Sparse random graphs with exchangeable point processes

François Caron

Department of Statistics, Oxford

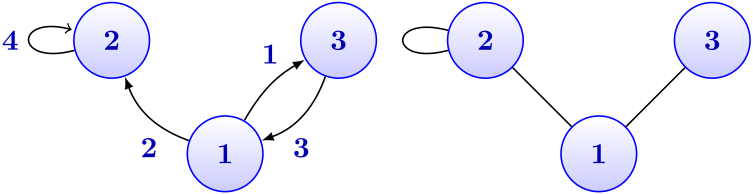
Statistics Seminar, Bocconi University
March 26, 2015

Joint work with Emily Fox (U. Washington)

MARIE CURIE ACTIONS

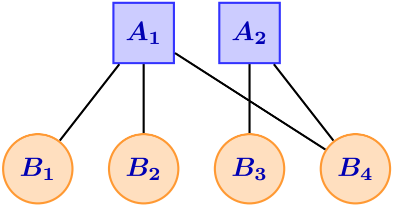
Outline

Introduction



- Multi-edges directed graphs
  - Emails
  - Citations
  - WWW
- Simple graphs
  - Social network
  - Protein-protein interaction

Introduction



- Bipartite graphs
  - Scientists authoring papers
  - Readers reading books
  - Internet users posting messages on forums
  - Customers buying items

Introduction

- Build a statistical model of the network to
  - Find interpretable structure in the network
  - Predict missing edges
  - Predict connections of new nodes

Introduction

- Properties of real world networks
  - Sparsity

$$\text{Dense graph: } n_e = \Theta(n^2)$$
$$\text{Sparse graph: } n_e = o(n^2)$$

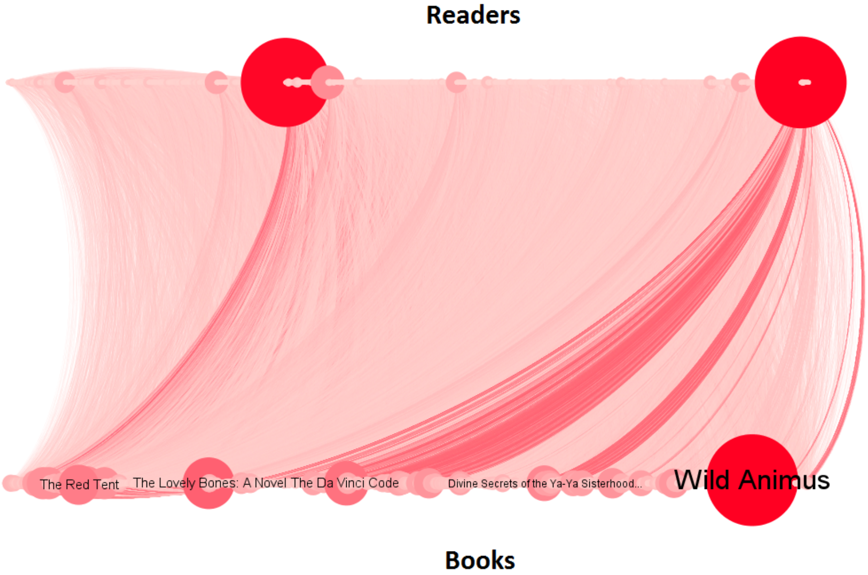
    with n_e the number of edges and n the number of nodes

  - Power-law degree distributions

[Newman, 2009, Clauset et al., 2009]

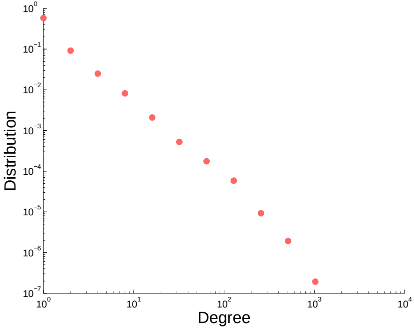
Book-crossing community network

5 000 readers, 36 000 books, 50 000 edges

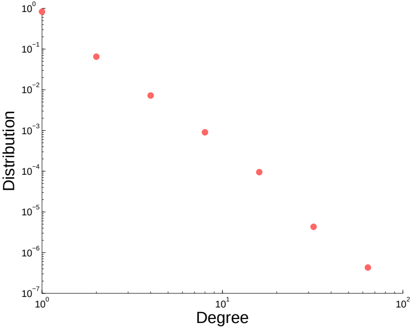


Book-crossing community network

Degree distributions on log-log scale



(a) Readers

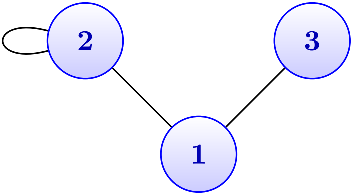


(b) Books

Outline

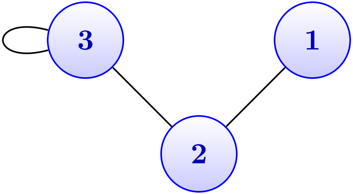
Introduction

- Statistical network modeling
- Probabilistic symmetry: exchangeability
- Ordering of the nodes is irrelevant



Introduction

- Statistical network modeling
- Probabilistic symmetry: exchangeability
- Ordering of the nodes is irrelevant



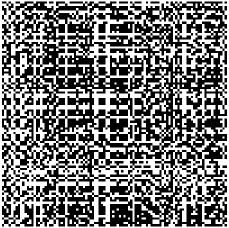
Introduction

- Graphs usually represented by a discrete structure
- Adjacency matrix $X_{ij} \in \{0,1\}, (i,j) \in \mathbb{N}^2$
- Joint exchangeability

$$(X_{ij}) \stackrel{d}{=} (X_{\pi(i)\pi(j)})$$

for any permutation π of $\mathbb{N}$

π



π

Introduction

- Aldous-Hoover representation theorem

$$(X_{ij}) = (F(U_i, U_j, U_{\{ij\}}))$$

  where U_i,$U_{\{ij\}}$ are uniform random variables and F is a random function from $[0,1]^3$ to $\{0,1\}$
- Several network models fit in this framework (e.g. stochastic blockmodel, infinite relational model, etc.)

[Hoover, 1979, Aldous, 1981, Lloyd et al., 2012]

Introduction

- Corollary of A-H theorem

  Exchangeable random graphs are either empty or dense

- To quote the survey paper of Orbanz and Roy

  "the theory [...] clarifies the limitations of exchangeable models. It shows, for example, that most Bayesian models of network data are inherently misspecified"

- Give up exchangeability for sparsity? e.g. preferential attachment model

[Barabási and Albert, 1999, Orbanz and Roy, 2015]

Outline

Point process representation

- Representation of a graph as a (marked) point process over $\mathbb{R}_+^{\mathbf{2}}$
- Representation theorem by Kallenberg for jointly exchangeable point processes on the plane
- Construction based on a completely random measure
- Properties of the model
  - Exchangeability
  - Sparsity
  - Power-law degree distributions (with exponential cut-off)
  - Interpretable parameters and hyperparameters
  - Reinforced urn process construction
- Posterior characterization
- Scalable inference

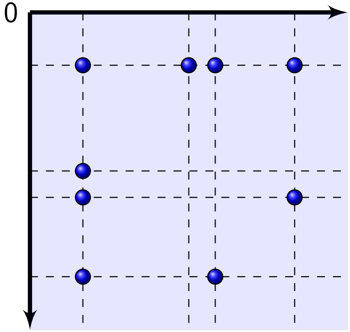
[Kallenberg, 2005, Caron and Fox, 2014]

Point process representation

- Undirected graph represented as a point process on $\mathbb{R}_+^2$

$$Z = \sum_{i,j} z_{ij} \delta_{(\theta_i, \theta_j)}$$

with $\theta_i \in \mathbb{R}$, $z_{ij} \in \{0, 1\}$ with $z_{ij} = z_{ji}$



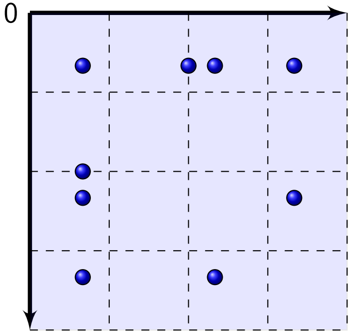
Point process representation

Joint exchangeability

Let $A_i = [h(i-1), hi]$ for $i \in \mathbb{N}$ then

$$(Z(A_i \times A_j)) \overset{d}{=} (Z(A_{\pi(i)} \times A_{\pi(j)}))$$

for any permutation π of $\mathbb{N}$ and any $h > 0$



Point process representation

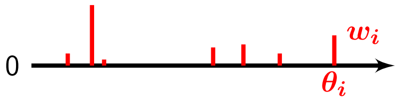
- Kallenberg derived a de Finetti style representation theorem for jointly and separately exchangeable point processes on the plane
- Representation via random transformations of unit rate Poisson processes and uniform variables
- Continuous-time equivalent of Aldous-Hoover for binary variables
- Our construction will fit into this framework

[Kallenberg, 1990, Kallenberg, 2005]

Completely random measures

- Nodes are embedded at some location $\theta_i \in \mathbb{R}_+$
- To each node is associated some sociability parameter w_i
- Homogeneous completely random measure on $\mathbb{R}_+$

$$W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \quad W \sim \mathrm{CRM}(\rho, \lambda).$$



- Lévy measure $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$ with λ the Lebesgue measure

[Kingman, 1967]

Completely random measures

- Lévy measure $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$ with λ the Lebesgue measure
- ρ is a measure on $\mathbb{R}_+$ such that

$$\int_0^\infty (1 - e^{-w})\rho(dw) < \infty. \quad (1)$$

which implies that $W([0, T]) < \infty$ for any $T < \infty$.

$$\int_0^\infty \rho(dw) = \infty \Longrightarrow \text{Infinite number of jumps in any interval } [0, T]$$

"Infinite activity CRM"

$$\int_0^\infty \rho(dw) < \infty \Longrightarrow \text{Finite number of jumps in any interval } [0, T]$$

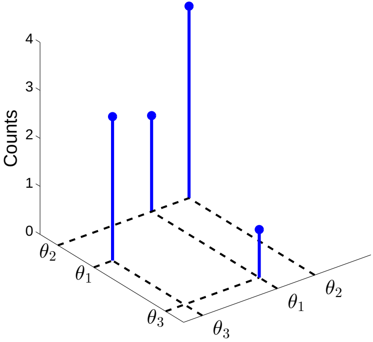
"Finite activity CRM"

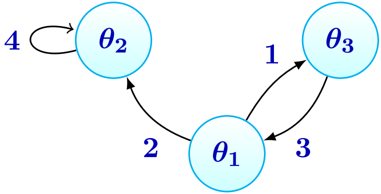
Model for multi-edges directed graphs

We represent the integer-weighted directed graph using an atomic measure on $\mathbb{R}_+^2$

$$D = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n_{ij} \delta_{(\theta_i, \theta_j)},$$

where n_{ij} counts the number of directed edges from node θ_i to node θ_j.

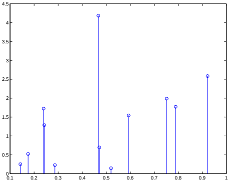




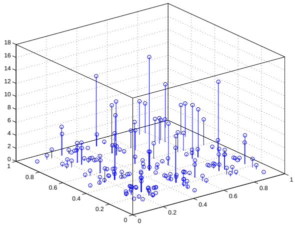
Model for multi-edges directed graphs

- Conditional Poisson process with intensity measure $\widetilde{\boldsymbol{W}} = \boldsymbol{W} \times \boldsymbol{W}$ on the product space $\mathbb{R}_+^{\mathbf{2}}$:

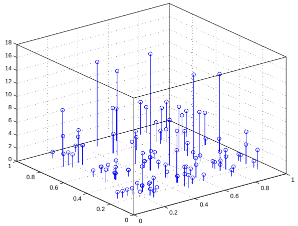
$$\boldsymbol{D} \mid \boldsymbol{W} \sim \mathrm{PP}(\boldsymbol{W} \times \boldsymbol{W})$$



(c) CRM $\boldsymbol{W}$



(d) Intensity measure $\widetilde{\boldsymbol{W}}$



(e) Poisson process $\boldsymbol{D}$

Model for multi-edges directed graphs

- By construction, for any bounded intervals $\boldsymbol{A}$ and $\boldsymbol{B}$ of $\mathbb{R}_+$, $\widetilde{\boldsymbol{W}}(\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{W}(\boldsymbol{A})\boldsymbol{W}(\boldsymbol{B}) < \infty$
- Finite number of counts over $\boldsymbol{A} \times \boldsymbol{B} \subset \mathbb{R}_+^{\mathbf{2}}$

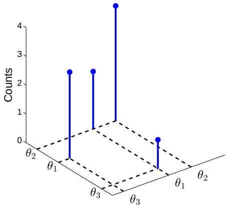
$$\boldsymbol{D}(\boldsymbol{A} \times \boldsymbol{B}) < \infty$$

Model for undirected graphs

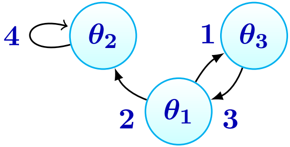
- Point process

$$Z = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} z_{ij} \delta_{(\theta_i, \theta_j)},$$

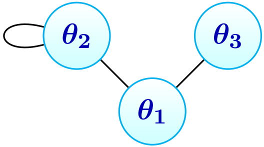
  with the convention $z_{ij} = z_{ji} \in \{0, 1\}$
- Constructed from D by setting $z_{ij} = z_{ji} = 1$ if $n_{ij} + n_{ji} > 0$ and $z_{ij} = z_{ji} = 0$ otherwise



(a) D



(b) Integer-valued directed graph



(c) Undirected graph

Model for undirected graphs

- Hierarchical model

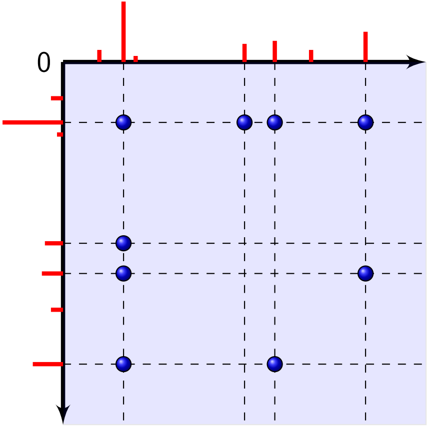
$$\begin{aligned}
&W = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} && W \sim \mathsf{CRM}(\rho, \lambda) \\
&D = \sum_{ij} n_{ij} \delta_{(\theta_i, \theta_j)} && D \sim \mathsf{PP}\left(W \times W\right) \\
&Z = \sum_{ij} \min(n_{ij} + n_{ji}, 1) \delta_{(\theta_i, \theta_j)} &&
\end{aligned}$$

Model for undirected graphs

- Equivalent direct formulation for $i \leq j$

$$\Pr(z_{ij} = 1 \mid w) = \begin{cases} 1 - \exp(-2w_i w_j) & i \neq j \\ 1 - \exp(-w_i^2) & i = j \end{cases}$$

and $z_{ji} = z_{ij}$



Outline

Properties: Exchangeability

Exchangeability

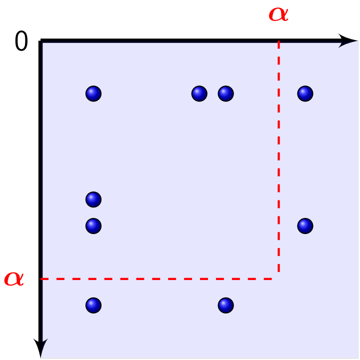
Let $h > 0$ and $A_i = [h(i-1), hi]$, $i \in \mathbb{N}$. By construction,

$$(Z(A_i \times A_j)) \overset{d}{=} (Z(A_{\pi(i)} \times A_{\pi(j)}))$$

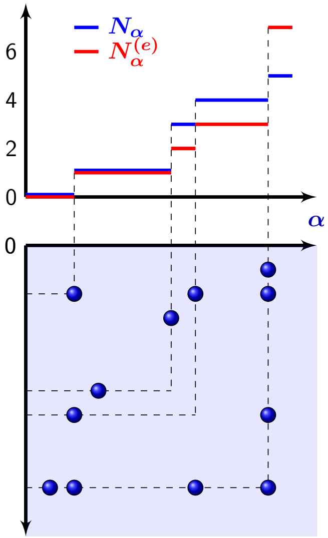
for any permutation π of $\mathbb{N}$.

Properties: Sparsity

- $W(\mathbb{R}_+) = \infty$, so infinite number of edges on $\mathbb{R}_+^2$
- Restrictions D_α and Z_α of D and Z, respectively, to the box $[0, \alpha]^2$.
- N_α number of nodes, and $N_\alpha^{(e)}$ number of edges



Properties: Sparsity



Properties: Sparsity

Definition

(Regular variation) Let $W \sim \mathrm{CRM}(\rho, \lambda)$. The (infinite-activity) CRM is said to be *regularly varying* if the tail Lévy intensity verifies

$$\int_x^\infty \rho(dw) \overset{x \downarrow 0}{\sim} \ell(1/x) x^{-\sigma}$$

for $\sigma \in (0, 1)$ where ℓ is a slowly varying function satisfying $\lim_{t \to \infty} \ell(at)/\ell(t) = 1$ for any $a > 0$.

Properties: Sparsity

Assume $\rho \neq 0$ and $\mathbb{E}[W([0,1])] < \infty$.

Theorem

Let N_α be the number of nodes and $N_\alpha^{(e)}$ the number of edges in the undirected graph restriction, Z_α. Then

$$N_\alpha^{(e)} = \begin{cases} \Theta\left(N_\alpha^2\right) & \text{if } W \text{ is finite-activity} \\ o\left(N_\alpha^2\right) & \text{if } W \text{ is infinite-activity} \\ O\left(N_\alpha^{2/(1+\sigma)}\right) & \text{if } W \text{ is regularly varying}^1 \text{ with } \sigma \in (0,1) \end{cases}$$

almost surely as $\alpha \to \infty$.

[1] with $\lim_{t\to\infty} \ell(t) > 0$

Outline

Generalized Gamma Process

- Lévy intensity
$$\frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\tau w}$$
with $\sigma \in (-\infty, 0]$ and $\tau > 0$
or $\sigma \in (0, 1)$ and $\tau \geq 0$
- Special cases:
  - Gamma process ($\sigma = 0$)
  - Stable process ($\tau = 0, \sigma \in (0, 1)$)
  - Inverse Gaussian process ($\sigma = 1/2, \tau > 0$)
- Infinite activity for $\sigma \geq 0$
- Regularly varying for $\sigma \in (0, 1)$
- Exact sampling of the graph via an urn process
- Power-law degree distribution

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

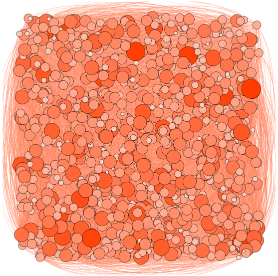
Theorem

Let N_α *be the number of nodes and* $N_\alpha^{(e)}$ *the number of edges in the undirected graph restriction,* Z_α*. Then*

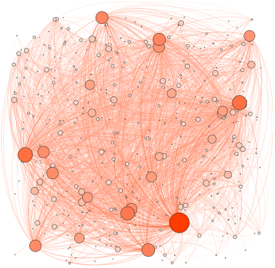
$$N_\alpha^{(e)} = \begin{cases} \Theta\left(N_\alpha^2\right) & \text{if } \sigma < 0 \\ o\left(N_\alpha^2\right) & \text{if } \sigma \in [0,1), \tau > 0 \\ O\left(N_\alpha^{2/(1+\sigma)}\right) & \text{if } \sigma \in (0,1), \tau > 0 \end{cases}$$

almost surely as $\alpha \to \infty$*. That is, the underlying graph is sparse if* $\sigma \geq 0$ *and dense otherwise.*

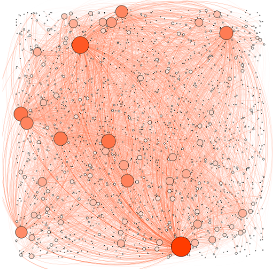
Particular cases: Generalized Gamma Process



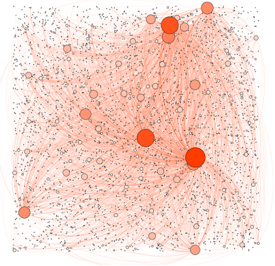
Erdös-Rényi $G(1000, 0.05)$



Gamma Process

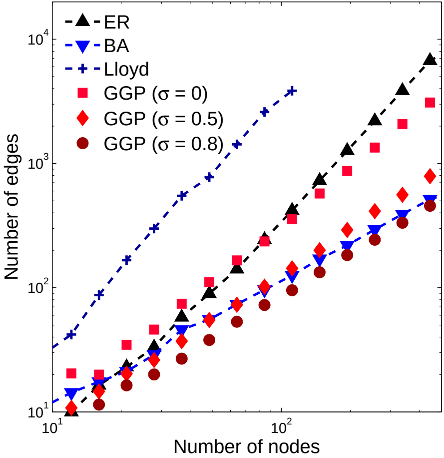


GGP ($\sigma = 0.5$)



GGP ($\sigma = 0.8$)

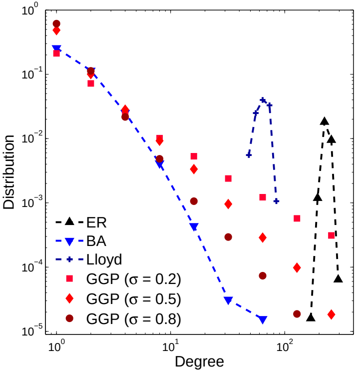
Particular cases: Generalized Gamma Process

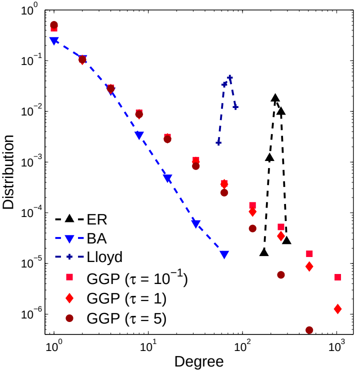


Particular cases: Generalized Gamma Process

Power-law degree distributions

- Power-law like behavior providing a heavy-tailed degree distribution
- Higher power-law exponents for larger σ
- The parameter τ tunes the exponential cut-off in the tails.





Particular cases: Generalized Gamma Process

Outline

Posterior characterization

Conditional distribution of W_α given D_α.

Theorem

Let $(\theta_1, \ldots, \theta_{N_\alpha})$, $N_\alpha \geq 0,$ *be the set of support points of* D_α *such that* $D_\alpha = \sum_{1 \leq i,j \leq N_\alpha} n_{ij} \delta_{(\theta_i, \theta_j)}$. *Let* $m_i = \sum_{j=1}^{N_\alpha} (n_{ij} + n_{ji}) > 0$ *for* $i = 1, \ldots, N_\alpha$. *The conditional distribution of* W_α *given* D_α *is equivalent to the distribution of*

$$w_* \sum_{i=1}^{\infty} \widetilde{P}_i \delta_{\widetilde{\theta}_i} + \sum_{i=1}^{N_\alpha} w_i \delta_{\theta_i}$$

where $\widetilde{\theta}_i \sim \mathrm{Unif}([0, \alpha])$, *and* $(\widetilde{P}_i)|w_* \sim \mathrm{PK}(\rho|w_*)$ *are from a Poisson-Kingman distribution . The weights* $(w_1, \ldots, w_{N_\alpha}, w_*)$ *are jointly dependent conditional on* D_α, *with* $p(w_1, \ldots, w_{N_\alpha}, w_* | D_\alpha) \propto$

$$\left[\prod_{i=1}^{N_\alpha} {w_i}^{m_i}\right] e^{-\left(\sum_{i=1}^{N_\alpha} w_i + w_*\right)^2} \left[\prod_{i=1}^{N_\alpha} \rho(w_i)\right] \times g_\alpha^*(w_*)$$

where g_α^* *is the probability density function of the random variable* $W_\alpha^* = W_\alpha([0, \alpha])$.

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

Posterior inference for undirected graphs

- Let $\phi = (\alpha, \sigma, \tau)$ with flat priors
- We want to approximate

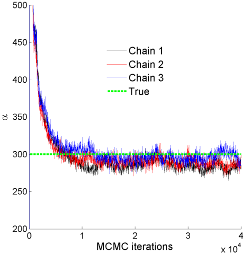
$$p(w_1, \dots, w_{N_\alpha}, w_*, \phi | (z_{ij})_{1\leq i,j\leq N_\alpha})$$

- Latent count variables $\overline{n}_{ij} = n_{ij} + n_{ji}$
- Markov chain Monte Carlo sampler
  1. Update the weights $(w_1, \dots, w_{N_\alpha})$ given the rest using an Hamiltonian Monte Carlo update
  2. Update the total mass w_* and hyperparameters $\phi = (\alpha, \sigma, \tau)$ given the rest using a Metropolis-Hastings update
  3. Update the latent counts $(\overline{n}_{ij})$ given the rest from a truncated Poisson distribution

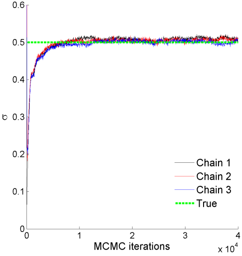
Outline

Simulated data

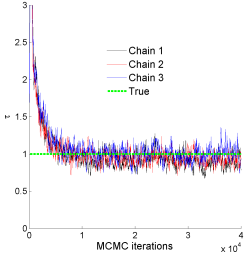
- Simulation of a GGP graph with $\boldsymbol{\alpha = 300, \sigma = 1/2, \tau = 1}$
- 13,995 nodes and 76,605 edges
- MCMC sampler with 3 chains and 40,000 iterations
- Takes 10min on a standard desktop with Matlab



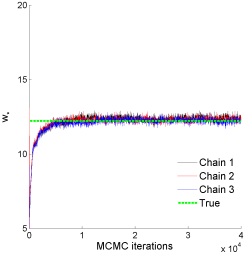
(a) $\boldsymbol{\alpha}$



(b) $\boldsymbol{\sigma}$

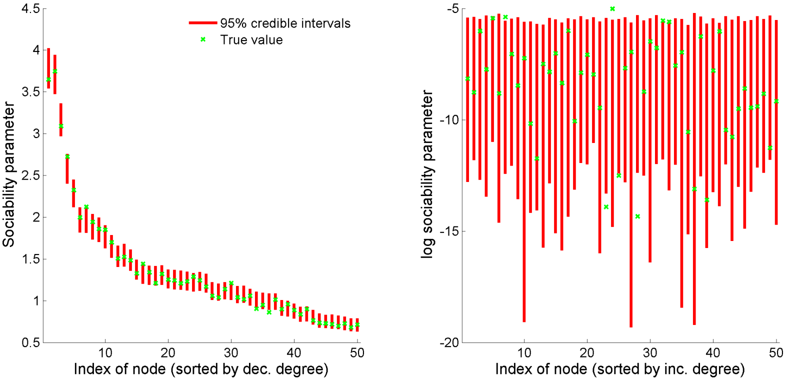


(c) $\boldsymbol{\tau}$



(d) $\boldsymbol{w_*}$

Simulated data



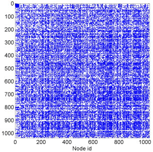
(a) 50 nodes with highest degree

(b) 50 nodes with lowest degree

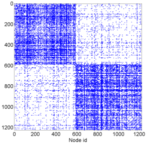
Figure: 95 % posterior intervals of (a) the sociability parameters w_i of the 50 nodes with highest degree and (b) the log-sociability parameter $\log w_i$ of the 50 nodes with lowest degree. True values are represented by a green star.

Real network data

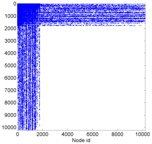
- Assessing the sparsity of the network
- We aim at reporting $\mathbf{Pr}(\boldsymbol{\sigma} \geq \mathbf{0}|\boldsymbol{z})$ based on a set of observed connections $(\boldsymbol{z})$
- 12 different networks
- $\sim \mathbf{1,000 - 300,000}$ nodes and $\mathbf{10,000 - 1,000,000}$ edges



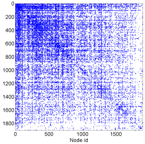
(a) facebook107



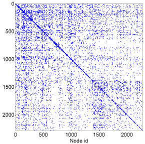
(b) polblogs



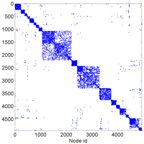
(c) USairport



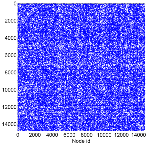
(d) UCirvine



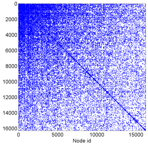
(e) yeast



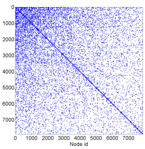
(f) USpower



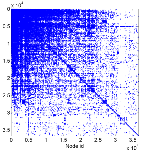
(g) IMDB



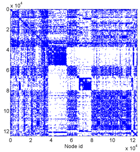
(h) cond-mat1



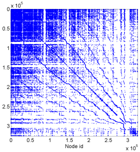
(i) cond-mat2



(j) enron



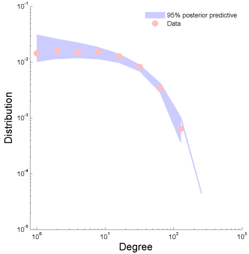
(k) internet



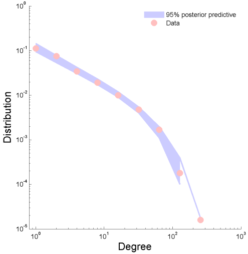
(l) www

Real network data

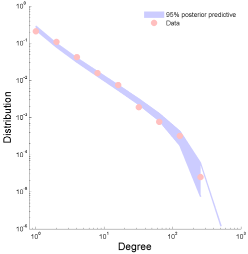
Name	Nb nodes	Nb edges	Time (min)	$\mathbf{Pr(\sigma > 0 \mid z)}$	99% CI $\boldsymbol{\sigma}$
facebook107	1,034	26,749	1	0.00	$\mathbf{[-1.06, -0.82]}$
polblogs	1,224	16,715	1	0.00	$\mathbf{[-0.35, -0.20]}$
USairport	1,574	17,215	1	1.00	$\mathbf{[\ 0.10,\ 0.18]}$
UCirvine	1,899	13,838	1	0.00	$\mathbf{[-0.14, -0.02]}$
yeast	2,284	6,646	1	0.28	$\mathbf{[-0.09, 0.05]}$
USpower	4,941	6,594	1	0.00	$\mathbf{[-4.84, -3.19]}$
IMDB	14,752	38,369	2	0.00	$\mathbf{[-0.24, -0.17]}$
cond-mat1	16,264	47,594	2	0.00	$\mathbf{[-0.95, -0.84]}$
cond-mat2	7,883	8,586	1	0.00	$\mathbf{[-0.18, -0.02]}$
Enron	36,692	183,831	7	1.00	$\mathbf{[\ 0.20,\ 0.22]}$
internet	124,651	193,620	15	0.00	$\mathbf{[-0.20, -0.17]}$
www	325,729	1,090,108	132	1.00	$\mathbf{[0.26, 0.30]}$



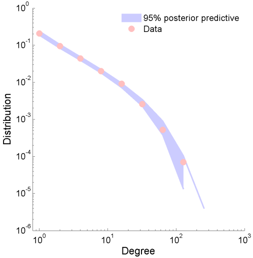
(a) facebook107



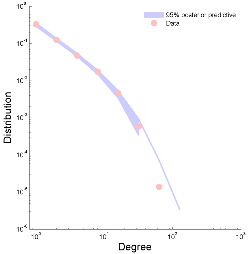
(b) polblogs



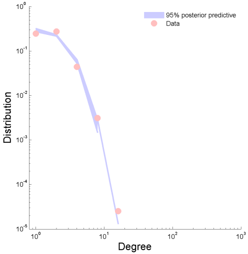
(c) USairport



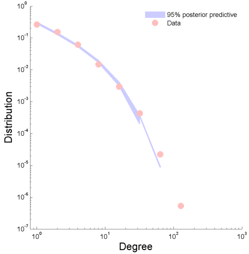
(d) UCirvine



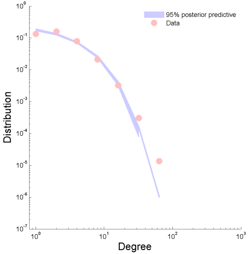
(e) yeast



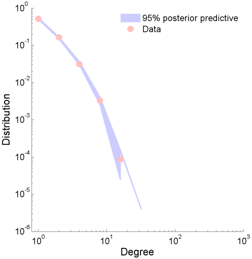
(f) USpower



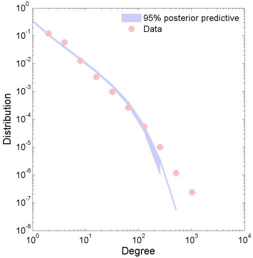
(g) IMDB



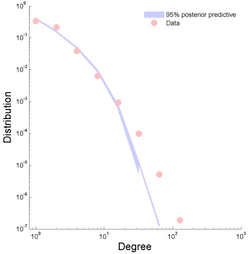
(h) cond-mat1



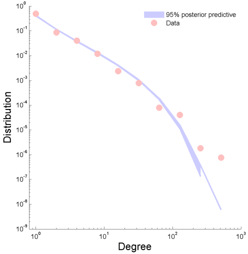
(i) cond-mat2



(j) enron



(k) internet



(l) www

Conclusion

- Statistical network models
- Build on exchangeable random measures
- Sparsity and power-law properties
- Scalable inference
- Similar construction for bipartite graphs
- Extensions to more structured models: low-rank, block-model, covariates, dynamic networks,etc

Bibliography I

Aldous, D. J. (1981).
Representations for partially exchangeable arrays of random variables.
Journal of Multivariate Analysis, 11(4):581–598.

Barabási, A. L. and Albert, R. (1999).
Emergence of scaling in random networks.
Science, 286(5439):509–512.

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

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

Caron, F. and Fox, E. B. (2014).
Sparse graphs using exchangeable random measures.
Technical report, arXiv:1401.1137.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009).
Power-law distributions in empirical data.
SIAM review, 51(4):661–703.

Bibliography II

Hoover, D. N. (1979).
Relations on probability spaces and arrays of random variables.
Preprint, Institute for Advanced Study, Princeton, NJ.

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

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

Kallenberg, O. (1990).
Exchangeable random measures in the plane.
Journal of Theoretical Probability, 3(1):81–136.

Kallenberg, O. (2005).
Probabilistic symmetries and invariance principles.
Springer.

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

Bibliography III

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.

Lloyd, J., Orbanz, P., Ghahramani, Z., and Roy, D. (2012).
Random function priors for exchangeable arrays with applications to graphs and relational data.
In *NIPS*, volume 25, pages 1007–1015.

Newman, M. (2009).
Networks: an introduction.
OUP Oxford.

Orbanz, P. and Roy, D. M. (2015).
Bayesian models of graphs, arrays and other exchangeable random structures.
IEEE Trans. Pattern Anal. Mach. Intelligence (PAMI), 37(2):437–461.

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