

Bayesian nonparametric models for bipartite graphs

François Caron

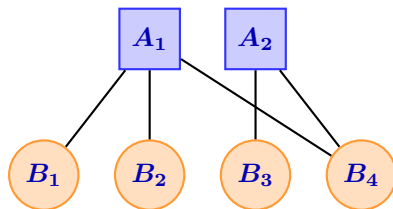
Department of Statistics, Oxford

Statistics Colloquium, Harvard University

November 11, 2013

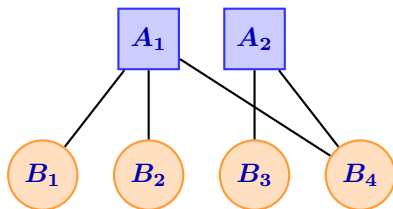


Bipartite networks



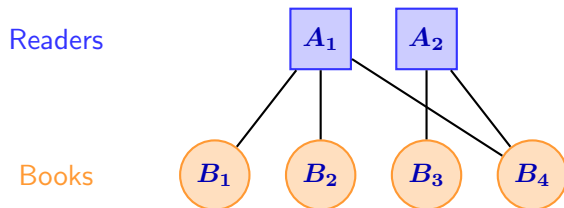
- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks



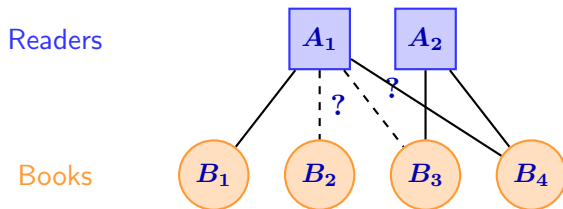
- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks



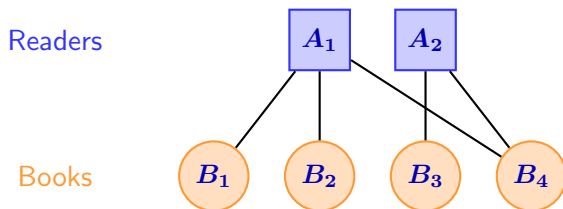
- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks



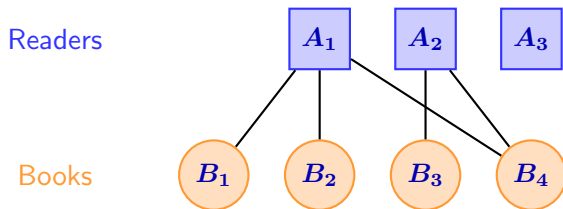
- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks



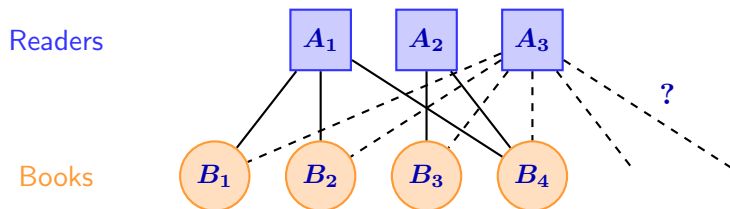
- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks



- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Bipartite networks

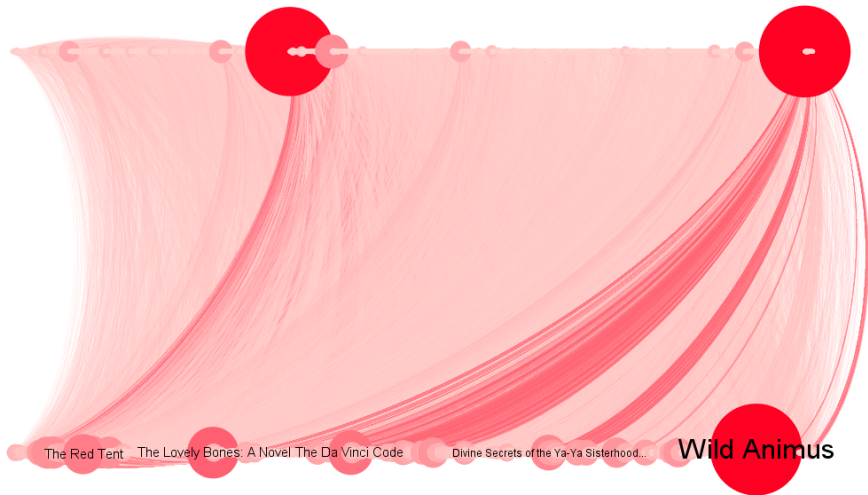


- ▶ Scientists authoring papers
- ▶ Readers reading books
- ▶ Internet users posting messages on forums
- ▶ Customers buying items
- ▶ Objects sharing a set of features

Book-crossing community network

5 000 readers, 36 000 books, 50 000 edges

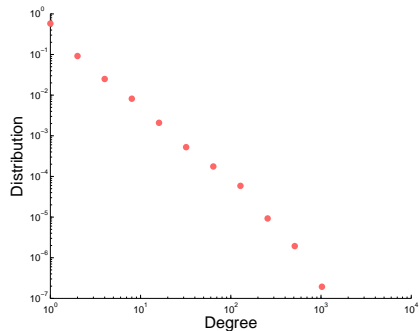
Readers



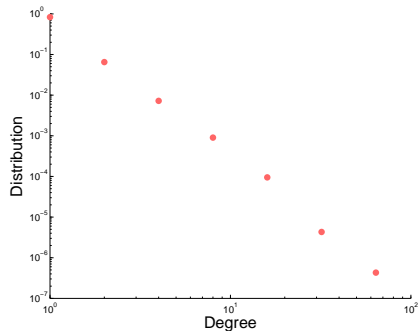
Books

Book-crossing community network

Degree distributions on log-log scale



(a) Readers



(b) Books

Statistical network models

- ▶ Statistics literature
 - ▶ Exponential random graph, stochastic block-models, Rasch models, etc
 - ▶ Do not capture power-law behavior
 - ▶ Inference do not scale well with the number of nodes
- ▶ Physics literature
 - ▶ Preferential attachment
 - ▶ Lacks interpretable parameters, non-exchangeability

Bayesian nonparametrics

- ▶ Parameter of interest is **infinite-dimensional**
- ▶ Allows the complexity of the model to adapt to the data
 - ▶ Dirichlet Process Mixtures: Clustering/density estimation with unknown number of modes
- ▶ Attractive power-law properties
 - ▶ Language modeling, image segmentation

[Teh, 2006; Sudderth and Jordan, 2008; Blunsom and Cohn, 2011]

BNP for networks

- ▶ Models with some latent structure (e.g. infinite relational model)
 - ▶ Number of nodes is fixed and dimension of the latent structure unknown
- ▶ Here: Infinite number of nodes
- ▶ (stable) Beta-Bernoulli/Indian Buffet Process
 - ▶ Can capture power-law degree distributions for books
 - ▶ Poisson degree distribution for readers

Bipartite networks

Aims

- ▶ **Bayesian nonparametric** model for bipartite networks with a potentially **infinite** number of nodes of each type
- ▶ Each node is modelled using a positive rating parameter that represents its ability to connect to other nodes
- ▶ Captures **power-law** behavior
- ▶ Simple generative model for **network growth**
- ▶ Develop efficient computational procedure for posterior simulation.

Hierarchical model

- ▶ Represent a bipartite network by a collection of atomic measures Z_i , $i = 1, 2, \dots$ such that

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

- ▶ $z_{ij} = 1$ if reader i has read book j , 0 otherwise
- ▶ $\{\theta_j\}$ is the set of books
- ▶ Each book j is assigned a positive “popularity” parameter w_j
- ▶ Each reader i is assigned a positive “interest in reading” parameter γ_i
- ▶ The probability that reader i reads book j is

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

Hierarchical model

- ▶ Represent a bipartite network by a collection of atomic measures Z_i , $i = 1, 2, \dots$ such that

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

- ▶ $z_{ij} = 1$ if reader i has read book j , 0 otherwise
- ▶ $\{\theta_j\}$ is the set of books
- ▶ Each book j is assigned a positive “popularity” parameter w_j
- ▶ Each reader i is assigned a positive “interest in reading” parameter γ_i
- ▶ The probability that reader i reads book j is

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

Hierarchical model

- ▶ Represent a bipartite network by a collection of atomic measures Z_i , $i = 1, 2, \dots$ such that

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

- ▶ $z_{ij} = 1$ if reader i has read book j , 0 otherwise
 - ▶ $\{\theta_j\}$ is the set of books
- ▶ Each book j is assigned a positive “popularity” parameter w_j
- ▶ Each reader i is assigned a positive “interest in reading” parameter γ_i
- ▶ The probability that reader i reads book j is

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

Hierarchical model

- ▶ Represent a bipartite network by a collection of atomic measures Z_i , $i = 1, 2, \dots$ such that

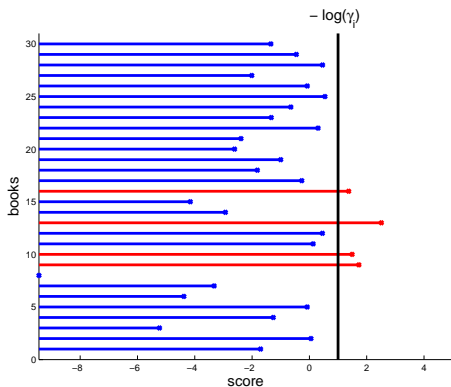
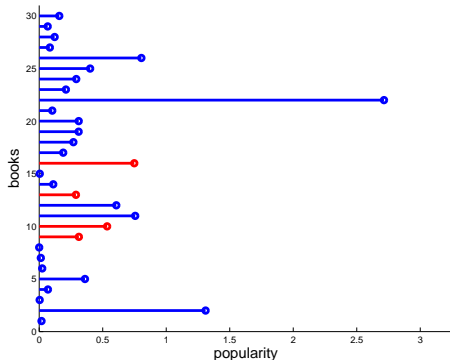
$$Z_i = \sum_{j=1}^{\infty} z_{ij} \delta_{\theta_j}$$

- ▶ $z_{ij} = 1$ if reader i has read book j , 0 otherwise
 - ▶ $\{\theta_j\}$ is the set of books
- ▶ Each book j is assigned a positive “popularity” parameter w_j
- ▶ Each reader i is assigned a positive “interest in reading” parameter γ_i
- ▶ The probability that reader i reads book j is

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

Data Augmentation

- ▶ Latent variable formulation
 - ▶ Latent scores $s_{ij} \sim \text{Gumbel}(\log(w_j), 1)$
 - ▶ All books with a score **above** $-\log(\gamma_i)$ are retained, others are discarded



Model for the book popularity parameters

- ▶ Random atomic measure

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

- ▶ Construction: two-dimensional Poisson process $N = \{w_j, \theta_j\}_{j=1, \dots}$
- ▶ **Completely Random Measure** $G \sim \text{CRM}(\lambda, h)$ characterized by a Lévy measure $\lambda(w)h(\theta)dw d\theta$

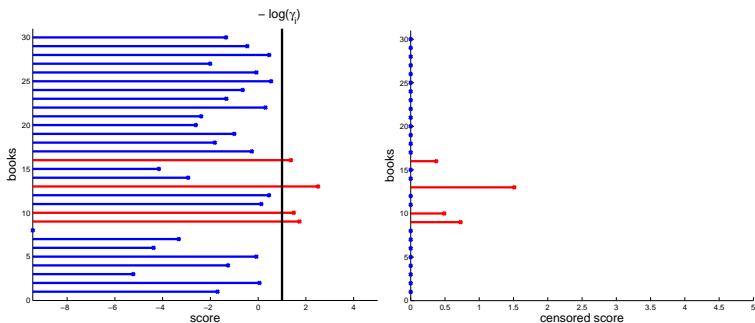
$$\int_0^{\infty} (1 - e^{-w}) \lambda(w) dw < \infty \quad \Rightarrow \quad \text{finite total } \sum_{j=1}^{\infty} z_{ij}.$$

Posterior characterization

- ▶ Observed bipartite network Z_1, \dots, Z_n
- ▶ n readers and K books with degree at least one
- ▶ **Cannot** derive directly the conditional of G given Z_1, \dots, Z_n nor the predictive of Z_{n+1} given Z_1, \dots, Z_n
- ▶ Let

$$X_i = \sum_{j=1}^{\infty} x_{ij} \delta_{\theta_j}$$

where $x_{ij} = \max(0, s_{ij} + \log(\gamma_i)) \geq 0$ are latent positive scores.



Posterior Characterization

The conditional distribution of G given X_1, \dots, X_n can be expressed as

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

where G^* and (w_j) are mutually independent with

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

and the masses are

$$P(w_j | \text{other}) \propto \lambda(w_j) w_j^{m_j} \exp\left(-w_j \sum_{i=1}^n \gamma_i e^{-x_{ij}}\right)$$

Characterization related to that for normalized random measures

[Prünster, 2002, James, 2002, James et al., 2009]

Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

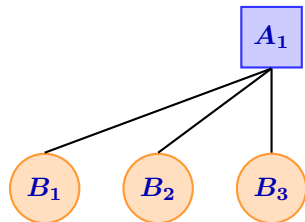
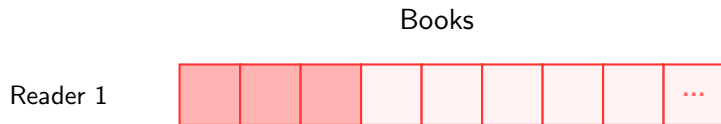
Books

Reader 1



Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

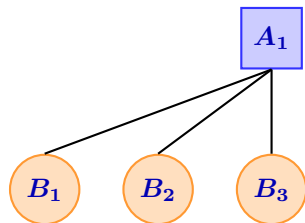


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1



Generative Process for network growth

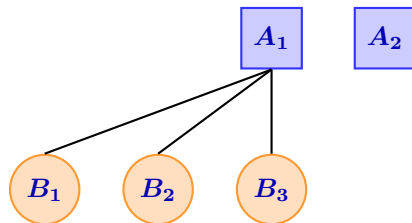
Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1



Reader 2

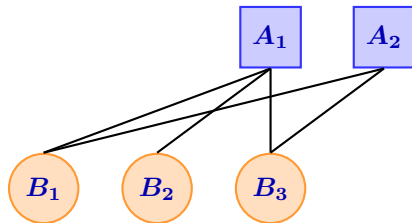


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14						...
Reader 2									

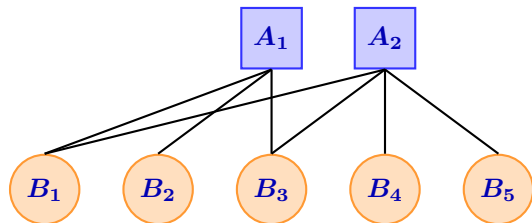


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14					...
Reader 2								...

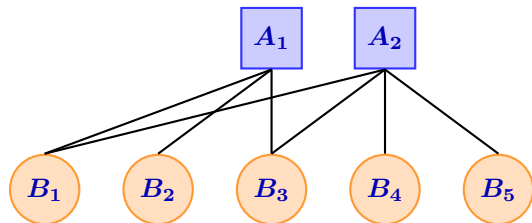


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14					...
Reader 2	12	0	8	13	4			...

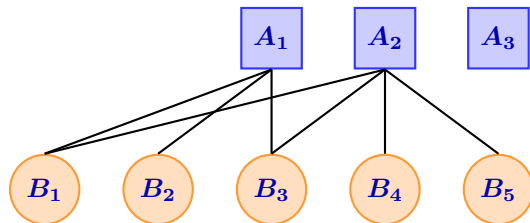


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14					...
Reader 2	12	0	8	13	4			...
Reader 3								

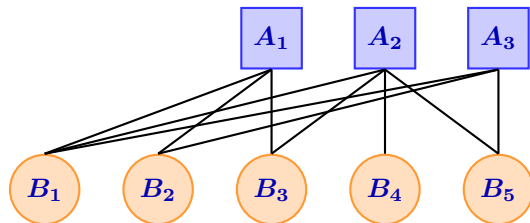


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14						...
Reader 2	12	0	8	13	4				...
Reader 3									

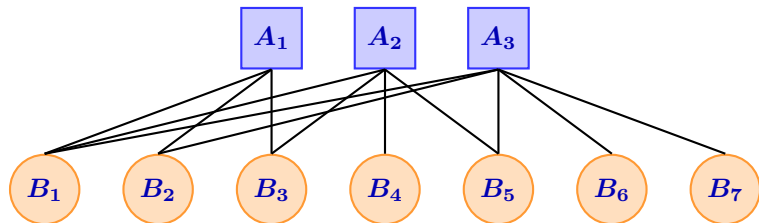


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

Books

Reader 1	18	4	14					...
Reader 2	12	0	8	13	4			...
Reader 3								...

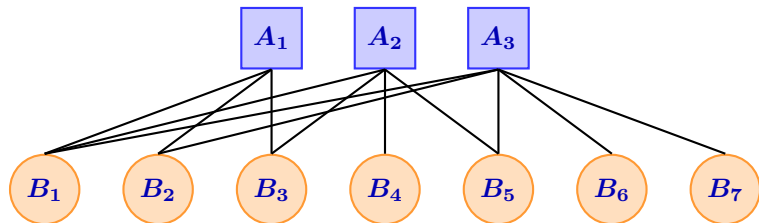


Generative Process for network growth

Predictive distribution of Z_{n+1} given the latent process X_1, \dots, X_n

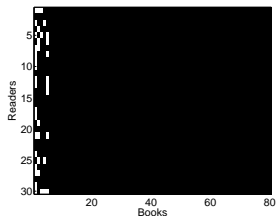
Books

Reader 1	18	4	14					...
Reader 2	12	0	8	13	4			...
Reader 3	16	10	0	0	14	9	6	...

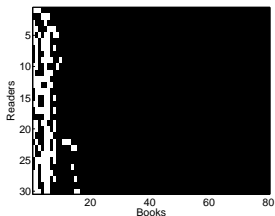


Prior Draws

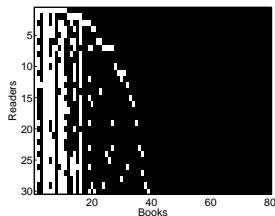
Generalized Gamma process with $\lambda(w) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1} e^{-\tau w}$, $\tau = 1$, $\gamma_i = 2$.



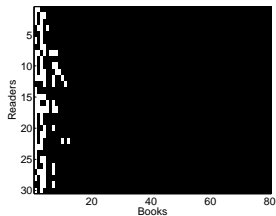
(c) $\alpha = 1, \sigma = 0$



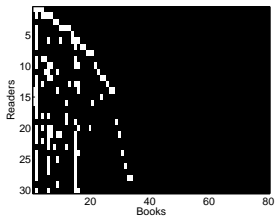
(d) $\alpha = 5, \sigma = 0$



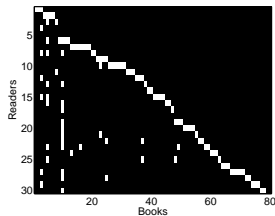
(e) $\alpha = 10, \sigma = 0$



(f) $\alpha = 2, \sigma = 0.1$



(g) $\alpha = 2, \sigma = 0.5$

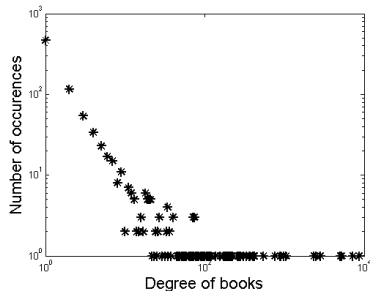
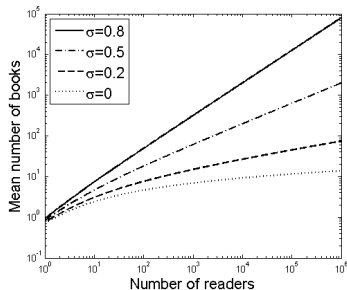


(h) $\alpha = 2, \sigma = 0.9$

[Brix, 1999, Lijoi et al., 2007]

Properties of the model

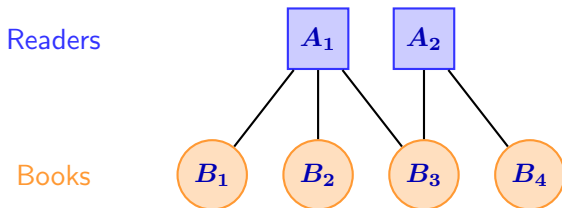
- ▶ **Power-law behavior** for the generalized gamma process with $\sigma > 0$
 - ▶ The total number of books read by n readers is $O(n^\sigma)$
 - ▶ Asympt., the proportion of books read by m readers is $O(m^{-1-\sigma})$



- ▶ (stable) Beta-Bernoulli/Indian Buffet process as a special case when

$$\lambda(w) = \frac{\alpha \Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(\sigma+c)} \gamma (1 - e^{-\gamma w})^{-\sigma-1} e^{-\gamma w(c+\sigma)}$$

Bayesian Inference via Gibbs Sampling



- ▶ Popularity parameters w_j of **observed books**.
- ▶ Sum w_* of popularity parameters of **unobserved books**.
- ▶ Latent scores x_{ij} associated to **observed edges**.
- ▶ Posterior distribution $P(\{w_j\}, w_*, \{x_{ij}\} | Z_1, \dots, Z_n)$

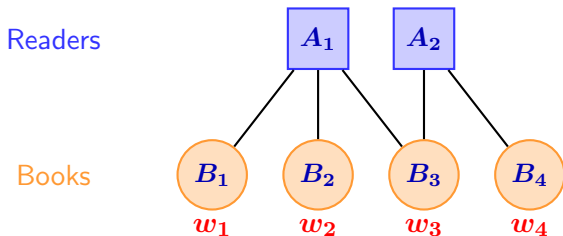
Gibbs sampler for the GGP

$x_{ij} | \text{rest} \sim \text{Truncated Gumbel}$

$w_j | \text{rest} \sim \text{Gamma}$

$w_* | \text{rest} \sim \text{Exponentially tilted stable}$

Bayesian Inference via Gibbs Sampling



- ▶ Popularity parameters w_j of **observed books**.
- ▶ Sum w_* of popularity parameters of **unobserved books**.
- ▶ Latent scores x_{ij} associated to **observed edges**.
- ▶ Posterior distribution $P(\{w_j\}, w_*, \{x_{ij}\} | Z_1, \dots, Z_n)$

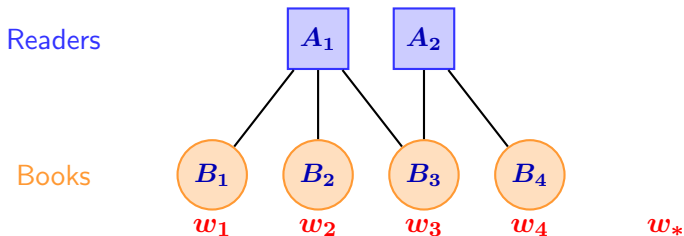
Gibbs sampler for the GGP

$x_{ij} | \text{rest} \sim$ Truncated Gumbel

$w_j | \text{rest} \sim$ Gamma

$w_* | \text{rest} \sim$ Exponentially tilted stable

Bayesian Inference via Gibbs Sampling



- ▶ Popularity parameters w_j of **observed books**.
- ▶ Sum w_* of popularity parameters of **unobserved books**.
- ▶ Latent scores x_{ij} associated to **observed edges**.
- ▶ Posterior distribution $P(\{w_j\}, w_*, \{x_{ij}\} | Z_1, \dots, Z_n)$

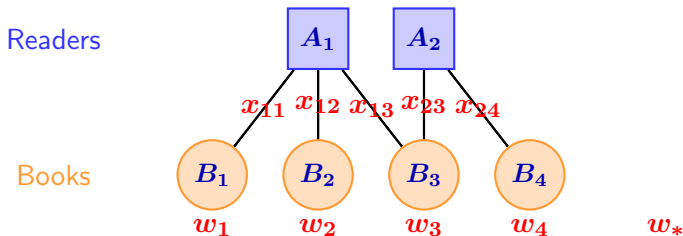
Gibbs sampler for the GGP

$x_{ij} | \text{rest} \sim$ Truncated Gumbel

$w_j | \text{rest} \sim$ Gamma

$w_* | \text{rest} \sim$ Exponentially tilted stable

Bayesian Inference via Gibbs Sampling



- ▶ Popularity parameters w_j of **observed books**.
- ▶ Sum w_* of popularity parameters of **unobserved books**.
- ▶ Latent scores x_{ij} associated to **observed edges**.
- ▶ Posterior distribution $P(\{w_j\}, w_*, \{x_{ij}\} | Z_1, \dots, Z_n)$

Gibbs sampler for the GGP

$x_{ij} | \text{rest} \sim$ Truncated Gumbel

$w_j | \text{rest} \sim$ Gamma

$w_* | \text{rest} \sim$ Exponentially tilted stable

Model for the “interest in reading” parameters

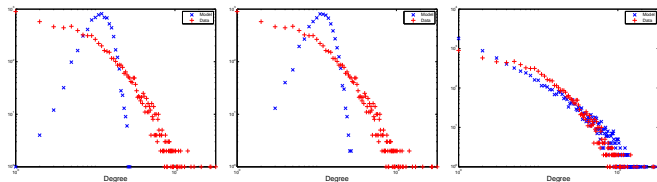
- ▶ Still Poisson degree distribution for readers
- ▶ **Parametric:** γ_i are indep. and identically distributed from a gamma distribution
- ▶ **Nonparametric:** γ_i are the points of a random atomic measure Γ
- ▶ Gibbs sampler can be derived in the same way as for books

Application

- ▶ Evaluate the fit of three models
 - ▶ Stable Indian Buffet Process
 - ▶ Proposed model where G follows a **Generalized Gamma process** of unknown parameters $(\alpha_w, \sigma_w, \tau_w)$
 - ▶ with shared and unknown $\gamma_i = \gamma$
 - ▶ with nonparametric prior where Γ follows a generalized gamma process of unknown parameters $(\alpha_\gamma, \tau_\gamma, \sigma_\gamma)$

Application: IMDB Movie Actor network

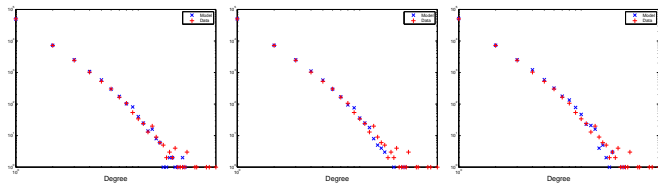
280 000 movies, 178 000 actors, 341 000 edges



(a) S-IBP

(b) GS

(c) GGP



(d) S-IBP

(e) GS

(f) GGP

Figure: Degree distributions for movies (a-d) and actors (e-h) for the IMDB movie-actor dataset with three different models. Data are represented by red plus and samples from the model by blue crosses.

Application: Book-crossing community network

5 000 readers, 36 000 books, 50 000 edges

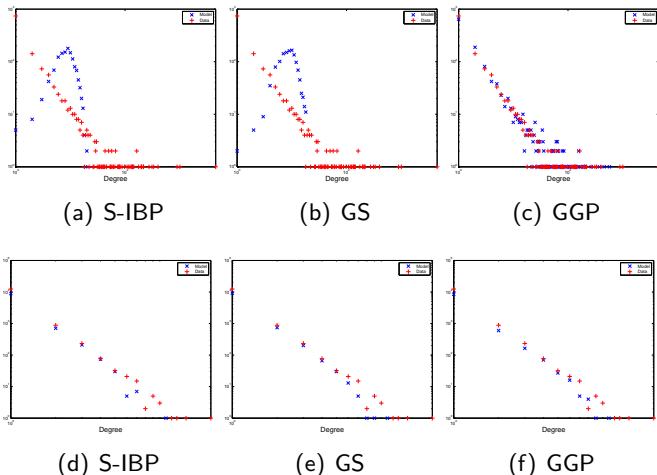
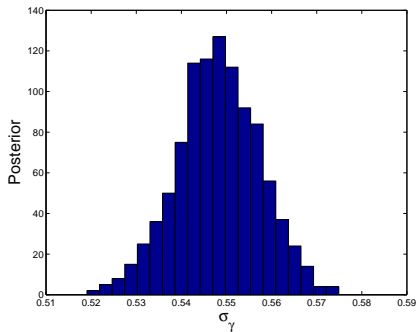


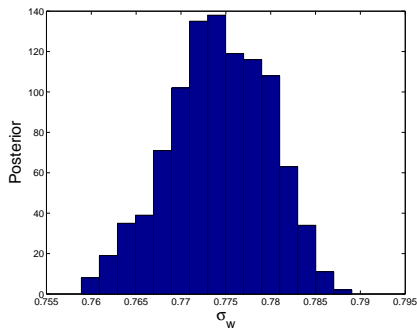
Figure: Degree distributions for readers (a-d) and books (e-h) for the book crossing dataset with three different models. Data are represented by red plus and samples from the model by blue crosses.

Application: Book-crossing community network

5 000 readers, 36 000 books, 50 000 edges



(a) σ_γ (Readers)



(b) σ_w (Books)

Figure: Posterior distributions of the power-law parameters σ_γ and σ_w

Application

- ▶ Log-likelihood on test dataset

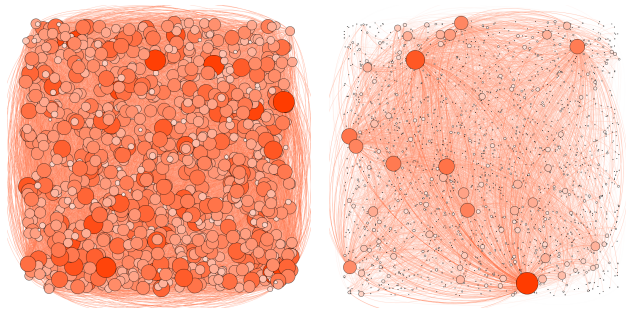
Dataset	S-IBP	SG	GGP
Board	9.82(29.8)	8.3(30.8)	-68.6 (31.9)
Forum	-6.7e3	-6.7e3	-5.6e3
Books	83.1	214	4.4e4
Citations	-3.7e4	-3.7e4	-3.4e4
Movielens100k	-6.7e4	-6.7e4	-5.5e4
IMDB	-1.5e5	-1.5e5	-1.1e5

Summary

- ▶ **Bayesian nonparametric** model for bipartite networks with a potentially **infinite** number of nodes
- ▶ Captures **power-law** behavior
- ▶ Simple generative model for **network growth**
- ▶ Simple computational procedure for posterior simulation.
- ▶ Displays a **good fit** on a variety of social networks

Future work

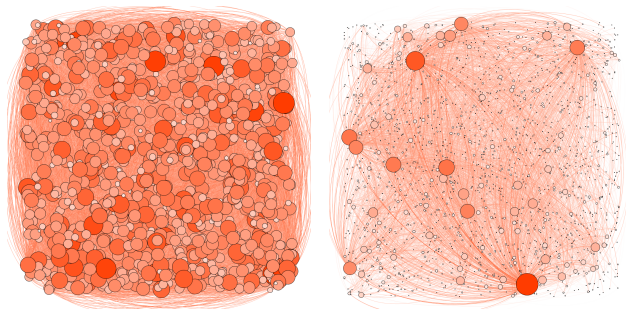
- ▶ BNP model for general (non-bipartite) networks



- ▶ BNP (dynamic) recommender systems
- ▶ Latent factorial models and dictionary learning

Future work

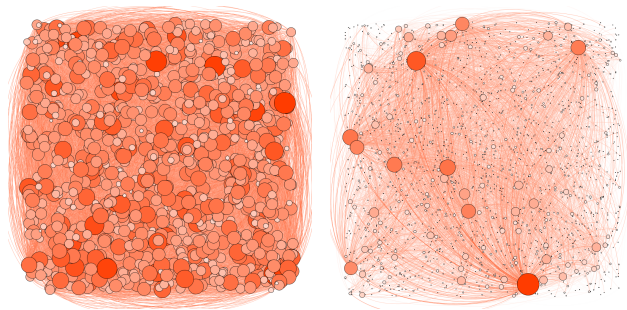
- ▶ BNP model for general (non-bipartite) networks



- ▶ BNP (dynamic) recommender systems
- ▶ Latent factorial models and dictionary learning


Future work


- ▶ BNP model for general (non-bipartite) networks





- ▶ BNP (dynamic) recommender systems
- ▶ Latent factorial models and dictionary learning


Bibliography I


 Brix, A. (1999).
Generalized gamma measures and shot-noise Cox processes.
Advances in Applied Probability, 31(4):929–953.

 Devroye, L. (2009).
Random variate generation for exponentially and polynomially tilted stable distributions.
ACM Transactions on Modeling and Computer Simulation (TOMACS), 19(4):18.

 Griffiths, T. and Ghahramani, Z. (2005).
Infinite latent feature models and the Indian buffet process.
In *NIPS*.

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

 James, L. F. (2002).
Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics.
arXiv preprint math/0205093.

 Kingman, J. (1967).
Completely random measures.
Pacific Journal of Mathematics, 21(1):59–78.

Bibliography II



Lijoi, A., Mena, R. H., and Prünster, I. (2007).
Controlling the reinforcement in bayesian non-parametric mixture models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology),
69(4):715–740.



Lijoi, A. and Prünster, I. (2010).
Models beyond the Dirichlet process.
In N. L. Hjort, C. Holmes, P. M. S. G. W., editor, *Bayesian Nonparametrics*. Cambridge
University Press.



Prünster, I. (2002).
*Random probability measures derived from increasing additive processes and their
application to Bayesian statistics*.
PhD thesis, University of Pavia.



Regazzini, E., Lijoi, A., and Prünster, I. (2003).
Distributional results for means of normalized random measures with independent
increments.
The Annals of Statistics, 31(2):560–585.



Teh, Y. and Görür, D. (2009).
Indian buffet processes with power-law behavior.
In *Neural Information Processing Systems (NIPS'2009)*.