	0.0	**97.43**	2.43

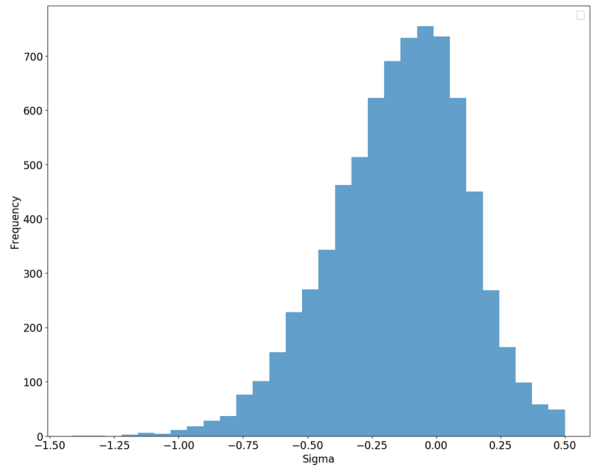
5.2 Political blogs

The polblogs network (Adamic and Glance 2005) is the network of the American political blogosphere in February 2005. It is a directed unweighted graph, where there is an edge (i, j) if blog i cites blog j. It is composed of 1490 nodes and 19025 edges. For each node, some ground truth information about its political affiliation (republican/democrat) is known.

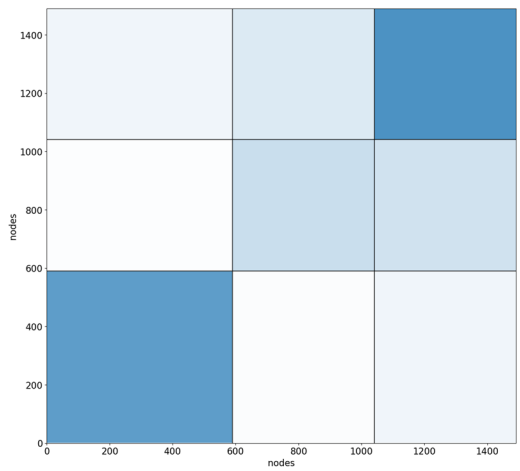
We will use this dataset in order to illustrate the role of the parameter α in the model. As indicated in Sect. 3, this parameter tunes the amount of overlapping between the communities. A smaller value enforces less overlap between communities. We run three chains with 500, 000 iterations. The posterior samples of σ for three different values of α



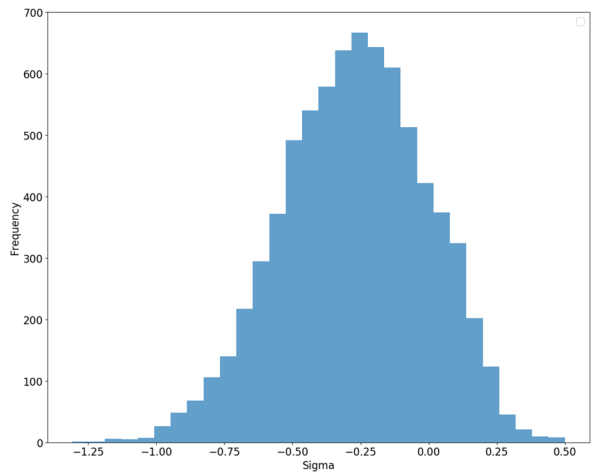
(a) Block densities of reordered adjacency matrix



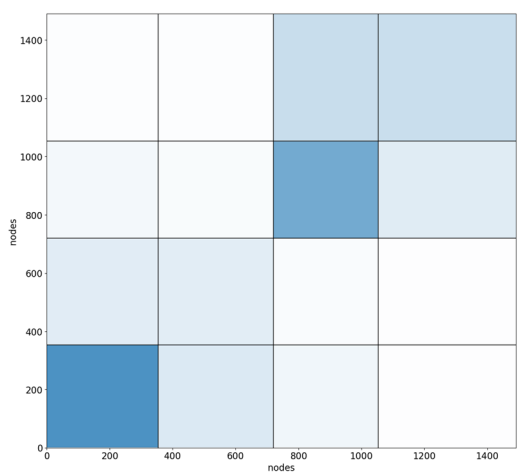
(b) Histogram of σ



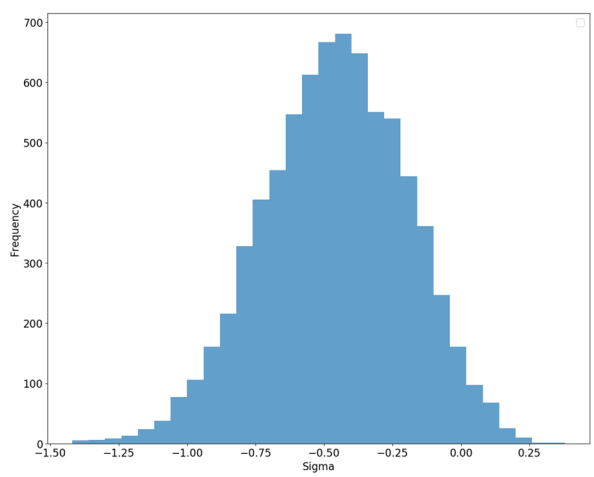
(c) Block densities of reordered adjacency matrix



(d) Histogram of σ



(e) Block densities of reordered adjacency matrix

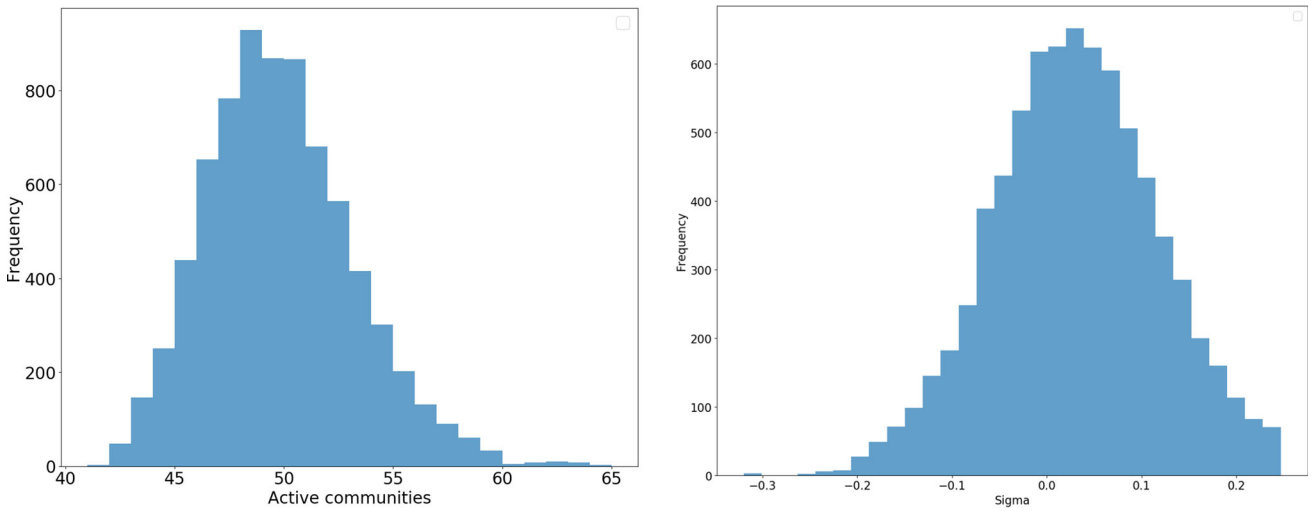


(f) Histogram of σ

Fig. 3 (Left) Estimated communities and (right) posterior on σ for the polblogs dataset with (top row) $\alpha = 0.8$, (middle row) $\alpha = 0.4$ and (bottom row) $\alpha = 0.2$

Table 3 Proportion of the interactions of the features in each block for different values of overlapping

	$\alpha = 0.8$		$\alpha = 0.4$			$\alpha = 0.2$			
	1	2	1	2	3	1	2	3	4
Dem/Dem	91.5	0.7	93.2	1.1	0.8	95	84	1.5	0.1
Dem/Rep	8.0	9.5	6.5	13	5.1	5	14.5	15.5	4.8
Rep/Rep	0.5	89.8	0.3	85.9	94.1	0	1.5	83	95.1



(a) Histogram of number of communities **(b)** Histogram of σ

Fig. 4 Posterior of K_n and σ for the Wiki-topcats dataset

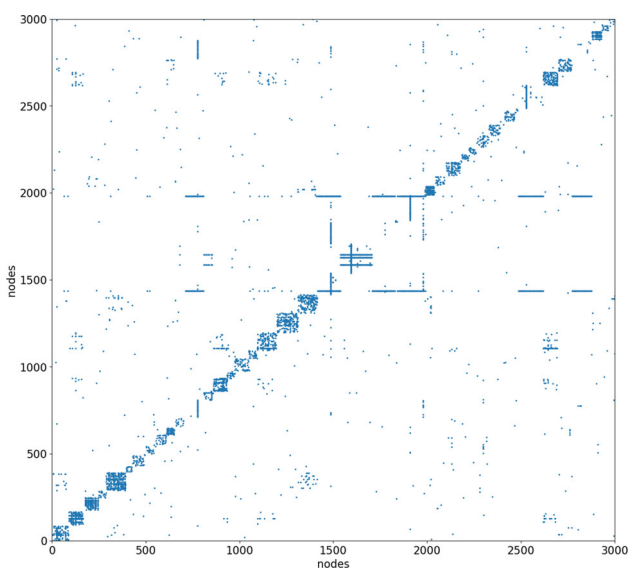


Fig. 5 Reordered adjacency matrix of the Wikipedia topcats dataset

are in also shown in Fig. 3. Nodes are reordered according to their estimated membership $\widehat{c}$ (through (29)), and Fig. 3 shows the densities of connection between and within clusters for three different values of α. Depending on the amount of overlapping, we obtain two ($\alpha = 0.8$), three ($\alpha = 0.4$) or four ($\alpha = 0.2$) communities. In order to interpret those communities, we calculate in Table 3 for each community the proportion of interactions between democrat blogs, between a democrat and a republican blog, and between two republican blogs. For $\alpha = 0.8$, there are two estimated communities which can clearly be identified as democrat (community #1) and republican (community 2). For $\alpha = 0.4$, we have three communities. One is mostly associated with democrat blogs (#1), while the other two correspond to a split of the republican blogs into right (#2) and center-right (#3) groups. For $\alpha = 0.2$, we obtain a further split of the democrat blogs into left (#1) and center-left (#2) groups. Increasing the value of α therefore leads to a finer and finer partition of the nodes.

5.3 Wikipedia topcast

The network is a partial web graph of Wikipedia hyperlinks collected in September 2011 (Klymko et al. 2014). It is a directed unweighted graph where an edge (i, j) corresponds to a citation from a page i to page j. We restrict it to the first 3000 nodes, and the associated 5687 edges. We run three MCMC chains for 200, 000 iterations. Trace plots of the number of active communities and parameter σ are given in Fig. 4. Figure 5 shows the adjacency matrix reordered by communities, as explained in the previous section. In order to

check that the learnt communities/features are meaningful, we report in Figures the proportions of webpages associated with a given category within a given community/feature (note that a webpage can be associated with multiple categories hence the proportion do not sum to 1).

Note that, while the approach is able to estimate the latent block structure, this dataset has the particularity of having star nodes, a feature that is not captured by our model.

5.4 Deezer

The dataset was collected from the music streaming service Deezer in November 2017 (Rozemberczki et al. 2018). It represents the friendship network of a subset of Deezer users from Romania. It is an undirected unweighted graph where nodes represent the users and edges are the mutual friendships. There are 41773 nodes and 125826 edges. We run three chains with 100000 iterations each. Posterior histograms of the number of active communities and σ are given in Fig. 8. The algorithms find around 45 communities/features for this dataset. The reordered adjacency matrix and block densities based on the point estimate of the partition are given in Fig. 9.

Now we can reorder the nodes using approximate MAP clustering as previously. We obtain the following adjacency matrix

For each individual in the network, a list of musical genres liked by that person are available. There are in total 84 distinct genres. We represent in Fig. 10 the proportion of individuals who liked a subset of the 84 genres for three different communities where the interpretation in terms of genres is quite clear. The overall proportion of individuals liking a given genre is shown at the bottom of Fig. 10. If the bar is red, this indicates that the proportion is 10% higher in the community than in the population. If the bar is blue, this means the proportion is 10% lower. Community 11 can be interpreted as $R\&B$, Community 8 as Dance, and Community 3 as Rock music. For some of the communities, not reported here, the interpretation in terms of the liked genres is less clear, and may be due to other covariates.

6 Discussion

The model presented in this paper assumed the same parameter β for each node. We can also consider a degree corrected version of the model, similarly to Zhou (2015), where each node is assigned a different parameter $\beta_i > 0$ and then defining $Z_{ijk} \sim \text{Poisson}(\frac{r_k v_{ik} v_{jk}}{\beta_i \beta_j})$. It is unclear however if a MCMC sampler targeting the exact posterior distribution could be implemented, and one may need to resort to some truncation approximation as in Zhou (2015).

The count matrix (A_{ij}) is infinitely exchangeable; hence, the model presented in this article lead to asymptotically dense graphs. That is, $\sum_{1\le i,j\le n} A_{ij} \asymp n^2$ as n tends to infinity. In order to obtain sparse graphs, we could consider two different strategies. The first solution consists in dropping the infinite exchangeability property and take $\beta_i^{(n)} \to +\infty$ with n, then the number of edges will behave as $(n/\beta^{(n)})^2$ (we can for instance take $\beta_i^{(n)} = \sqrt{n}$ for any node i to obtain a linear growth of the number of edges). The model would still be finitely exchangeable for any fixed n, but not projective anymore. The second solution would be to consider the different notion of infinite exchangeability developed in (Caron and Fox 2017) and consider $(\beta_i)_i$ as a realization of a Poisson point process.

Finally, we presented a model for count (and binary) data. The results build on the additive contributions of the communities, which is why we chose the Poisson distribution on the entries of the adjacency matrix (A_{ij}). We can generalize to non-count data using other probability distributions which are closed under convolution. For example, one could consider $A_{ij} \sim \text{Gamma}(\sum_k r_k v_{ik} v_{jk}, 1)$ for $A_{ij} \in \mathbb{R}_+$ or $A_{ij} \sim \mathcal{N}(\sum_k r_k v_{ik} v_{jk}, 1)$ for $A_{ij} \in \mathbb{R}$.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

A Proofs

A.1 Technical Lemmas

Lemma 3 (Gnedin et al. (2007), Propositions 17 and 19) *Let ρ be a Lévy measure, let $\overline{\rho}(x) = \int_x^\infty \rho(r)dr$ be the tail Levy intensity and $\psi(t) = \int (1-e^{-rt})\rho(dr)$ its Laplace exponent. Then, the two following conditions are equivalent:*

$$\overline{\rho}(x) \overset{x\to 0_+}{\asymp} \ell(1/x)x^{-\sigma} \tag{30}$$

$$\psi(t) \overset{x\to +\infty}{\asymp} \Gamma(1-\sigma)t^\sigma \ell(t) \tag{31}$$

with ℓ a slowly varying function and $0 \le \sigma < 1$.

Besides, if we let $\psi_d(t) = \frac{t^d}{d!}\int r^d e^{-rt}\rho(dr)$

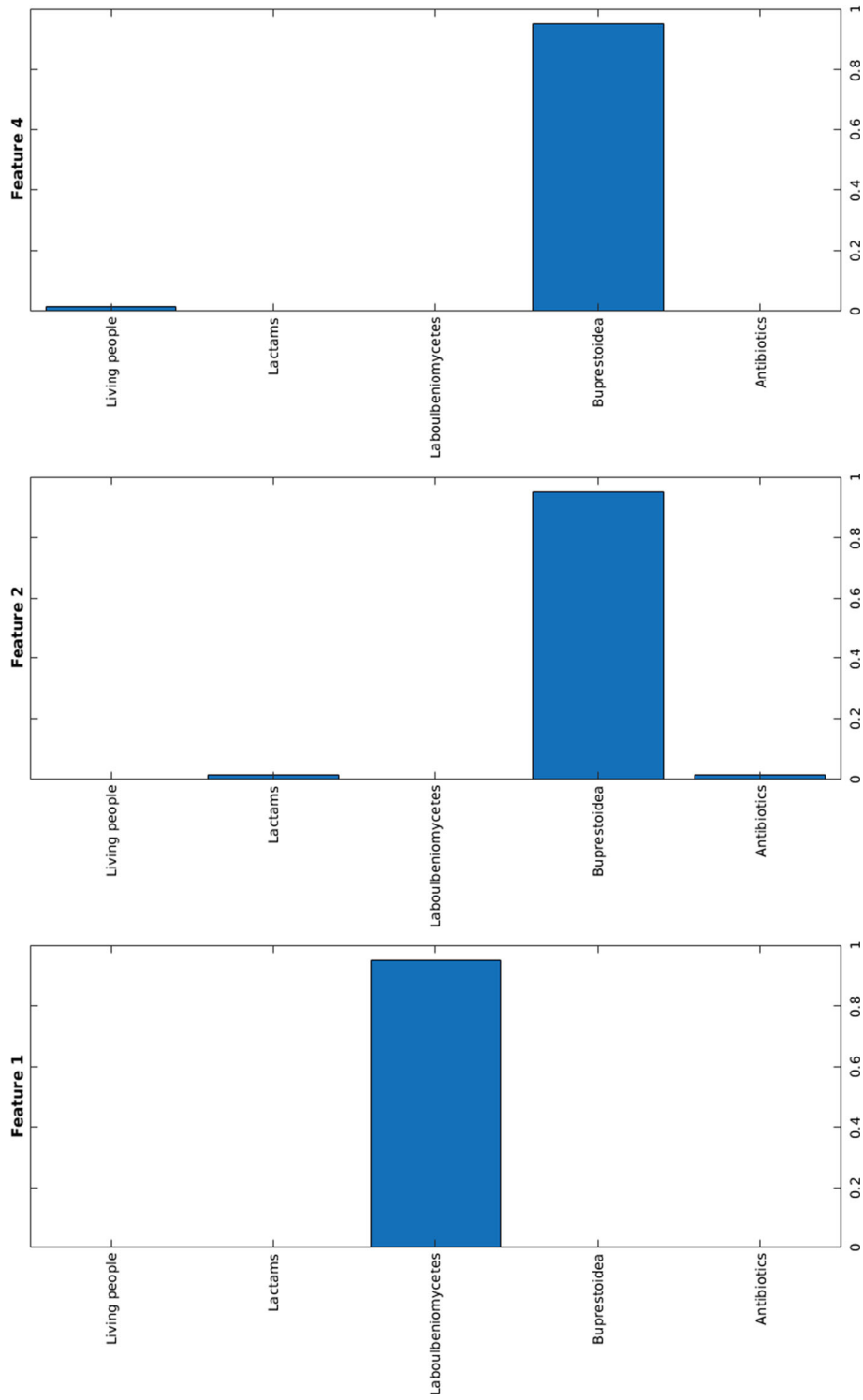


Fig. 6 Features compared to categories for the Wikipedia dataset

1. *if* $\sigma > 0$, *then* (30) *implies that* $\psi_d(t) \overset{t\to+\infty}{\asymp} \frac{\sigma\Gamma(d-\sigma)}{d!} t^\sigma \ell(t)$
2. *if* $\sigma = 0$, *then* (30) *implies that* $\psi_d(t) = o(\ell(t))$

Lemma 4 *(Pollard 2015, Exercise 15) Let X be a Poisson random variable with parameter* λ *. For any* $t > 0$

$$\mathbb{P}(|X - \lambda| \geq \lambda t) \leq 2e^{-\frac{\lambda t^2}{2(1+t)}}. \tag{32}$$

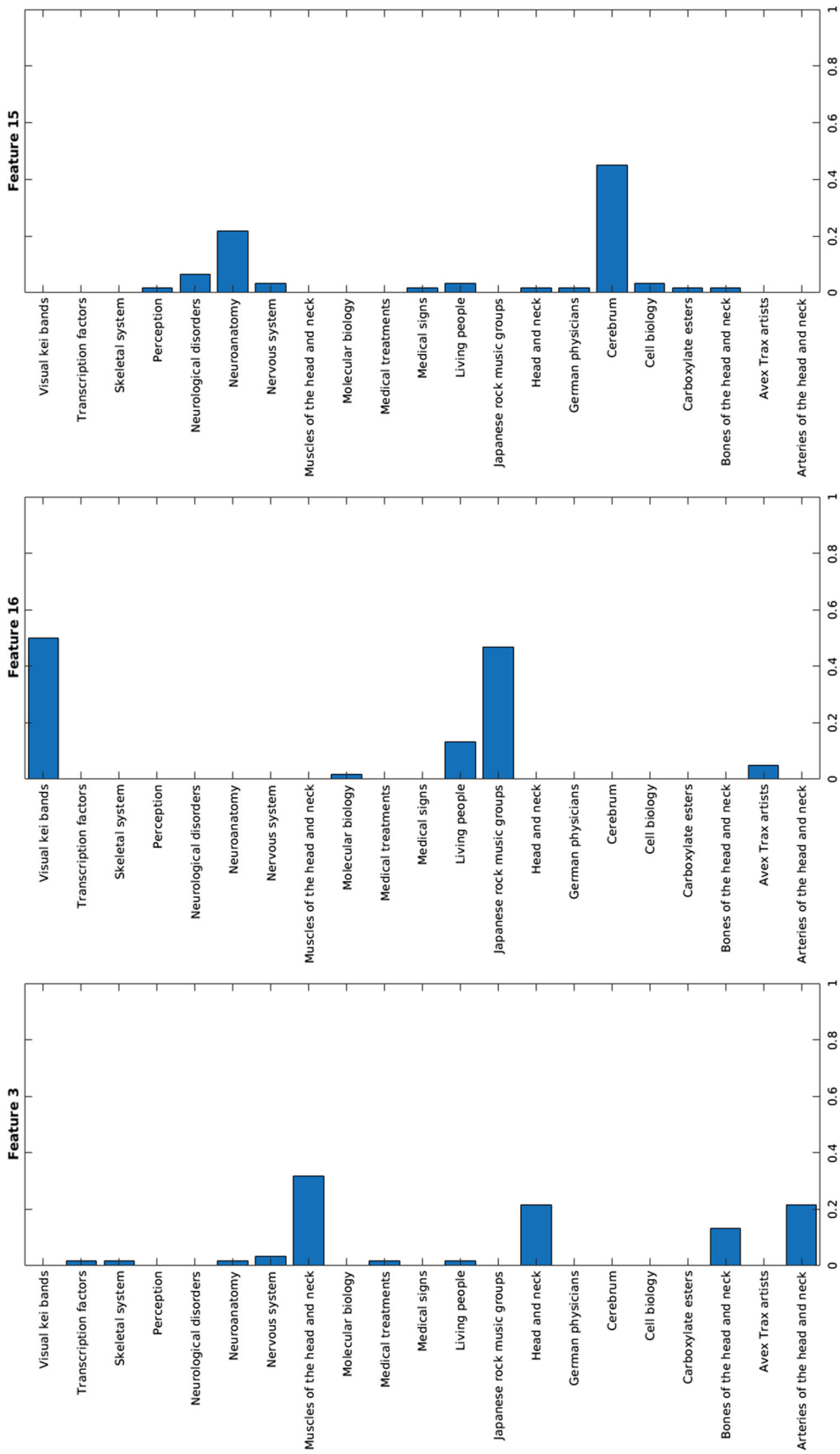
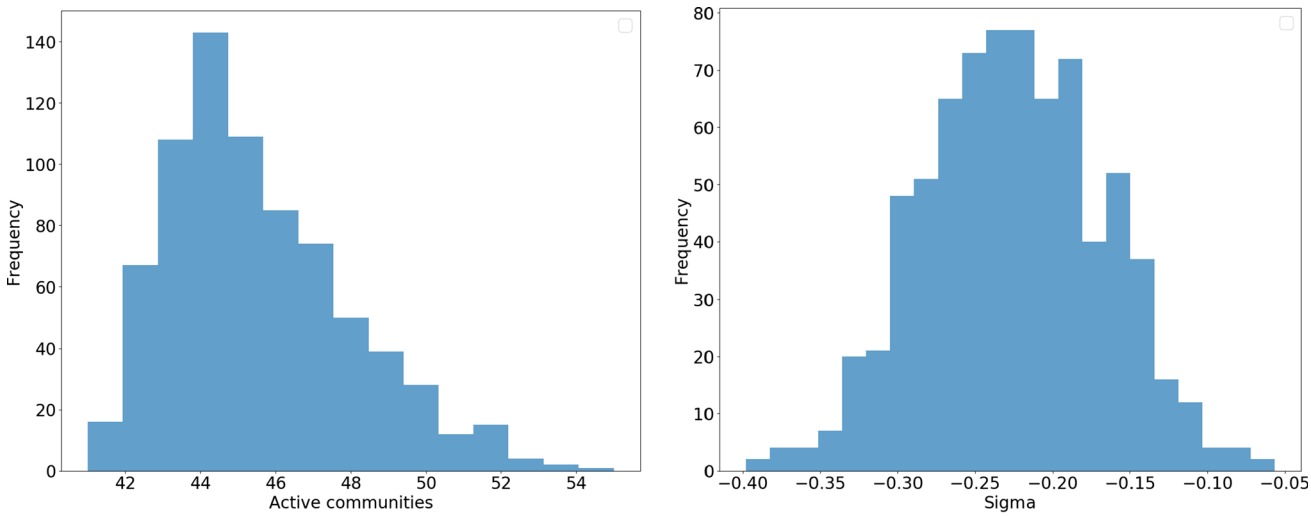


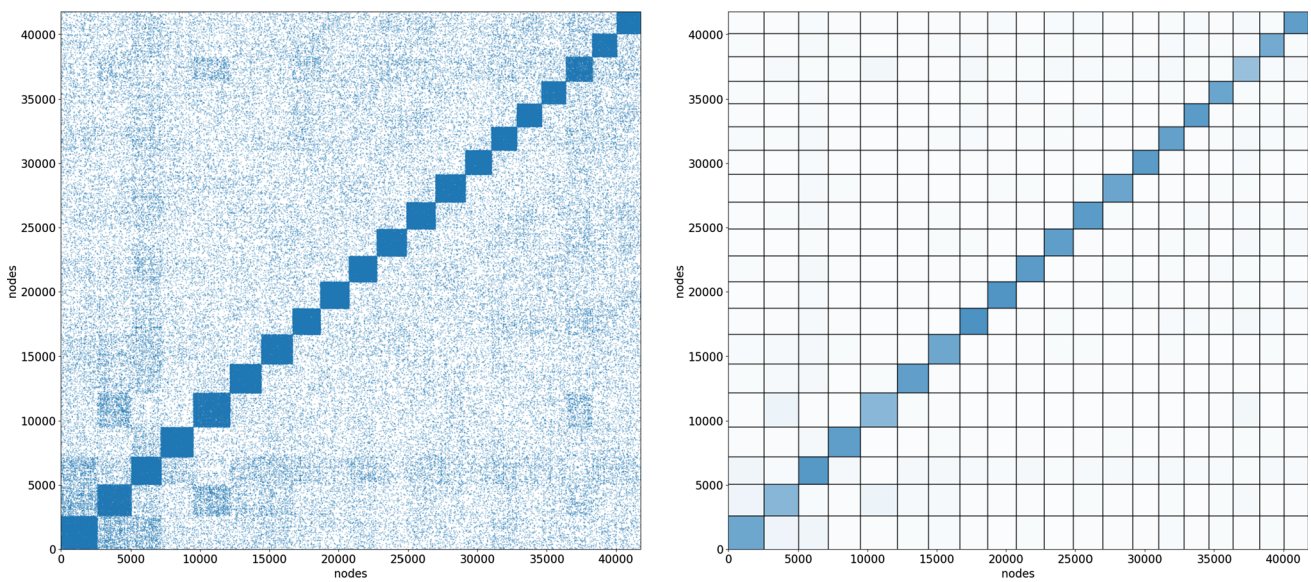
Fig. 7 Features compared to categories for the Wikipedia dataset



(a) Histogram of number of communities

(b) Histogram of σ

Fig. 8 Posterior of K_n and σ on Deezer's dataset



(a) Reordered Adjacency matrix

(b) Block densities of reordered adjacency matrix

Fig. 9 Reordered adjacency matrix and block densities for Deezer's dataset

Lemma 5 *Let $(X_n)_{n\geq 1}$ be a sequence of Poisson random variables with mean $(\mu_n)_{n\geq 1}$. If $\log n = o(\mu_n)$ then $X_n \asymp \mu_n$ almost surely as n tends to infinity.*

Proof Let $0 < \epsilon < 1/2$. Using Lemma 4, we have

$$\mathbb{P}\left(\left|\frac{X_n}{\mu_n} - 1\right| \geq \epsilon\right) \leq 2e^{-\frac{\epsilon^2\mu_n}{4}} = 2n^{-\frac{\epsilon^2\mu_n}{4\log n}} \tag{33}$$

Using the assumption, we have that $-\frac{\epsilon^2\mu_n}{4\log n} \to -\infty$. Therefore, the RHS of (33) is summable. The almost sure result follows from Borel–Cantelli lemma. □

Lemma 6 *For any $x, y \geq 0$, we have the following bound*

$$1 - e^{-xy} \leq \max(1, y)(1 - e^{-x}) \tag{34}$$

Proof The bound is trivial when $y \leq 1$. Consider the case $y > 1$. For all x, the function $y \to e^{-xy}$ is convex hence $f_x(y) = \frac{e^{-xy}-1}{y}$ is a monotonically non-decreasing function of y therefore

$$\frac{e^{-xy} - 1}{y} \geq e^{-x} - 1$$

for $y \geq 1$. □

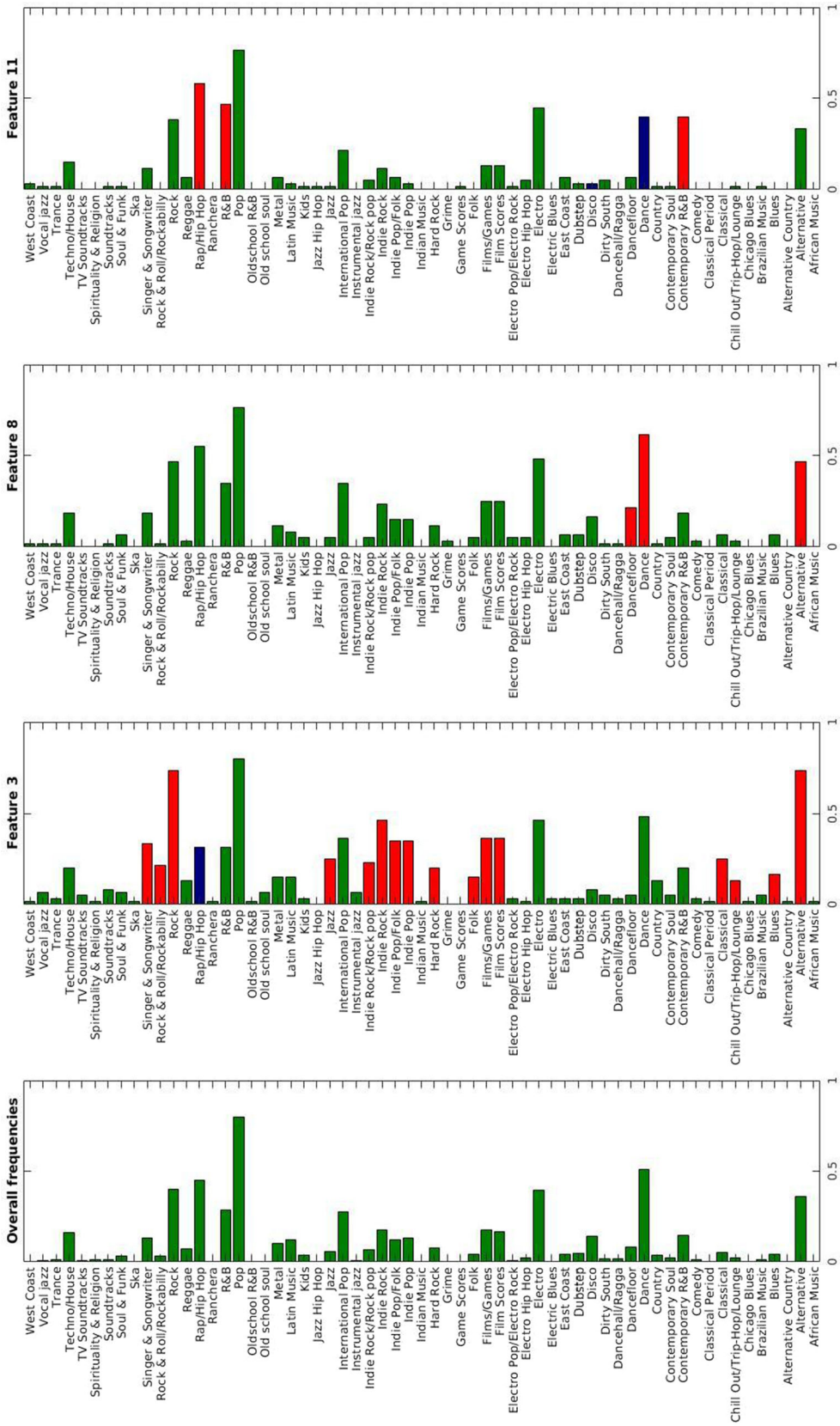


Fig. 10 Features compared to genres for Deezer's dataset

A.2 Proofs of Section 3

Proof of Proposition 1 The result for K_n is proved in James (2014) in the general context of GIBP. We provide here the details of the proof for K_n, which can be straightforwardly adapted to $K_{n,j}$.

First, let us remark that the bound (34) together with assumptions (A1) and (A2) imply $\Psi(n) < \infty$. For $s < 0$,

$$\begin{aligned}\mathbb{E}[e^{sK_n}] &= \mathbb{E}[\prod_{k=1}^{\infty}\mathbb{E}[e^{s\mathbb{1}_{\sum_{1\le i,j\le n} Z_{ijk}\ge 1}}|G]\,]\\ &= \mathbb{E}[\prod_{k=1}^{\infty}[\,e^s+(1-e^s)e^{-r_k(\sum_{i=1}^n v_{ik})^2}\,]\\ &= \mathbb{E}[\,e^{\sum_k \log[e^s+(1-e^s)e^{-r_k(\sum_{i=1}^n v_{ik})^2}]}\,]\end{aligned}$$

Then, since

$$\begin{aligned}&\log[e^s+(1-e^s)e^{-r_k(\sum_{i=1}^n v_{ik})^2}]\\ &= \log[e^{-r_k(\sum_{i=1}^n v_{ik})^2}+e^s(1-e^{-r_k(\sum_{i=1}^n v_{ik})^2})]\\ &\le \log[1+e^s(1-e^{-r_k(\sum_{i=1}^n v_{ik})^2})]\\ &\le 1-e^{-r_k(\sum_{i=1}^n v_{ik})^2}\end{aligned}$$

and the last part is integrable, we can use Campbell's theorem (Kingman 1993) to get:

$$\begin{aligned}\mathbb{E}[e^{sK_n}] &= \exp[\int(e^{\log[e^s+(1-e^s)e^{-r(\sum_{i=1}^n v_i)^2}]}-1)\\ &\quad\prod_{i=1}^n F(dv_i)\,\rho(dr)\,]\\ &= \exp[\,(e^s-1)\Psi(n)\,]\end{aligned}$$

We can prove similarly that $K_{n,d}$ is a Poisson random variable with mean $\Psi_d(n)$ and that $\sum_{d\ge D} K_{n,d}$ is Poisson distributed with mean $\sum_{d\ge D}\Psi_d(n)$. The assumption $\Psi_d(n) < \infty$ is also sufficient in this case to apply Campbell's theorem.

Proof of Proposition 2 From Proposition 1, we get that

$$\mathbb{E}(K_n) = \int(1-e^{-r(\sum_{i=1}^n v_i)^2})\prod_{i=1}^n F(dv_i)\,\rho(dr) = \Psi(n) \tag{35}$$

Let $(V_i)_{i\in\mathbb{N}}$ be i.i.d random variables with distribution F. By assumption, $0 < \mathbb{E}[V_i] = \mu < +\infty$ and $\mathbb{V}ar[V_i] = \tau^2 < +\infty$. Let $\epsilon > 0$. Let $A(r)$ be defined for $r > 0$ by

$$A(r) = \mathbb{E}[1-e^{-r(\sum_{i=1}^n V_i)^2}]$$

Since $v \mapsto 1-e^{-rv}$ is concave, using successively Jensen's inequality and the independence of (V_i), we obtain

$$\begin{aligned}A(r) &\le 1-e^{-r\mathbb{E}[(\sum_{i=1}^n V_i)^2]}\\ &\le 1-e^{-r(n^2\mu^2+n\tau^2)}\\ &\le 1-e^{-(1+\epsilon)rn^2\mu^2}\end{aligned}$$

where the last inequality holds for any $\epsilon > 0$ when $n > \frac{\tau^2}{\epsilon\mu^2}$. Therefore, for $\epsilon > 0$ and $n > \frac{\tau^2}{\epsilon\mu^2}$

$$\Psi(n) \le \int(1-e^{-(1+\epsilon)rn^2\mu^2})\rho(dr).$$

Besides, since $v \mapsto 1-e^{-rv}$ is increasing, by Markov's inequality we have for any $\epsilon > 0$

$$\begin{aligned}A(r) &= \mathbb{E}\left[1-e^{-rn^2(\frac{1}{n}\sum_{i=1}^n V_i)^2}\right]\\ &\ge \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n V_i \ge \frac{\mu}{(1+\epsilon)}\right)\left(1-e^{-\frac{n^2\mu^2}{(1+\epsilon)^2}r}\right)\end{aligned}$$

Hence,

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n V_i \ge \frac{\mu}{(1+\epsilon)}\right)\psi\left(\frac{n^2\mu^2}{(1+\epsilon)^2}\right)\\ &\le\ \Psi(n)\ \le \psi\left((1+\epsilon)n^2\mu^2\right)\end{aligned}$$

where $\psi(t) = \int(1-e^{-rt})\rho(dr)$ is the Laplace exponent. Furthermore, by the law of large numbers,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n V_i \ge \frac{\mu}{(1+\epsilon)}\right) \to 1.$$

Therefore, under Assumption (A4), Lemma 3 implies

$$\mathbb{E}(K_n) \asymp \Gamma(1-\sigma)\mu^{2\sigma}n^{2\sigma}\ell(n^2)$$

as n tends to infinity.

In the finite-activity case, that is $\sigma = 0$ and $\ell(t) \to \overline{\rho}(0) < \infty$, we have $\mathbb{E}(K_n) \to \overline{\rho}(0)$ hence K_n tends in distribution to Poisson$(\overline{\rho}(0))$.

Now, for $\sigma > 0$, the almost sure result (15) follows from Lemma 5 and the fact that for every slowly varying function ℓ_0 and every $\epsilon > 0$

$\lim_{x\to\infty} \ell_0(x)x^{-\epsilon} = 0.$

Finally, assume that $(K_n)_{n\geq 1}$ is non-decreasing. We only need to prove the asymptotic behavior for $\sigma = 0$. In that setting, $\Psi(n) \asymp \ell(n^2)$. Using the assumption that $\int \rho(dr) = \infty$, we therefore have $\lim_{n\to\infty} \Psi(n) = \infty$. Let $n \geq 1$,

$$\Psi(n+1) - \Psi(n) = \int \mathbb{E}\left[e^{-r(\sum_{i=1}^{n} V_i)^2} - e^{-r(\sum_{i=1}^{n+1} V_i)^2} \right] \rho(r)dr$$

Since $V_i \geq 0$ a.s, it comes that $(\Psi(n))_n$ is non-decreasing. Now extend the sequence $(\Psi(n))_n$ to a non-decreasing and continuous function Ψ on $\mathbb{R}_+$ (by linear interpolation for instance). Let $t > 1$, then

$$1 \leq \frac{\Psi(t+1)}{\Psi(t)} \leq \frac{\Psi(3\lfloor t\rfloor)}{\Psi(\lfloor t\rfloor)} \asymp \frac{\ell(9\lfloor t\rfloor^2)}{\ell(\lfloor t\rfloor^2)} \to 1$$

Hence $\lim \frac{\Psi(t+1)}{\Psi(t)} = 1$

Now, for every integer $m \geq \Psi(0)$, choose t_m such that $\Psi(t_m) = m$. We have that (t_m) is non-decreasing and diverges. Since Ψ is increasing, it comes

$$\Psi(t_{m+1} - 1) \leq \Psi(\lfloor t_{m+1}\rfloor - 1) \leq \Psi(t_{m+1}) = m + 1$$

Hence, $\Psi(\lfloor t_{m+1}\rfloor - 1) \asymp m$. Then, using Lemma 5, we get that

$$K_{\lfloor t_{m+1}\rfloor - 1} \asymp \Psi(\lfloor t_{m+1}\rfloor - 1) \asymp m \quad a.s.$$

Finally, let $n \geq \Psi(0)$, let $m_n = \min\{m \mid n \in \{\lfloor t_m\rfloor, \ldots, \lfloor t_{m+1}\rfloor - 1\}\}$,

$$\frac{K_{\lfloor t_{m_n}\rfloor}}{\Psi(\lfloor t_{m_n+1}\rfloor - 1)} \leq \frac{K_n}{\Psi(n)} \leq \frac{K_{\lfloor t_{m_n+1}\rfloor - 1}}{\Psi(\lfloor t_{m_n}\rfloor)}$$

Since $t_{m_n} \to \infty$, both bounds converge to 1 almost surely, which gives the result. □

Proof of Proposition 3 As for Proposition 2, we only need to show that for $d \geq 1$,

$$\Psi_d(n) \asymp \frac{\sigma\Gamma(d-\sigma)}{d!} n^{2\sigma}\mu^{2\sigma}\ell(n^2)$$

Therefore, the proof is very similar to the one of Proposition 2. However, there are some technicalities we need to address since here $v \mapsto \frac{v^d}{d!}e^{-rv}$ is neither convex nor decreasing. Like previously, we will lower bound and upper bound $\Psi_d(n)$ by two quantities that are equivalent to $\frac{\sigma\Gamma(d-\sigma)}{d!} n^{2\sigma}\mu^{2\sigma}\ell(n^2)$.

Let us first introduce some notations. Let $(V_i)_{i\in\mathbb{N}}$ i.i.d variables with distribution F. Let $S_n = \sum_{i=1}^{n} V_i$ and $\varphi_d(r, v)$ defined as

$$\varphi_d(r, v) = \frac{r^d v^d}{d!} e^{-rv}.$$

Now, define $A_d(r, n)$ for $r > 0$ by

$$A_d(r, n) = \mathbb{E}[\varphi_d(r, S_n^2)]$$

Suppose that $0 < \sigma < 1$, let $\epsilon > 0$, recalling that $\mathbb{E}V_i = \mu$, define $B_\epsilon = [\frac{n\mu}{\epsilon+1}, (1+\epsilon)n\mu]$. Let us notice that the law of large numbers gives us that $\mathbb{P}(S_n \in B_\epsilon) \to 1$.

1. Lower bound: We have that

   $$\mathbb{E}[\mathbb{1}_{S_n\in B_\epsilon}\varphi_d(r, S_n^2)] \leq A_d(r, n)$$

   Besides, for all $v \in B_\epsilon$ and all $r > 0$

   $$\phi_d(r, v) \geq \frac{r^d\mu^{2d}n^{2d}}{(1+\epsilon)^{2d}d!} e^{-rn^2(1+\epsilon)^2\mu^2}$$

   hence

   $$\mathbb{P}(S_n \in B_\epsilon)\frac{r^d\mu^{2d}n^{2d}}{(1+\epsilon)^{2d}d!} e^{-rn^2(1+\epsilon)^2\mu^2} \leq \mathbb{E}[\mathbb{1}_{S_n\in B_\epsilon}\varphi_d(r, S_n^2)].$$

   Therefore, using Lemma 3 and since $\mathbb{P}(S_n \in B_\epsilon) \to 1$, we have for n large enough

   $$\frac{\sigma\Gamma(d-\sigma)}{d!(1+\epsilon)^{4d+1-2\sigma}} n^{2\sigma}\mu^{2\sigma}\ell(n^2) \leq \int A_d(r, n)\rho(dr).$$

2. Upper bound: We have that

   $$A_d(r, n) = \mathbb{E}[\mathbb{1}_{S_n\in B_\epsilon}\varphi_d(r, S_n^2)] + \mathbb{E}[\mathbb{1}_{S_n\notin B_\epsilon}\varphi_d(r, S_n^2)]$$

   Like previously, since

   $$\mathbb{E}[\mathbb{1}_{S_n\in B_\epsilon}\varphi_d(r, nS_n)] \leq \mathbb{P}(S_n \in B_\epsilon)\frac{r^d(1+\epsilon)^{2d}\mu^{2d}n^{2d}}{d!} e^{-r\frac{n^2\mu^2}{(1+\epsilon)^2}}$$

   We find that for n large enough,

   $$\int A_d(r, n)\rho(dr) \leq \frac{\sigma\Gamma(d-\sigma)(1+\epsilon)^{4d+1-2\sigma}}{d!} n^{2\sigma}\mu^{2\sigma}\ell(n^2) + \int \mathbb{E}[\mathbb{1}_{S_n\notin B_\epsilon}\varphi_d(r, S_n^2)]\rho(dr)$$

Therefore, we only need to prove that

$$\int \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \varphi_d(r, S_n^2)]\rho(r)dr = o(n^{2\sigma}\ell(n^2)).$$

In order to do so, we split the integral with respect to r in two parts, an integral over $(0, \frac{1}{n^2})$ and an integral over $(\frac{1}{n^2}, \infty)$ and show that both are $o(n^{2\sigma}\ell(n^2))$. Since $\varphi_d(r, v) \leq 1$,

$$\begin{aligned}
&\int_{1/n^2}^{\infty} \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \varphi_d(r, S_n^2)]\rho(dr) \\
&\leq \mathbb{P}(S_n \notin B_\epsilon) \int_{1/n^2}^{\infty} \rho(dr) \\
&= \mathbb{P}(S_n \notin B_\epsilon)\, \overline{\rho}(1/n^2) \\
&= o(n^{2\sigma}\ell(n^2))
\end{aligned}$$

where the last line follows from the law of large numbers and Assumption (A4). Besides,

$$\begin{aligned}
&\int_0^{1/n^2} \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \varphi_d(r, S_n^2)]\rho(dr) \\
&= \int_0^{1/n^2} \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \frac{rS_n^2}{d} \varphi_{d-1}(r, S_n^2)]\rho(dr) \\
&\leq \int_0^{1/n^2} \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} rS_n^2]\rho(dr) \\
&= \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \frac{S_n^2}{n^2}] \int_0^{1/n^2} rn^2\rho(dr) \\
&\leq \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \frac{S_n^2}{n^2}]\, e \int_0^{1/n^2} rn^2 e^{-rn^2}\rho(dr) \\
&\leq 8\mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \frac{S_n^2}{n^2}]\, \frac{\sigma\Gamma(d-\sigma)}{d!} n^{2\sigma}\ell(n^2)
\end{aligned}$$

where the last inequality holds for n large enough by Assumption (A4) and Lemma 3. Now, we have that

$$\frac{S_n^2}{n^2} < \frac{1}{n}\sum_{i=1}^{n} V_i^2$$

Since $(V_i^2)_i$ are i.i.d random variables in $\mathcal{L}_1$, we know that $(\frac{1}{n}\sum_{i=1}^{n} V_i^2)_{n\geq 1}$ is uniformly integrable. Therefore, $(\mathbb{1}_{S_n \notin B_\epsilon} \frac{S_n^2}{n^2})_{n\geq 1}$ is uniformly integrable. Besides, using the law of large numbers, the sequence converges almost surely, and hence in probability, to 0. Therefore, $\lim_n \mathbb{E}[\mathbb{1}_{S_n \notin B_\epsilon} \frac{S_n^2}{n^2}] = 0$, which concludes the proof.
For $\sigma = 0$, the previous computations for the upper bound give that almost surely, $\Psi_d(n) = o(\ell(n^2)) = o(\Psi(n))$. Now let $D > 1$,

$$\mathbb{E}\sum_{d\geq D} K_{n,d} = \Psi(n) - \sum_{d=1}^{D-1} \Psi_d(n) \asymp \ell(n^2)$$

And since $x \mapsto \sum_{d\geq D} \varphi_d(1, x)$ is non-decreasing, $(\mathbb{E}\sum_{d\geq D} K_{n,d})_n$ is non-decreasing, therefore, similarly to the proof for $\sigma = 0$ for $(K_n)_n$, we find that

$$\sum_{d\geq D} K_{n,d} \asymp \mathbb{E}\sum_{d\geq D} K_{n,d} \asymp \ell(n^2) \quad a.s$$

Therefore, we finally find that

$$\frac{K_{n,D}}{K_n} = \frac{\sum_{d\geq D} K_{n,d} - \sum_{d\geq D+1} K_{n,d}}{K_n} \to 0 \quad a.s$$

□

B Gibbs sampler

As mentioned in the main text, the observed graph can be directed or undirected, binary or count, and can have missing entries we would like to predict. Denote by B the observed graph. Here, we describe the steps of a Gibbs algorithm with stationary distribution

$$p(K_n, (\widetilde{r}_k, \widetilde{v}_{1:n,k})_{k=1,\ldots,K_n}, \theta \mid B).$$

Notice that observing the full matrix $B = A$ corresponds to a weighted and directed graph with no missing entry. Let $\mathcal{I}$ denote the set of all possible edges. In the directed case, $\mathcal{I} = \{(i, j) \mid 1 \leq i, j \leq n\}$ and on the undirected case $\mathcal{I} = \{(i, j) \mid 1 \leq i \leq j \leq n\}$. We say that (i, j) is not observed if we do not know the value of $A_{i,j}$. Remark that (i, j) can be observed and still $A_{i,j} = 0$. Denote $\mathcal{O}$ the set of all observed entries and $\mathcal{O}^c = \mathcal{I} \setminus \mathcal{O}$, the set on unobserved entry. For all unobserved entry $(i, j) \in \mathcal{O}^c$, set $B_{i,j} = -1$

Additionally, to deal with the unknown number of active communities K_n, we use auxiliary slice variables $s_{i,j}$ for all $(i, j) \in \mathcal{I}$, details are given in the following paragraphs. Denote s the smallest non-zero slice variable $s_{i,j}$ for $(i, j) \in \mathcal{I}$. By definition of the slice variables, $\widetilde{r}_k \geq s$ for all $k = 1, \ldots, K_n$. Let

$$\overline{G} = \sum_k r_k \delta_{v_{1:n,k}} \mathbb{1}_{r_k \geq s} := \sum_{k=1}^{\overline{K}_n} \overline{r}_k \delta_{\overline{v}_{1:n,k}}$$

be the CRM corresponding to the set of active or inactive communities with weight $r_k \geq s$, of (almost surely finite) cardinality $\overline{K}_n \geq K_n$. Denote $\overline{Z}_{ijk} \geq 0$ the associated community interactions, and $\overline{Z}_k = (\overline{Z}_{ijk})$.

B.1 Directed graph

For each observed pair $(i, j) \in \mathcal{O}$, we define the slice variable as

$$s_{ij}|(\widetilde{r}_k, \widetilde{Z}_{ijk})_{k=1,\dots,K_n} \sim \text{Unif}\left(0, \min_{\{k|\widetilde{Z}_{ijk}\geq 1\}} \widetilde{r}_k\right) \tag{36}$$

if $A_{ij} > 0$ and $s_{ij} = 0$ otherwise. For each non-observed entry $(i, j) \in \mathcal{O}^c$, we define $s_{i,j}$ by (38) if $\{k \mid \widetilde{Z}_{ijk} \geq 1\} \neq \emptyset$ and

$$s_{ij}|(\widetilde{r}_k, \widetilde{Z}_{ijk})_{k=1,\dots,K_n} \sim \text{Unif}(0, 1) \tag{37}$$

otherwise.

B.1.1 Gibbs sampler step 1 for weighted graph on observed entries

Updating $(\widetilde{Z}_k)_{k=1,\dots,K_n}|(s, \overline{G}), \theta, B$ on observed entries indexes

We sample $(\overline{Z}_l)_{l=1,\dots,\overline{K}_n}$ associated with all atoms of $\overline{G}$ and keep only the non-empty communities. For every $(i, j) \in \mathcal{O}$ such that $A_{i,j} > 0$. define the random variable $m_{ij} = \min_{\{l|\widetilde{Z}_{ijl}\geq 1\}} \widetilde{r}_l$. Then, writing the joint distribution it comes that independently for every such (i, j),

$$\mathbb{P}((\overline{Z}_{ijl})_{l=1,\dots,\overline{K}_n}|(s, G^+), \theta, B_{ij}) \propto \prod_{i,j} \frac{1}{m_{ij}} \mathbb{1}_{s_{ij}<m_{ij}} \text{Mult}((\overline{Z}_{ijl})_l; B_{ij}, (\overline{p}_{ijl})_l)$$

where Mult is the multinomial distribution and $\overline{p}_{ijl} = \frac{r_l v_{il} v_{jl}}{\sum_{t=1}^{\overline{K}_n} r_t v_{it} v_{jt}}$. Let $p_{ijl} = r_l v_{il} v_{jl}$. To simplify the notations, let us suppose that the atoms of $\overline{G}$ are in decreasing order. Remark that the indexing of $\widetilde{Z}$ is different from the one of $\overline{Z}$, the second corresponds to the one of the truncated random measure. For each observed edge (i, j) independently, we can proceed in 4 phases for this step.

1. Sample m_{ij} from the locations of $\overline{G}$ such that $\mathbb{P}(m_{ij} = r_L) \propto \frac{(\sum_{l=1}^{L} p_{ijl})^{B_{ij}} - (\sum_{l=1}^{L-1} p_{ijl})^{B_{ij}}}{r_L} \mathbb{1}_{s_{ij}<r_L}$.
2. For $l > L$, set $\overline{Z}_{ijl} = 0$
3. Sample $\overline{Z}_{ijL} \sim \text{tBin}(B_{ij}, \frac{p_{ijL}}{\sum_{l=1}^{L} p_{ijl}})$, where tBin is the zero truncated binomial distribution
4. Sample $(\overline{Z}_{ij1}, \dots, \overline{Z}_{ijL-1}) \sim \text{Mult}(B_{ij} - \overline{Z}_{ijL}, (\frac{p_{ijl}}{\sum_{t=1}^{L-1} p_{ijt}})_{l\leq L-1})$

B.1.2 Gibbs sampler step 1 for unweighted graph on observed entries

In this setting, we observe a binary matrix $B_{ij} = 1_{A_{ij}>0}$. Then, the first step of the Gibbs sampler is modified and becomes:

Updating $(\widetilde{Z}_k)_{k=1,\dots,K_n}|(s, \overline{G}), \theta, B$ on observed entries indexes

For each observed edge $(i, j) \in \mathcal{O}$ independently do

1. Sample m_{ij} from the locations of $\overline{G}$ such that

$$\mathbb{P}(m_{ij} = r_L) \propto \frac{e^{\sum_{k=1}^{L} p_{ijl}} - e^{\sum_{k=1}^{L-1} p_{ijk}}}{r_L} \mathbb{1}_{s_{ij}<r_L}$$

   Suppose $m_{ij} = r_L$
2. For $l > L$, set $\overline{Z}_{ijl} = 0$
3. Sample $\overline{Z}_{ijL} \sim \text{tPoisson}(p_{ijL})$, where tPoisson is the zero truncated Poisson distribution
4. For $l < L$, sample $Z_{ijl} \sim Poiss(p_{ijl})$

B.1.3 Gibbs sampler step 1 on unobserved entries

For each unobserved entry $(i, j) \in \mathcal{O}^c$, knowing s_{ij}, we define $L_0 = \max\{k \mid r_k > s_{ij}\}$.

1. Draw $1_{A_{ij}=0}$, which is a Bernoulli with parameter

$$p = \frac{1}{1 + \sum_{L=1}^{L_0} \frac{e^{\sum_{k=1}^{L} p_{ijk}} - e^{\sum_{k=1}^{L-1} p_{ijk}}}{r_L}}$$

2. If $A_{i,j} \neq 0$, then use subsection B.1.2. Otherwise, set all counts of that entry to zero

B.2 Undirected graph

In the undirected graph, we suppose that for $i \neq j$, $B_{ij} = A_{ij} + A_{ji}$ and $B_{ii} = A_{ii}$. Besides, in this setting we actually do not need to sample $\widetilde{Z}_{ijk}$ for all (i, j, k) but only $\widetilde{Z}_{ijk} + \widetilde{Z}_{jik}$. For each observed pair $(i, j) \in \mathcal{O}$, we define the slice variable as

$$s_{ij}|(\widetilde{r}_k, \widetilde{Z}_{ijk} + \widetilde{Z}_{jik})_{k=1,\dots,K_n} \sim \text{Unif}\left(0, \min_{\{k|\widetilde{Z}_{ijk}+\widetilde{Z}_{jik}\geq 1\}} \widetilde{r}_k\right) \tag{38}$$

if $B_{ij} > 0$ and $s_{ij} = 0$ otherwise. For each non-observed entry $(i, j) \in \mathcal{O}^c$, we define s_{ij} by (38) if $\{k \mid \widetilde{Z}_{ijk} + \widetilde{Z}_{jik} \geq 1\} \neq \emptyset$ and

$$s_{ij}|(\widetilde{r}_k, \widetilde{Z}_{ijk} + \widetilde{Z}_{jik})_{k=1,\dots,K_n} \sim \text{Unif}(0,1) \tag{39}$$

otherwise. Then, step 2 and step 3 remain unchanged. For step 1, simply replace p_{ijk} by $2p_{ijk}$ for $i \neq j$.

B.3 Proofs for the Gibbs sampler step 1

B.3.1 Weighted graph

Here, will give the posterior distribution of the count matrices and show that

$$\begin{aligned}&(\widetilde{v}_k, \widetilde{Z}_k)_{k=1,\dots,K_n}|(s, \overline{G}), \theta, B\\ &\overset{dist}{=} (\widetilde{v}_k, \widetilde{Z}_k)_{k=1,\dots,K_n}|s, G, \theta, B\end{aligned}$$

In order to do so, we derive the RHS posterior distribution. Let us first notice that given G, sampling the nonzero counts and corresponding locations is equivalent to sampling $(Z_k)_k$ for $k \in \mathbb{N}$. As stated previously, we can treat each edge (i, j) independently. Therefore, we sample the sequence $(Z_{ijk})_k$. Here, we suppose that the communities come with decreasing activity order. Let the random variable $L = \max\{k \mid Z_{ijk} > 0\}$ (supposing that the $(r_k)_k$ are decreasing). And let $p_{ijk} = r_k v_{ik} v_{jk}$

$$\begin{aligned}&\mathbb{P}((Z_{ijk})_k|s, G, \theta, A_{ij})\\ &\propto \mathbb{P}((Z_{ijk})_k|G, \theta, A_{ij}) \times \mathbb{P}(s_{ij}|(Z_{ijk})_k, G, \theta, A_{ij})\\ &\propto \mathbb{1}_{\sum_k Z_{ijk}=A_{ij}} \frac{A_{ij}!}{\prod_{k=1}^{L} Z_{ijk}!} \prod_{k=1}^{L} p_{ijk}^{Z_{ijk}} \times \frac{1}{r_L}\mathbb{1}_{s_{ij}<r_L}\\ &\propto \frac{(\sum_{k=1}^{L} p_{ijk})^{A_{ij}}(1-(\frac{\sum_{k=1}^{L-1} p_{ijk}}{\sum_{k=1}^{L} p_{ijk}})^{A_{ij}})}{r_L}\mathbb{1}_{s_{ij}<r_L}\\ &\quad\times \mathbb{1}_{1\le Z_{ijL}\le A_{ij}} \frac{A_{ij}!}{Z_{ijL}!(A_{ij}-Z_{ijL})!}\\ &\quad\times \frac{(\frac{p_{ijL}}{\sum_{k=1}^{L} p_{ijk}})^{Z_{ijL}}(\frac{\sum_{k=1}^{L-1} p_{ijk}}{\sum_{k=1}^{L} p_{ijk}})^{A_{ij}-Z_{ijL}}}{1-(\frac{\sum_{k=1}^{L-1} p_{ijk}}{\sum_{k=1}^{L} p_{ijk}})^{A_{ij}}}\\ &\quad\times \mathbb{1}_{\sum_{k=1}^{L-1} Z_{ijk}=A_{ij}-Z_{ijL}} \frac{(A_{ij}-Z_{ijL})!}{\prod_{k=1}^{L-1} Z_{ijk}!} \prod_{k=1}^{L-1}(\frac{p_{ijk}}{\sum_{k=1}^{L-1} p_{ijk}})^{Z_{ijk}}\end{aligned}$$

This shows how we can sample in three steps these variables. Let us remark that the second part corresponds to the distribution of a zero truncated binomial and that the third part corresponds to the distribution of a multinomial. We also notice that only the elements of $\overline{G}$ are actually needed.

B.3.2 Unweighted Graph

We proceed similarly for the unweighted graph

$$\begin{aligned}&\mathbb{P}((Z_{ijk})_k|s, G, \theta, B_{ij}=1)\\ &\propto \mathbb{P}((Z_{ijk})_k|G, \theta, B_{ij}=1)\\ &\quad\times \mathbb{P}(s_{ij}|(Z_{ijk})_k, G, \theta, B_{ij}=1)\\ &\propto \mathbb{1}_{Z_{ijL}\neq 0} \prod_{k=1}^{L} \frac{p_{ijk}^{Z_{ijk}}}{Z_{ijk}!} e^{-p_{ijk}} \times e^{\sum_{k=1}^{L} p_{ijk}} \times \frac{1}{r_L}\mathbb{1}_{s_{ij}<r_L}\\ &\propto \frac{e^{\sum_{k=1}^{L} p_{ijk}} - e^{\sum_{k=1}^{L-1} p_{ijk}}}{r_L}\mathbb{1}_{s_{ij}<r_L}\\ &\quad\times \mathbb{1}_{Z_{ijL}\neq 0} \frac{1}{1-e^{-p_{ijL}}} \frac{p_{ijk}^{Z_{ijL}}}{Z_{ijL}!} e^{-p_{ijL}}\\ &\quad\times \mathbb{1}_{Z_{ij(L+1)},\dots=0} \prod_{k=1}^{L-1} \frac{p_{ijk}^{Z_{ijk}}}{Z_{ijk}!} e^{-p_{ijk}}\end{aligned}$$

B.3.3 Prediction

Here, we show how to update the missing entries we try to predict. Let us recall that for a predicted count, if it is positive, we define the slice variable as previously. However, if the count is equal to zero, then the slice variable is simply uniform over [0, 1]. Now let $L_0 = \max\{k \mid r_k > s_{ij}\}$

$$\begin{aligned}&\mathbb{P}((Z_{ijk})_k|s, G, \theta)\\ &\propto \mathbb{P}((Z_{ijk})_k|G, \theta) \times \mathbb{P}(s_{ij}|(Z_{ijk})_k, G, \theta)\\ &\propto \mathbb{1}_{Z_{ij(L_0+1)},\dots=0}\, e^{\sum_{k=1}^{L_0} p_{ijk}} \prod_{k=1}^{L_0} \frac{p_{ijk}^{Z_{ijk}}}{Z_{ijk}!} e^{-p_{ijk}}\\ &\quad\times (\mathbb{1}_{A_{ij}\neq 0}\frac{1}{r_L}\mathbb{1}_{s_{ij}<r_L} + \mathbb{1}_{A_{ij}=0})\end{aligned}$$

Now let

$$\begin{aligned}f((Z_{ijk})_k) = \mathbb{1}_{Z_{ij(L_0+1)},\dots=0}\, e^{\sum_{k=1}^{L_0} p_{ijk}} \prod_{k=1}^{L_0} \frac{p_{ijk}^{Z_{ijk}}}{Z_{ijk}!} e^{-p_{ijk}}\\ \times (\mathbb{1}_{A_{ij}\neq 0}\frac{1}{r_L}\mathbb{1}_{s_{ij}<r_L} + \mathbb{1}_{A_{ij}=0})\end{aligned}$$

Using B.3.2, it comes that

$$\mathbb{E}(f((Z_{ijk})_k) \mid A_{ij} \neq 0) = e^{\sum_{k=1}^{L_0} p_{ijk}} \sum_{L=1}^{L_0} \frac{e^{\sum_{k=1}^{L} p_{ijk}} - e^{\sum_{k=1}^{L-1} p_{ijk}}}{r_L}$$

Besides,

$$\mathbb{E}(f((Z_{ijk})_k) \mid A_{ij} = 0) = f(0) = 1$$

Therefore, here we proceed in two steps, first we sample the binomial $\mathbb{1}_{A_{ij}=0}$ with parameter

$$p = \frac{1}{1 + e^{\sum_{k=1}^{L_0} p_{ijk}} \sum_{L=1}^{L_0} \frac{e^{\sum_{k=1}^{L} p_{ijk}} - e^{\sum_{k=1}^{L-1} p_{ijk}}}{r_L}}$$

Then, conditioning on the event $A_{ij} \neq 0$, we use B.3.2 to proceed.

B.4 Proof for the Gibbs step 2

Here, we show how we can update the parameters $\theta = (\kappa, \sigma, \tau, \alpha, \beta)$ using a Metropolis–Hastings update. First, let us derive the posterior distribution of the hyperparameters.

We write $G = G' + \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}$ where G' is the non-observed part. And we note $\overline{G'}$ the restriction of G' to the locations which intensity is larger than $\min s$.

$$\begin{aligned}
&p(\theta|(s, \overline{G}), \widetilde{v}, \widetilde{Z}) \\
&\propto p(\theta \mid \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, s, \overline{G'}, \widetilde{Z}) \\
&\propto p(\theta, \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, s, \overline{G'}, \widetilde{Z}) \\
&\propto p(\theta, \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, \widetilde{Z})\ p(s, \overline{G'} \mid \theta, \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, \widetilde{Z})
\end{aligned}$$

Now let us derive consider the first part

$$p(\theta, \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, \widetilde{Z}) \propto p(\theta)\ p(\sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c} | \theta)\ p(\widetilde{Z} | \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c})$$

$$\begin{aligned}
&\propto p(\theta) p(\sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c} | \theta) \\
&\propto p(\theta) e^{-\Psi(n)} \prod_{c=1}^{K} \rho_{\kappa,\sigma,\tau}(\tilde{r}_c) f_{\alpha,\beta}(\tilde{v}_c)
\end{aligned}$$

Now let us consider the second part

$$\begin{aligned}
&p(s, \overline{G'} \mid \theta, \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, \widetilde{Z}) \\
&\propto p(s | \sum_{c=1}^{K} \tilde{r}_c \delta_{\tilde{v}_c}, \widetilde{Z})\ p(\overline{G'} | \theta, s) \\
&\propto p(\overline{G'} | \theta, s) \\
&\propto e^{-\Psi'(\min s, n)} \prod_{k} \rho_{\kappa,\sigma,\tau}(r'_k) f_{\alpha,\beta}(v'_k)
\end{aligned}$$

where (r'_k) and (v'_k) are, respectively, the intensities and locations of $\overline{G'}$

Let

$$\pi(\theta) = e^{-\Psi(n) - \Psi'(\min s, n)} \prod_{t=1}^{T} \mu_\theta(r_t, v_t),$$

where we are taking the product over the T atoms and jumps of $\overline{G}$ and

$$\Psi'(s, n) = \int_{r > s, v} e^{-r|v|^2} \prod_{i=1}^{n} (f(v_i) dv_i)\ \rho(r) dr.$$

The posterior satisfies $p(\theta|(s, \overline{G})) \propto p(\theta)\pi(\theta)$. With our particular choice of distribution of the CRM, the multivariate integrals are reduced to one dimensional integrals, which makes the algorithm tractable. Indeed, we find that

$$\begin{aligned}
&\Psi(n) + \Psi'(\min s, n) \\
&= \frac{\kappa}{\sigma} \int_0^{+\infty} (\tau + \varsigma^2)^\sigma [\frac{\sigma}{\Gamma(1-\sigma)} \Gamma(-\sigma, (\tau + \varsigma^2) \min s) + 1] \\
&\quad f_{n\alpha,\beta}(\varsigma) d\varsigma - \frac{\kappa \tau^\sigma}{\sigma}
\end{aligned}$$

$$\Psi(n) = \begin{cases} \frac{\kappa}{\sigma} \int_0^{+\infty} (\tau + \varsigma^2)^\sigma f_{n\alpha,\beta}(\varsigma) d\varsigma - \frac{\kappa \tau^\sigma}{\sigma}, & \text{if } \sigma > 0 \text{ or } \sigma < 0 \\ \kappa \int_0^{+\infty} \log(\tau + \varsigma^2) f_{n\alpha,\beta}(\varsigma) d\varsigma - \kappa \log \tau, & \text{if } \sigma = 0 \end{cases}$$

and

$$\Psi'(\min s, n) = \begin{cases} \frac{\kappa}{\Gamma(1-\sigma)} \int_0^{+\infty} (\tau + \varsigma^2)^\sigma \Gamma(-\sigma, (\tau + \varsigma^2) \min s) f_{n\alpha,\beta}(\varsigma) d\varsigma, & \text{if } \sigma > 0 \text{ or } \sigma < 0 \\ \kappa \int_0^{+\infty} \int_{(\tau+\varsigma^2) \min s}^{+\infty} \frac{e^{-r}}{r} dr f_{n\alpha,\beta}(\varsigma) d\varsigma, & \text{if } \sigma = 0 \end{cases}$$

We use the following priors:

$$\begin{aligned}
1 - 2\sigma &\sim Gamma(a_\sigma, b_\sigma) \\
\kappa &\sim Gamma(a_\kappa, b_\kappa) \\
\tau &\sim Gamma(a_\tau, b_\tau) \\
\alpha &\sim Gamma(a_\alpha, b_\alpha)
\end{aligned}$$

$$\beta \sim Gamma(a_\beta, b_\beta)$$

And proposals

$$\begin{aligned}
1-2\tilde{\sigma}|\sigma &\sim Lognormal(\log(1-2\sigma), \Sigma_\sigma)\\
\tilde{\kappa}|\kappa &\sim Lognormal(\log\kappa, \Sigma_\kappa)\\
\tilde{\tau}|\tau &\sim Lognormal(\log\tau, \Sigma_\tau)\\
\tilde{\alpha}|\alpha &\sim Lognormal(\log\alpha, \Sigma_\alpha)\\
\tilde{\beta}|\beta &\sim Lognormal(\log\beta, \Sigma_\beta)
\end{aligned}$$

We find that

$$\begin{aligned}
&\log\frac{p(\tilde{\theta})q(\theta|\tilde{\theta})}{p(\theta)q(\tilde{\theta}|\theta)}\\
&= a_\sigma \log\frac{1-2\tilde{\sigma}}{1-2\sigma} + 2b_\sigma(\tilde{\sigma}-\sigma)\\
&\quad + a_\kappa \log\frac{\tilde{\kappa}}{\kappa} - b_\kappa(\tilde{\kappa}-\kappa)\\
&\quad + a_\tau \log\frac{\tilde{\tau}}{\tau} - b_\tau(\tilde{\tau}-\tau)\\
&\quad + a_\alpha \log\frac{\tilde{\alpha}}{\alpha} - b_\alpha(\tilde{\alpha}-\alpha)\\
&\quad + a_\beta \log\frac{\tilde{\beta}}{\beta} - b_\beta(\tilde{\beta}-\beta)
\end{aligned}$$

And

$$\begin{aligned}
&\log\frac{\pi(\tilde{\theta})}{\pi(\theta)}\\
&= \Psi_\theta(n) + \Psi'_\theta(\min s, n) - \Psi_{\tilde{\theta}}(n) - \Psi'_{\tilde{\theta}}(\min s, n)\\
&\quad + T\log\frac{\tilde{\kappa}}{\kappa} - T\log\frac{\Gamma(1-\tilde{\sigma})}{\Gamma(1-\sigma)} - nT\log\frac{\Gamma(\tilde{\alpha})}{\Gamma(\alpha)}\\
&\quad - (\tilde{\sigma}-\sigma)\sum_t \log r_t - (\tilde{\tau}-\tau)\sum_t r_t\\
&\quad + (\tilde{\alpha}-\alpha)\sum_{t,i}\log v_{t,i} - (\tilde{\beta}-\beta)\sum_{t,i} v_{t,i}\\
&\quad + nT(\tilde{\alpha}\log\tilde{\beta} - \alpha\log\beta)
\end{aligned}$$

B.5 Sampling from the inhomogeneous CRM

In this section, we show how we can sample from the inhomogeneous CRM G' with measure:

$$\mu'(dr, dv) = e^{-r(\sum_{i=1}^{n} v_i)^2}\rho_{\kappa,\tau,\sigma}(r)\,[\prod_{i=1}^{n} f_{\alpha,\beta}(v_i)dv_i]\,dr$$

Let us recall that $f_{\alpha,\beta}$ is the gamma pdf and $\rho_{\kappa,\tau,\sigma}$ the GGP intensity. From Section 4, we know that if we make the following change of variables $(v_1, \ldots, v_n) \mapsto (\varsigma = \sum_i v_i, \nu_1 = v_1/s, \ldots, \nu_n = v_n/s)$, we get

$$\begin{aligned}
&\mu'(dr, d\varsigma, d\nu) = e^{-r\varsigma^2}\rho_{\kappa,\tau,\sigma}(r)\; f_{n\alpha,\beta}\\
&\quad (\varsigma)d\varsigma\; dr\; \mathrm{Dir}(d\nu_1, \ldots, d\nu_n; \boldsymbol{\alpha})
\end{aligned}$$

Hence, we can sample independently (r, ς) and ν. From one hand, $(\nu_1, \ldots, \nu_n)$ is sampled from a Dirichlet distribution with parameter $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$. On the other hand, the total sum ς and the intensity r are sampled from

$$\begin{aligned}
\mu(r, \varsigma) &= e^{-r\varsigma^2}\rho_{\kappa,\tau,\sigma}(r)\; f_{n\alpha,\beta}(\varsigma)\\
&= \frac{\kappa}{\Gamma(1-\sigma)}e^{-r(\varsigma^2+\tau)}r^{-1-\sigma} f_{n\alpha,\beta}(\varsigma)
\end{aligned}$$

Now, to reduce the problem to sampling from a homogeneous CRM, let us consider the change of variable $(r, \varsigma) \mapsto (\bar{r} = r[\tau + \varsigma^2], s)$ which determinant is $\tau + \varsigma^2$. We find finally that

$$\mu(\bar{r}, \varsigma) = \frac{\kappa}{\Gamma(1-\sigma)}e^{-\bar{r}}\,\bar{r}^{-1-\sigma}\,(\varsigma^2+\tau)^\sigma f_{n\alpha,\beta}(\varsigma)$$

Besides, since $\bar{r} \geq r\tau \;\; \forall(i, j)$, we only need to sample the points such that $\bar{r} \geq \tau \min s$. Therefore, since $\tau > 0$, we sample a finite number of atoms. Then, we only keep the points such as $r > \min s$. Finally, let us notice that in our setting, even with $\sigma \leq 0$

$$\int_\varsigma (\varsigma^2+\tau)^\sigma f_{n\alpha,\beta}(\varsigma)d\varsigma = \frac{\sigma}{\kappa}\Psi(n) + \tau^\sigma$$

Therefore, we first sample the jumps from the levy measure

$$\rho(\bar{r}) = \frac{\sigma\Psi(n) + \kappa\tau^\sigma}{\Gamma(1-\sigma)}e^{-\bar{r}}\,\bar{r}^{-1-\sigma}\mathbb{1}_{\bar{r}\geq\tau\min s}$$

using adaptive thinning (Favaro and Teh 2013). Then, we sample ς with pdf $\propto (\varsigma^2+\tau)^\sigma f_{n\alpha,\beta}(\varsigma)$ using rejection sampling.

References

Adamic, L. A., Glance, N.: The political blogosphere and the 2004 US election: divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 36–43. ACM, 2005

Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic blockmodels. J. Mach. Learn. Res. **9**, 1981–2014 (2008)

Arenas, A., Danon, L., Diaz-Guilera, A., Gleiser, P.M., Guimera, R.: Community analysis in social networks. Eur. Phys. J. B **38**(2), 373–380 (2004)

Ball, B., Karrer, B., Newman, M.E.J.: Efficient and principled method for detecting communities in networks. Phys. Rev. E **84**(3), 036103 (2011)

Brix, A.: Generalized gamma measures and shot-noise Cox processes. Adv. Appl. Probab. **31**(4), 929–953 (1999)

Caron, F., Fox, E.B.: Sparse graphs using exchangeable random measures. J. R. Statist. Soc. B**79**(5),(2017)
Cemgil, A.T.: Bayesian inference for nonnegative matrix factorisation models. Comput. Intell. Neurosci. **2009**,(2009)
Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**(6), 066111 (2004)
Dunson, D.B., Herring, A.H.: Bayesian latent variable models for mixed discrete outcomes. Biostatistics **6**(1), 11–25 (2005)
Durante, D., Dunson, D.B.: Nonparametric bayes dynamic modelling of relational data. Biometrika **101**(4), 883–898 (2014)
Favaro, S., Teh, Y. W.: MCMC for normalized random measure mixture models. Statist. Sci., 335–359, (2013)
Gnedin, A., Hansen, B., Pitman, J.: Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. Probab. Surv. **4**, 146–171 (2007)
Gopalan, P., Ruiz, F. J., Ranganath, R., Bleiand, D. M.: Bayesian nonparametric Poisson factorization for recommendation systems. In: AISTATS, pp. 275–283 (2014)
Gopalan, P., Hofman, J. M., Blei, D. M.: Scalable recommendation with hierarchical Poisson factorization. In: UAI, pp. 326–335 (2015)
Griffin, J.E., Leisen, F.: Compound random measures and their use in Bayesian non-parametrics. J. R. Statist. Soc. Ser. B Statist. Methodol. **79**(2), 525–545 (2017)
Hoff, P.D.: Multiplicative latent factor models for description and prediction of social networks. Comput. Math. Org. Theory **15**(4), 261 (2009)
Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. J. Am. Statist. Assoc. **97**(460), 1090–1098 (2002)
Hougaard, P.: Survival models for heterogeneous populations derived from stable distributions. Biometrika **73**(2), 387–396 (1986)
James, L. F.: Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics. arXiv preprint arXiv:math/0205093 (2002)
James, L. F.: Poisson latent feature calculus for generalized Indian buffet processes. arXiv preprint arXiv:1411.2936 (2014)
James, L. F.: Bayesian Poisson calculus for latent feature modeling via generalized Indian buffet process priors. Ann. Statist., 45(5):2016–2045, 10 2017. URL https://doi.org/10.1214/16-AOS1517
Kalli, M., Griffin, J.E., Walker, S.G.: Slice sampling mixture models. Statist. Comput. **21**(1), 93–105 (2011)
Kingman, J.F.C.: Completely random measures. Pacific J. Math. **21**(1), 59–78 (1967)
Kingman, J.F.C.: Poisson Processes, vol. 3. Oxford University Press, Cambridge (1993)
Klymko, C.E., Gleich, D., Kolda, T. G: Using triangles to improve community detection in directed networks. arXiv preprint arXiv:1404.5874 (2014)
Lau, J.W., Green, P.J.: Bayesian model-based clustering procedures. J. Comput. Graph. Statist. **16**(3), 526–558 (2007)
Lee, D. D., Seung, H. S.: Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems, pp. 556–562 (2001)
Lijoi, A., Prünster, I.: Models beyond the Dirichlet process. Bayes. Nonparametr. **28**(80), 3 (2010)
Ma, H., Liu, C., King, I., Lyu, M. R.: Probabilistic factor models for web site recommendation. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp. 265–274. ACM (2011)
Miller, J.W., Harrison, M.T.: A simple example of dirichlet process mixture inconsistency for the number of components. In: Advances in neural information processing systems, pp. 199–206 (2013)
Paatero, P., Tapper, U.: Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics **5**(2), 111–126 (1994)
Pitman, J.: Poisson–kingman partitions. Lecture Notes-Monograph Series, pp. 1–34 (2003)
Pollard, D.: Mini empirical. Manuscript. http://www. stat. yale. edu/pollard/Books/Mini (2015)
Radicchi, Filippo, Castellano, Claudio, Cecconi, Federico, Loreto, Vittorio, Parisi, Domenico: Defining and identifying communities in networks. Proc. Natl. Acad. Sci. **101**(9), 2658–2663 (2004)
Rozemberczki, B., Davies, R., Sarkar, R., Sutton, C.:. Gemsec: Graph embedding with self clustering. ArXiv e-prints, (2018)
Stegehuis, C., Van Der Hofstad, R., Van Leeuwaarden, J.S.H.: Power-law relations in random networks with communities. Phys. Rev. E **94**(1), 012302 (2016)
Titsias, M. K.: The infinite gamma-Poisson feature model. In: Advances in Neural Information Processing Systems, pp. 1513–1520 (2008)
Virtanen, T., Cemgil, A. T., Godsill, S.: Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 1825–1828. IEEE (2008)
Walker, S.G.: Sampling the Dirichlet mixture model with slices. Commun. Statist. Simul. Comput. **36**(1), 45–54 (2007)
Zhou, M.:. Infinite edge partition models for overlapping community detection and link prediction. In: Artificial Intelligence and Statistics, pp. 1135–1143 (2015)
Zhou, M., Hannah, L., Dunson, D., Carin, L.: Beta-negative binomial process and Poisson factor analysis. In: N. D. Lawrence and M. Girolami (eds.) Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, volume 22 of Proceedings of Machine Learning Research, pp. 1462–1471, La Palma, Canary Islands, 21–23 (2012). PMLR

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.