Bayesian nonparametric models for bipartite graphs

François Caron

Andrea Teruzzi

September 5, 2022

Table of Contents

Bipartite Networks

Definition

A **bipartite graph** is a graph $g = (V, E)$, where vertices V are divided in two sets A and B and edges E can occur only between elements of two different sets.

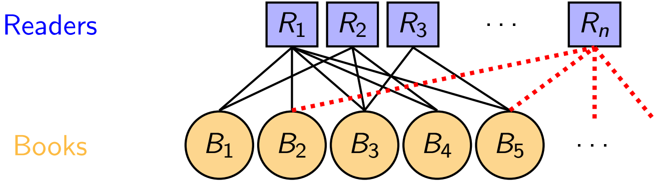
Real world examples:

- Scientists authoring papers
- Internet users posting messages on forums
- **Readers reading books**

Definition

We say **degree of a vertex** the number of edges connected to that vertex.

Bipartite Networks



Bayesian nonparametric (BNP) models for network growth:

- Parameter of interest is **infinite-dimensional** (i.e. infinite number of books)
- Bayesian nonparametric (BNP) models:
  - Indian Buffet Process (IBP), but does <u>not induce power-law behaviour</u>
  - Stable IBP, but <u>induces Poissonian distribution for the degree of readers</u>
- Flexible BNP model able to capture **power-law behaviour** for both books and readers, while retaining **computational tractability**

Table of Contents

Bayesian Model

Bipartite graph

- We represent a bipartite graph using a collection of atomic measure Z_i. For each reader $i = 1, \ldots, n$ with books $j = 1, \ldots, \infty$:

$$Z_i = \sum_{j=1}^{\infty} z_{ij} \delta_{\theta_j}$$

  where $\{\theta\} \subset \Theta$ the set of books and z_{ij} equal 1 if reader i has read book j, 0 otherwise.
- For each reader we consider the latent process V_i:

$$V_i = \sum_{j=1}^{\infty} v_{ij} \delta_{\theta_j}$$

  where v_{ij} (inversely) controls the **probability of the existence of the edge** between reader i and book j.

Bayesian Model

Latent process

- Assuming:

$$v_{ij} | w_j \sim Exp(w_j \gamma_i)$$

  - A positive **popularity parameter** w_j assigned to each book
  - A positive **interest-in-reading parameter** γ_i assigned to each reader

- Then, the probability that reader i reads book j is:

$$p(z_{ij} = 1 | w_j, \gamma_i) = 1 - exp(w_j \gamma_i)$$

For tractability issues, we consider $u_{ij} = \min(v_{ij}, 1)$ and the process U_i. Z_i can be obtained deterministically from U_i.

Bayesian Model

Book popularity parameter

Definition

Let Θ be a measurable space. A **completely random measure** (**CRM**) is a random measure G such that for any collection of disjoint measurable subsets $A_1, \ldots, A_n$ of Θ, the random masses of the subsets $G(A_1), \ldots, G(A_n)$ are independent.

G $\sim$ **CRM**$(\boldsymbol{\lambda}, \mathrm{h})$ with Levy measure:

$$\Lambda(\mathrm{d}w, \mathrm{d}\theta) = \lambda(w)h(\theta)\mathrm{d}w\mathrm{d}\theta$$

Realizations of G take the form of Poisson processes over $\{(w_j, \theta_j), j = 1, \ldots, \infty\} \subset \mathbb{R}_+ \times \Theta$:

$$G = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}$$

Bayesian Model

Book popularity parameter

An example of CRM is the **generalized gamma process** (GGP), which includes the gamma process (GP), the inverse Gaussian process (IGP) and stable process as special cases:

$$\lambda(w; \alpha, \sigma, \tau) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1} e^{-w\tau}$$

G is an **homogeneous CRM**:

- Atoms i.i.d from h (base density), independently from masses
- Masses distributed according to Poisson process over $\mathbb{R}^+$ with intensity λ (Levy intensity)

We assume:

$$\begin{cases} \int_0^\infty \lambda(w) dw = \infty \\ \int_0^\infty (1 - e^{-w}) \lambda(w) dw < \infty \end{cases} \Rightarrow \mathbf{G}(\Theta) = \sum_{j=1}^{\infty} w_j \textbf{ finite and positive}$$

Bayesian Model

Hierarchical model

Z_i **is a Poisson process**, obtained from transformations of Poisson processes.

Proposition

Z_i is marginally characterized by a Poisson process. Furthermore, **the total mass** $Z_i(\Theta) = \sum_{j=1}^{\infty} z_{ij}$, which corresponds to the total number of books read by reader i, **is finite with probability one and admits a Poisson($\psi_\lambda(\gamma_i)$) distribution**, with:

$$\psi_\lambda(\gamma_i) = \int_0^\infty (1 - e^{-\gamma_i w})\lambda(w)\mathrm{d}w$$

We can sum up the model in the following hierarchical form:

$$\begin{aligned} v_{ij} | \, G &\sim \mathrm{Exp}(w_j \gamma_i) \\ G &\sim \mathrm{CRM}(\lambda, h) \end{aligned}$$

We observe a set of edges $\{z_{ij}\}$ of a bipartite network $Z_1, \dots, Z_n$ of n reader:

- K books $\{\theta_1, \dots, \theta_K\}$
- $\mathrm{K_i} = Z_i(\Theta) = \sum_{j=1}^{\infty} z_{ij}$ **the degree of reader** i
- $\mathrm{m_j} = \sum_{i=1}^{n} Z(\{\theta_j\}) = \sum_{i=1}^{n} z_{ij}$ **the degree of book** j

Posterior distribution of the CRM given the latent process U coincides with the distribution of another CRM having a rescaled intensity and fixed observed points of discontinuity :

$$\mathbf{G} = \mathbf{G}^* + \sum_{j=1}^{K} w_j \delta_{\theta_j}$$

Bayesian Model

Posterior Characterization

- G* and $\{w_j\}$ are mutually independent with:

$$G^* \sim \text{CRM}(\lambda^*, h) \quad \text{and} \quad \lambda^*(w) = \lambda(w) \exp(-w \sum_{i=1}^{n} \gamma_i)$$

  and the masses:

$$p(w_j | \text{ rest}) \propto \lambda(w_j) w_j^{m_j} \exp(-w_j \sum_{i=1}^{n} \gamma_i U_{ij})$$

- For the GGP, G^* is still a GGP with parameters $\alpha^* = \alpha$, $\sigma^* = \sigma$ and $\tau^* = \tau + \sum_i^n \gamma_i$ and:

$$w_j | \text{ rest} \sim \text{Gamma}(m_j - \sigma, \tau + \sum_{i=1}^{n} \gamma_i u_{ij})$$

Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $\mathrm{x_{ij}} = -\log(u_{ij})$ positive latent score

Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

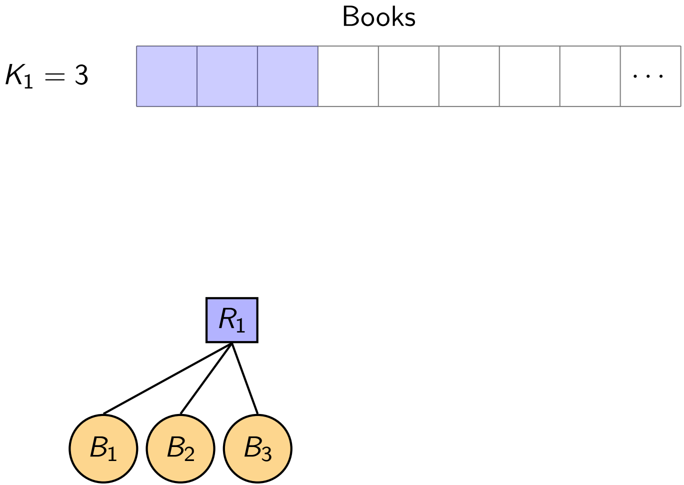
Books

K_1

$\boxed{R_1}$

Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

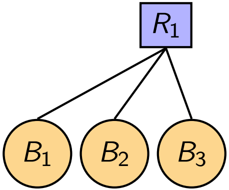


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

Books

$K_1 = 3$	18	4	14						$\cdots$

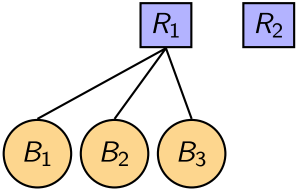


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $\mathrm{x_{ij}} = -\log(u_{ij})$ positive latent score

Books

$K_1 = 3$	18	4	14						$\cdots$
K_2									

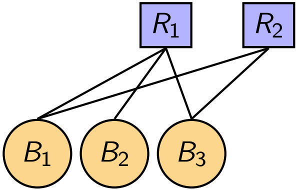


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

Books

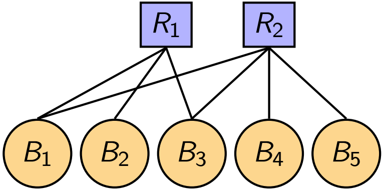
$K_1 = 3$	18	4	14						$\cdots$
K_2									



Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

	Books								
$K_1 = 3$	18	4	14						$\cdots$
$K_2^+ = 2$									$\cdots$

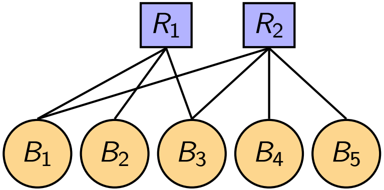


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $\mathrm{x_{ij}} = -\log(u_{ij})$ positive latent score

Books

$K_1 = 3$	18	4	14						$\cdots$
$K_2 = 4$	12	0	8	13	4				$\cdots$

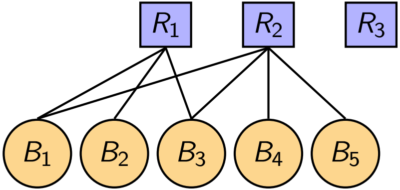


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

Books

$K_1 = 3$	18	4	14						$\cdots$
$K_2 = 4$	12	0	8	13	4				$\cdots$
K_3									

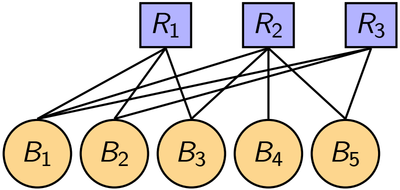


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

Books

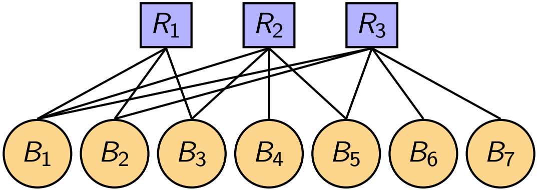
$K_1 = 3$	18	4	14						$\cdots$
$K_2 = 4$	12	0	8	13	4				$\cdots$
K_3									



Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

	Books								
$K_1 = 3$	18	4	14						$\cdots$
$K_2 = 4$	12	0	8	13	4				$\cdots$
$K_3^+ = 2$									$\cdots$

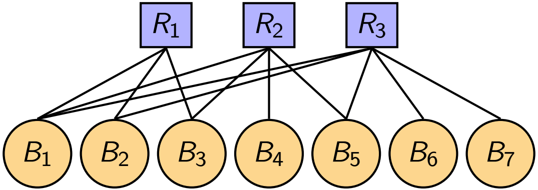


Generative Process

Distribution of $Z_n | U_1, \ldots, U_{n-1}$, with $x_{ij} = -\log(u_{ij})$ positive latent score

Books

$K_1 = 3$	18	4	14						$\cdots$
$K_2 = 4$	12	0	8	13	4				$\cdots$
$K_3 = 5$	16	10	0	0	14	9	6		$\cdots$



Gibbs sampling

We use Gibbs sampler to derive the posterior distribution of U, G| Z.

<u>For the GGP</u>:

1. For $i = 1, \dots, n$ and $j = 1, \dots, K$ set $u_{ij} = 1$ if $z_{ij} = 0$, otherwise:

$$u_{ij} | z_{ij}, w_j, \gamma_i \sim \mathrm{rExp}(\gamma_i w_j, 1)$$

2. For $j = 1, \dots, K$:

$$w_j | U, \gamma_i \sim \mathrm{Gamma}(m_j - \sigma, \tau + \sum_i^n \gamma_i u_{ij})$$

and

$$G^*(\Theta) \sim \text{Exponentially tilted stable}^1$$

[1] For general cases $G^*(\Theta)$ follows $g^*(w) \propto g(w) \exp^{-w \sum_i^n \gamma_i}$ with $g(w)$ the distribution of $G(\Theta)$

Table of Contents

Update of γ_i

1. **Parametric**: γ_i to be unknown and estimate them from the graph by assigning a prior $\gamma_i \sim \text{Gamma}(a_\gamma,\ b_\gamma)$ and update:

$$\gamma_i | \, G, U \sim \text{Gamma}\Big(a_\gamma + \sum_j^K z_{ij},\ b_\gamma + \sum_j^K w_j u_{ij} + G^*(\Theta)\Big)$$

   But $Z_i(\Theta)$ still have a (but more flexible) Poisson distribution!

2. **Nonparametric**: Let $\Gamma \sim \mathbf{CRM}(\lambda_\gamma, \mathrm{h}_\gamma)$ and a measurable space of readers $\tilde{\Theta}$, which we can represent in the form $\Gamma = \sum_{i=1}^{\infty} \gamma_i \delta_{\theta_j}$. Conditionally on $(U, w, G^*(\Theta))$, we update:

$$\Gamma = \Gamma^* + \sum_{i=1}^{n} \gamma_i \delta_{\tilde{\theta}_i}$$

   We have more of flexibility in the modelling of the distribution of the degree of readers (**power-law behavior**)!

Posterior characterization for GGP for w_i and γ_i

Let G and Γ GGP distributed with parameters (α, σ, τ) and $(\alpha_\gamma, \sigma_\gamma, \tau_\gamma)$:

- Reader update: $\Gamma = \Gamma^* + \sum_{i=1}^{n} \gamma_i \delta_{\tilde{\theta}_i}$ with:

$$\Gamma^* \sim \mathrm{CRM}(\lambda^*_\gamma, h_\gamma)$$

$$\gamma_i | \, U, G \sim \mathrm{Gamma}\Big(K_i - \sigma_\gamma, \tau_\gamma + \sum_{j=1}^{K} w_j u_{ij} + G^*(\Theta)\Big)$$

- Book update: $G = G^* + \sum_{i=1}^{K} w_i \delta_{\theta_i}$ with:

$$G^* \sim \mathrm{CRM}(\lambda^*, h)$$

$$w_i | \, U, \Gamma \sim \mathrm{Gamma}\Big(m_j - \sigma, \tau + \sum_{i=1}^{n} \gamma_i u_{ij} + \Gamma^*(\tilde{\Theta})\Big)$$

Table of Contents

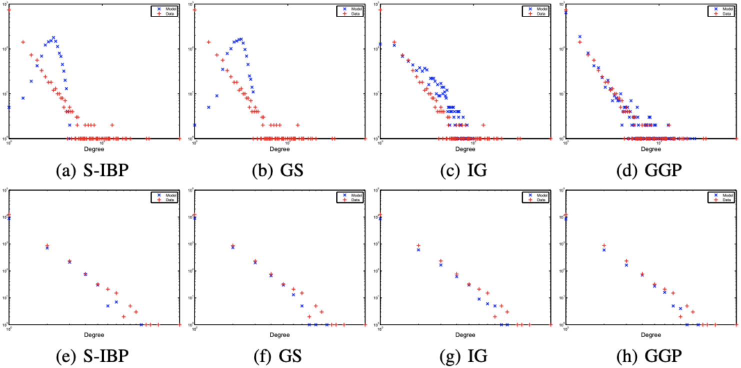
Power-law properties

For the GGP with $\sigma > 0$, we can achieve **power-law behavior of the network growth**:

- The total number of books read by n readers is $O(n^{\sigma})$
  $\rightarrow$ **Proof.** When $\gamma_i = \gamma$, the total number of books is Poisson($\psi_\lambda(n\gamma)$) distributed. Considering the GGP:
  $$\psi_\lambda(n\gamma) = \frac{\alpha}{\sigma}((n\gamma + \sigma)^{\sigma} - \tau^{\sigma})$$
  which for large n, is of order n^{σ}.

Similar results are achievable also with an S-IBP for the degree distribution of books, but not for readers for which it will always be Poisson!

Real world example – Book-crossing community network



(a) S-IBP (b) GS (c) IG (d) GGP

(e) S-IBP (f) GS (g) IG (h) GGP

Figure 1: Degree distribution for readers (a-d) and books (e-h) with 4 models: a stable Indian Buffet Process (S-IBP); our model with $\gamma_i = \gamma$ and flat prior assigned (GS); our model with $\gamma_i \sim$ Gamma (a_γ, a_γ) and flat prior assigned to the parameters (IG); our model with GGP prior for γ_i (GGP). Data are presented in red and samples from the models in blue.

Bibliography I

F. Caron.
Bayesian nonparametric models for bipartite graphs.
In *NIPS*, 2012.

E. Çınlar.
Probability and Stochastics.
Graduate Texts in Mathematics. Springer New York, 2011.

L. F. James, A. Lijoi, and I. Pruenster.
Posterior analysis for normalized random measures with independent increments.
Scandinavian Journal of Statistics, 36(1):76–97, 2009.

J. F. C. Kingman.
Poisson Processes.
Oxford Studies in Probability. Oxford University Press, 1993.