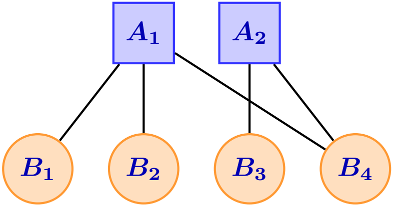
Bayesian Nonparametric Models for Bipartite Graphs

François Caron

INRIA

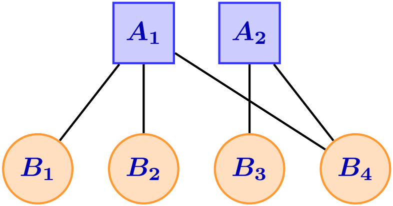
March 20, 2013
Workshop IHES

Bipartite networks



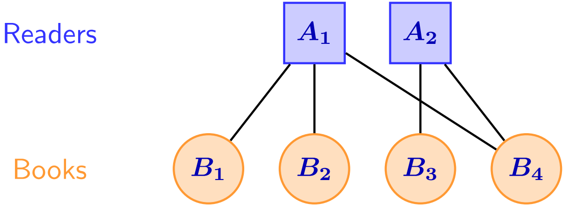
- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



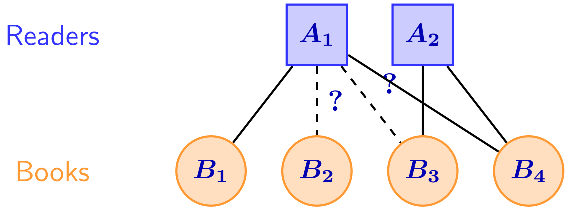
- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



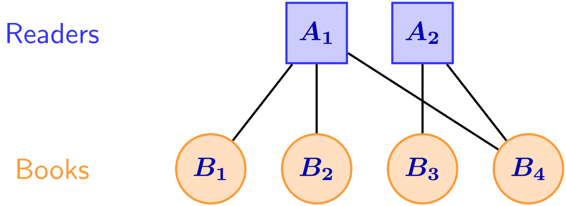
- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



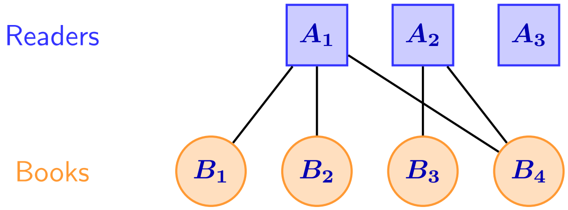
- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



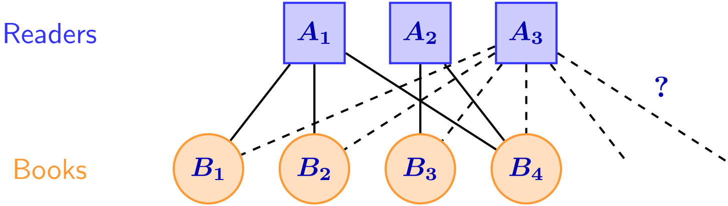
- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

Bipartite networks



- Scientists co-authoring the same paper
- Readers reading the same book
- Internet users posting a message on the same forum
- Customers buying the same item

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, $i = 1, 2, \ldots$ 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, $i = 1, 2, \ldots$ such that

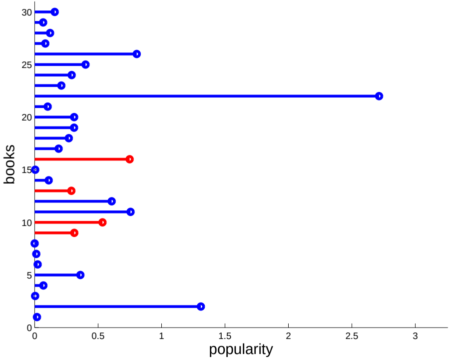
$$Z_i = \sum_{j=1}^{\infty} z_{ij} \delta_{\theta_j}$$

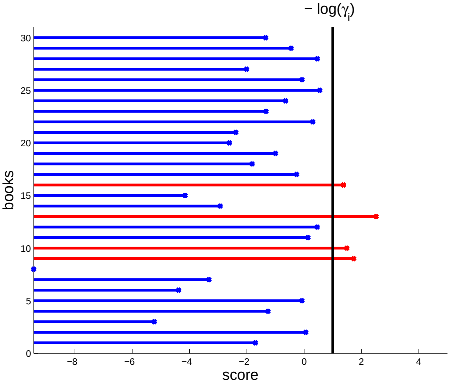
  - $z_{ij} = 1$ if reader i has read book j, 0 otherwise
  - $\{\theta_j\}$ is the set of books
- Each book j is assigned a positive "popularity" parameter w_j
- Each reader i is assigned a positive "interest in reading" parameter γ_i
- The probability that reader i reads book j is

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

Data Augmentation

- Latent variable formulation
  - Latent scores $s_{ij} \sim \mathbf{Gumbel}(\log(w_j), 1)$
  - 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 $G \sim \mathrm{CRM}(\lambda, h)$ characterized by a Lévy intensity $\lambda(w)$
- Conditions on Lévy intensity:

$$\int_0^\infty \lambda(w)dw = \infty \qquad \int_0^\infty (1 - e^{-w})\lambda(w)dw < \infty$$

$\Rightarrow$ infinitely many books

$\Rightarrow$ finite total $\sum_{j=1}^{\infty} w_j$

$\Rightarrow$ finite total $\sum_{j=1}^{\infty} z_{ij}$

[Kingman, 1967]

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 $G \sim \mathrm{CRM}(\lambda, h)$ characterized by a Lévy intensity $\lambda(w)$
- Conditions on Lévy intensity:

$$\int_0^\infty \lambda(w)dw = \infty \qquad \int_0^\infty (1 - e^{-w})\lambda(w)dw < \infty$$

$\Rightarrow$ infinitely many books

$\Rightarrow$ finite total $\sum_{j=1}^{\infty} w_j$

$\Rightarrow$ finite total $\sum_{j=1}^{\infty} z_{ij}$

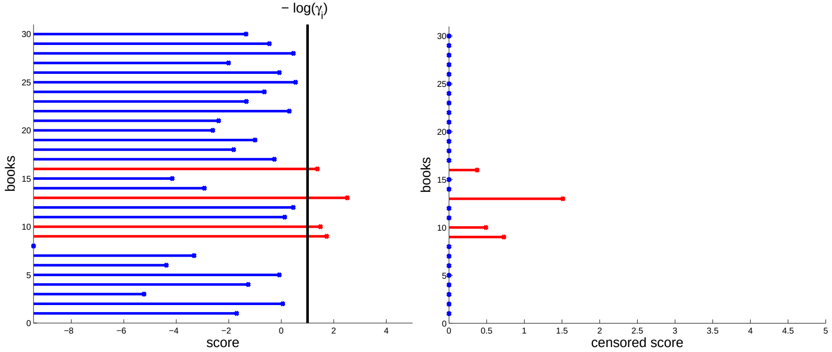
[Kingman, 1967]

Posterior characterization

- Observed bipartite network $Z_1, \dots, Z_n$
- Cannot derive directly 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

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

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 ranked data [Caron and Teh, 2012] and normalized random measures [James et al., 2009].

Generative Process for network growth

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

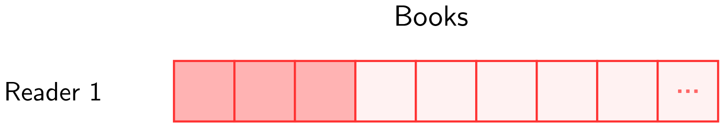
Books

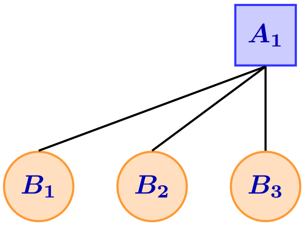
Reader 1

$\boxed{A_1}$

Generative Process for network growth

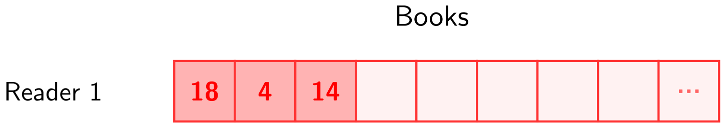
Predictive distribution of Z_{n+1} given the latent process $X_1, \ldots, X_n$

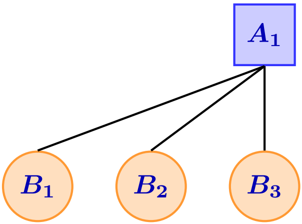




Generative Process for network growth

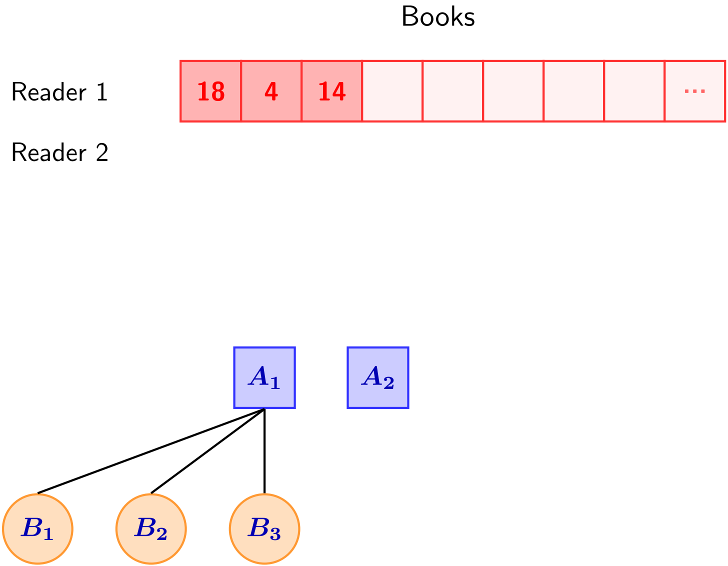
Predictive distribution of Z_{n+1} given the latent process $X_1, \ldots, X_n$





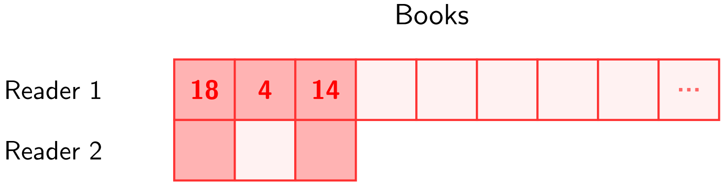
Generative Process for network growth

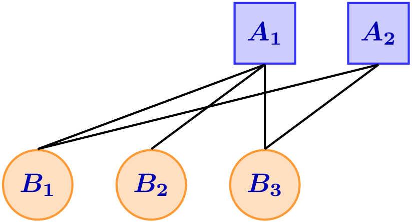
Predictive distribution of Z_{n+1} given the latent process $X_1, \ldots, X_n$



Generative Process for network growth

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



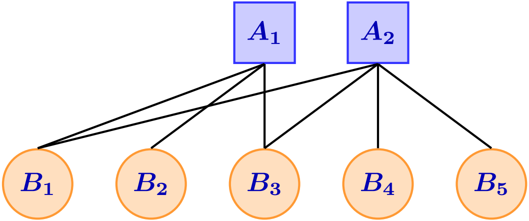


Generative Process for network growth

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

Books

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

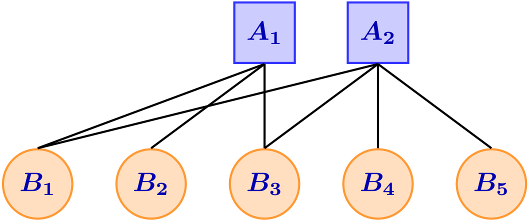


Generative Process for network growth

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

Books

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

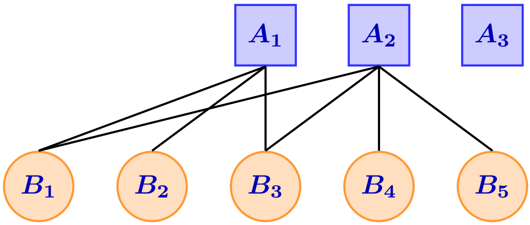


Generative Process for network growth

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

Books

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

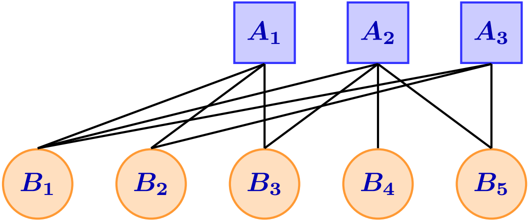


Generative Process for network growth

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

Books

Reader 1	**18**	**4**	**14**						⋯
Reader 2	**12**	**0**	**8**	**13**	**4**				⋯
Reader 3									

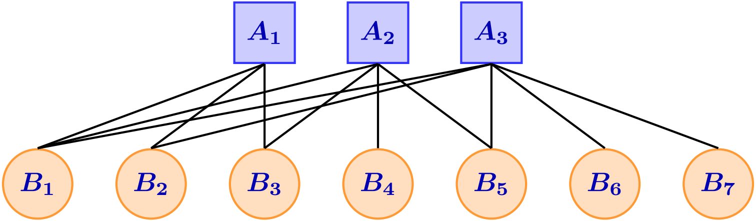


Generative Process for network growth

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

Books

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

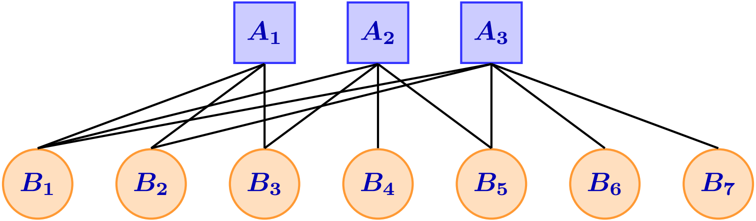


Generative Process for network growth

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

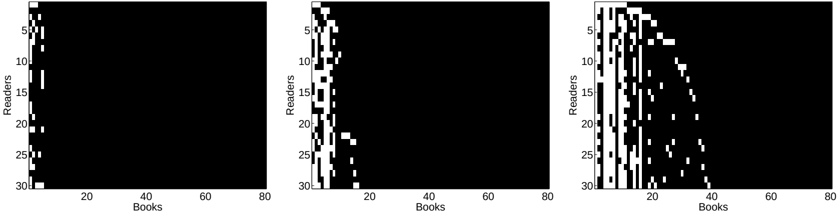
Books

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

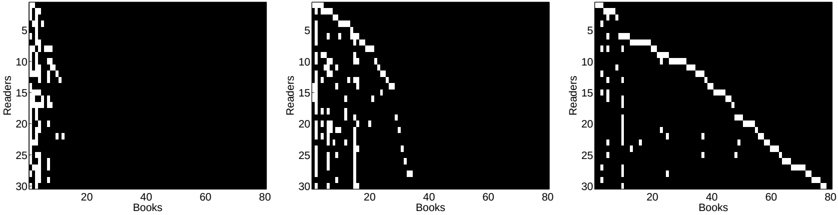


Prior Draws

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



(a) $\alpha = 1, \sigma = 0$ (b) $\alpha = 5, \sigma = 0$ (c) $\alpha = 10, \sigma = 0$

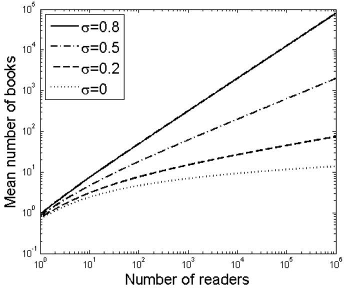


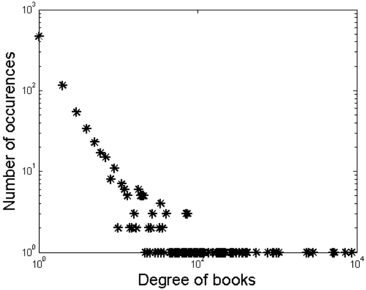
(d) $\alpha = 2, \sigma = 0.1$ (e) $\alpha = 2, \sigma = 0.5$ (f) $\alpha = 2, \sigma = 0.9$

[Brix, 1999]

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 $O(m^{-1-\sigma})$





Bayesian Inference via Gibbs Sampling

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

Gibbs sampler for the GGP

$$x_{ij}|\text{rest} \sim \text{Truncated Gumbel}$$
$$w_j|\text{rest} \sim \text{Gamma}$$
$$w_*|\text{rest} \sim \text{Exponentially tilted stable}$$

[Devroye, 2009]

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

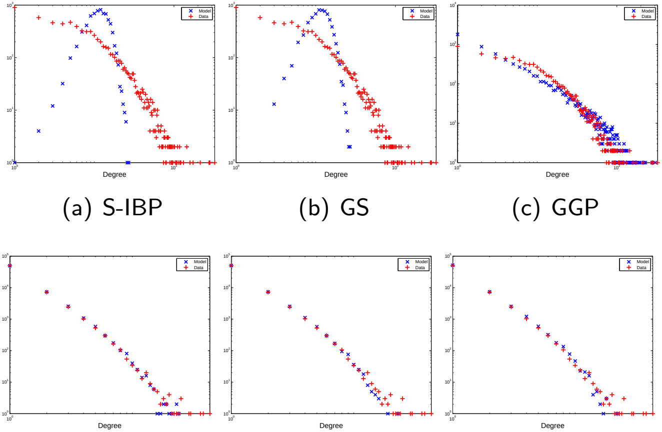
Application

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

[Teh and Görür, 2009, Griffiths and Ghahramani, 2005]

Application: IMDB Movie Actor network

280 000 movies, 178 000 actors, 341 000 edges



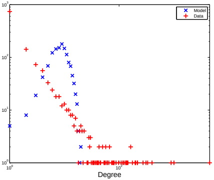
(a) S-IBP (b) GS (c) GGP

(d) S-IBP (e) GS (f) GGP

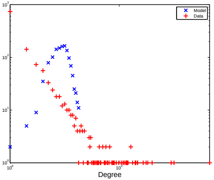
Figure: Degree distributions for movies (a-d) and actors (e-h) for the IMDB movie-actor dataset with three different models. Data are represented by red plus and samples from the model by blue crosses.

Application: Book-crossing community network

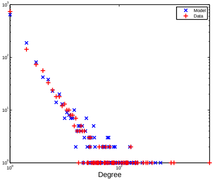
5 000 readers, 36 000 books, 50 000 edges



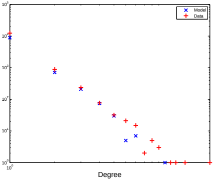
(a) S-IBP



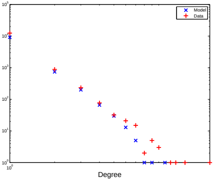
(b) GS



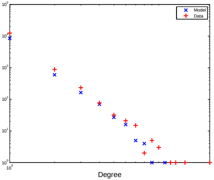
(c) GGP



(d) S-IBP



(e) GS



(f) GGP

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

- Log-likelihood on test dataset

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

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:
  - Latent feature model
  - Bayesian nonparametric (dynamic) recommender systems
  - BNP model for general (non-bipartite) networks

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:
  - Latent feature model
  - Bayesian nonparametric (dynamic) recommender systems
  - BNP model for general (non-bipartite) networks

Bibliography

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

Caron, F. and Teh, Y. W. (2012).
Bayesian nonparametric models for ranked data.
In *Neural Information Processing Systems (NIPS'2012)*.

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.

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

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