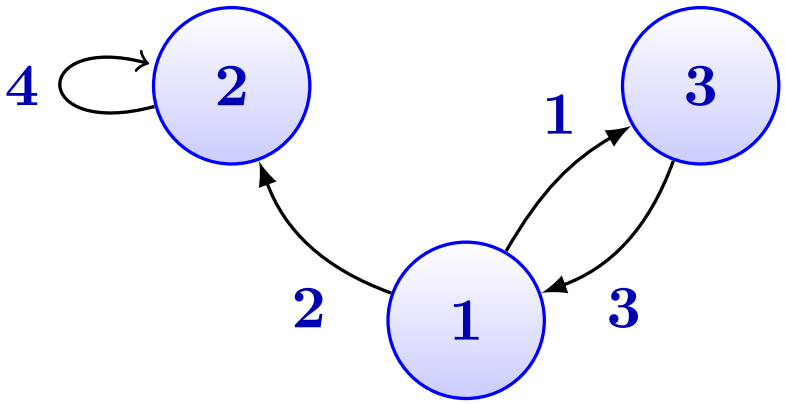
Sparse graphs using exchangeable random measures

François Caron

Department of Statistics, Oxford

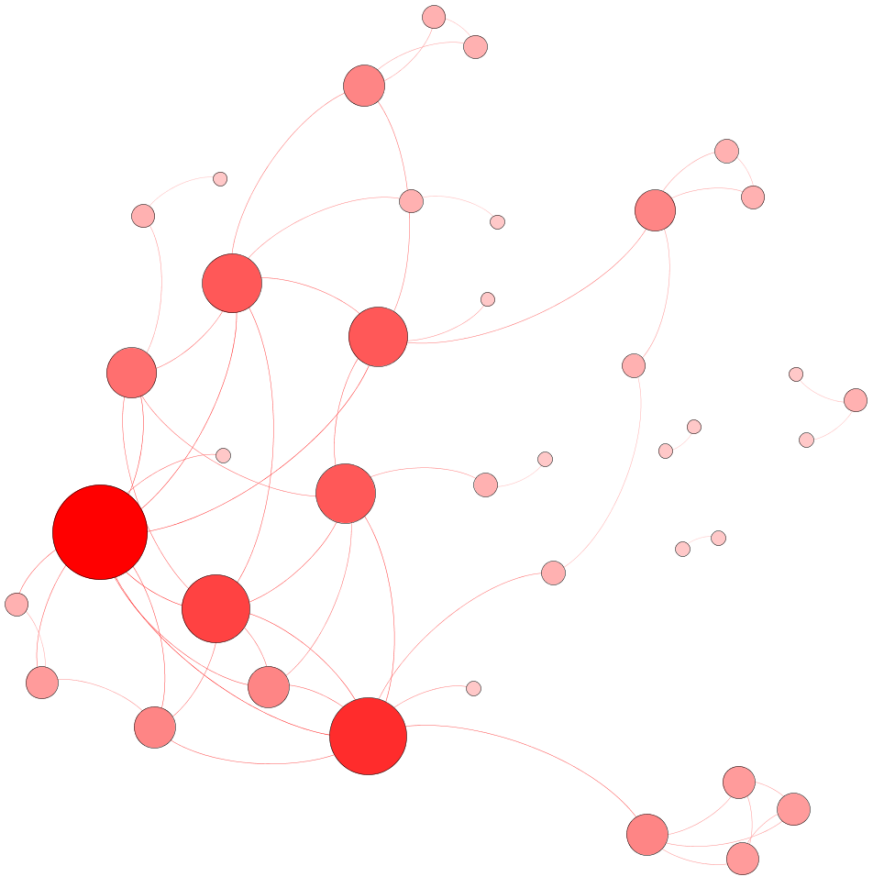
Oxford/Warwick workshop
Scalable Statistical Methods for Analysis of large and complex data sets
October 9, 2015

Introduction



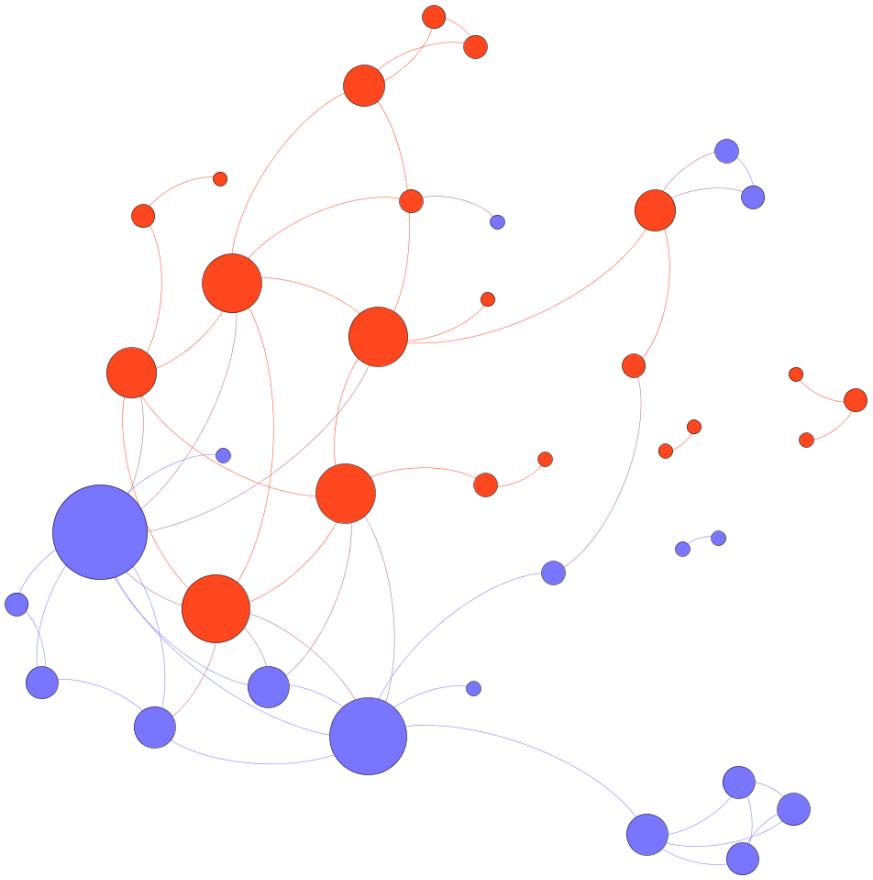
- Directed Multigraphs
  - Emails
  - Citations
  - WWW

Introduction



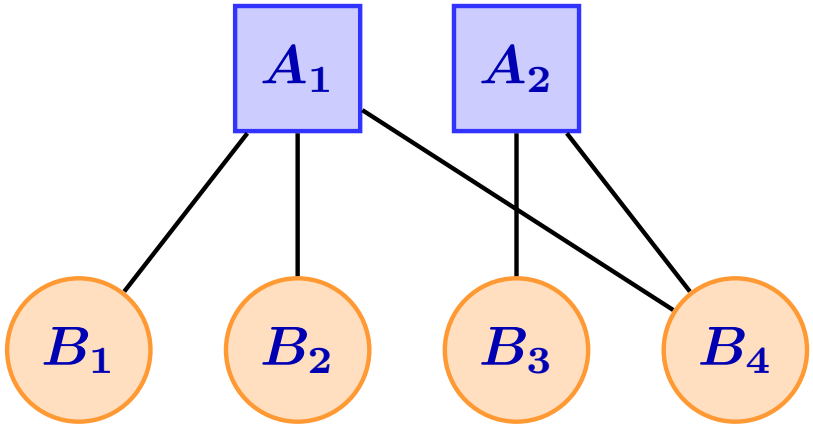
- Simple graphs
  - Social network
  - Protein-protein interaction

Introduction



- 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

- Massive networks
  - Linkedin: $\simeq$ 300 millions
  - Facebook: $\simeq$ billion
  - Twitter: $\simeq$ 300 millions
  - www: $\simeq$ billion
- Capture large-scale properties of networks
- Scalable inference algorithms

Introduction

- Properties of real-world networks
  - Sparsity

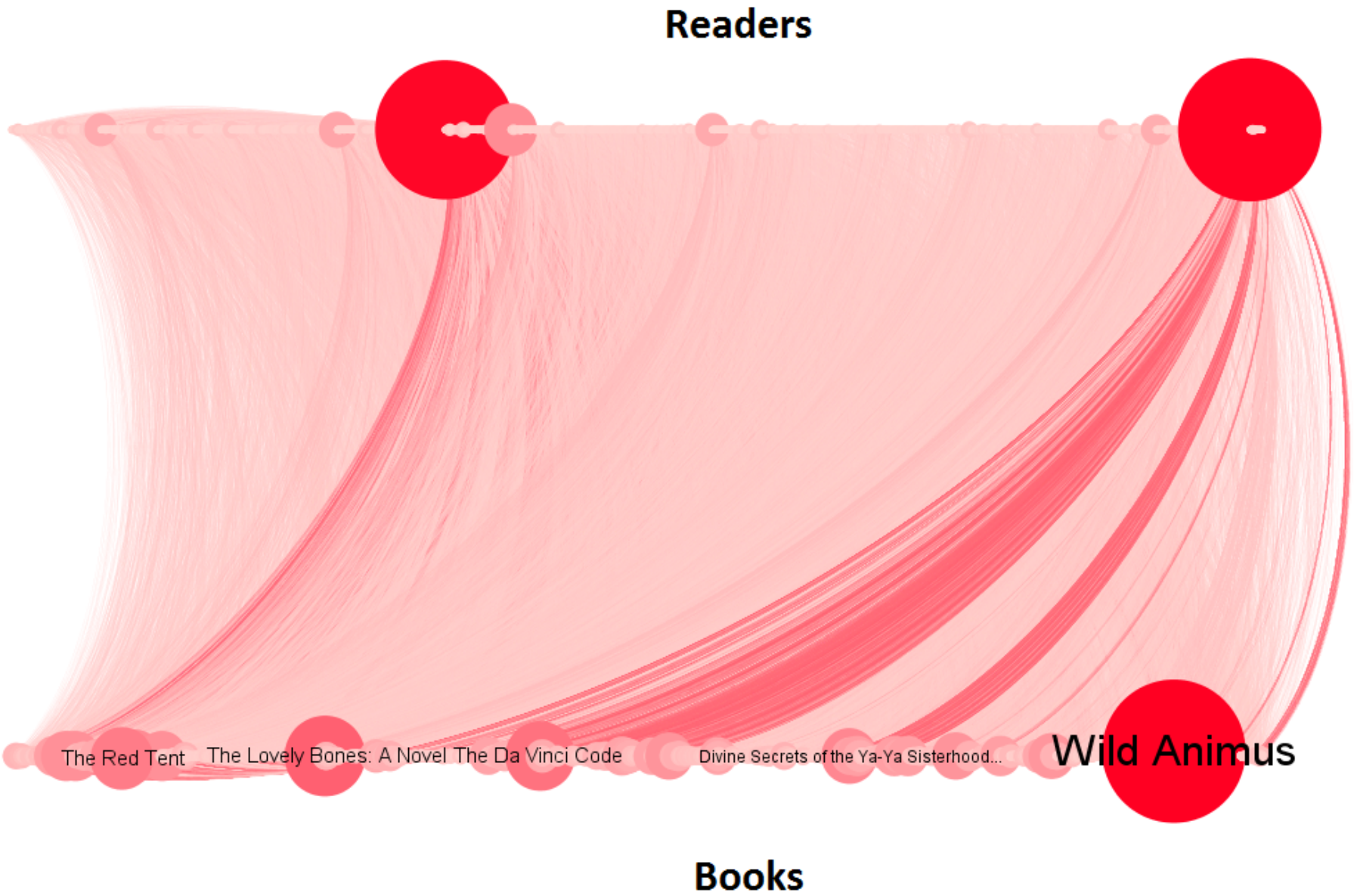
    Dense graph: $n_e = \Theta(n^2)$
    Sparse graph: $n_e = o(n^2)$

    with n_e the number of edges and n the number of nodes
  - Heavy-tailed degree distributions
  - Latent structure

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

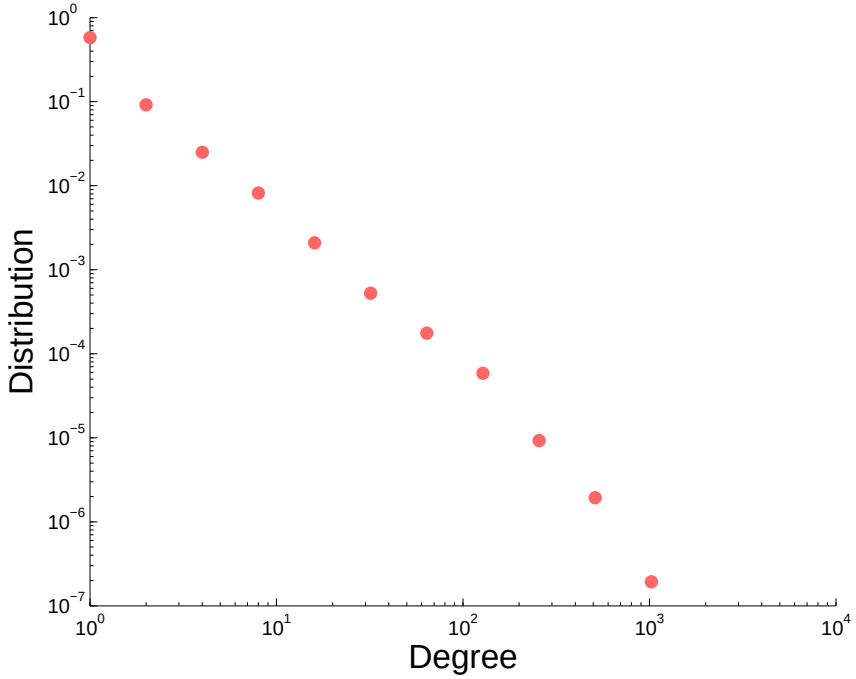
Book-crossing community network

5 000 readers, 36 000 books, 50 000 edges

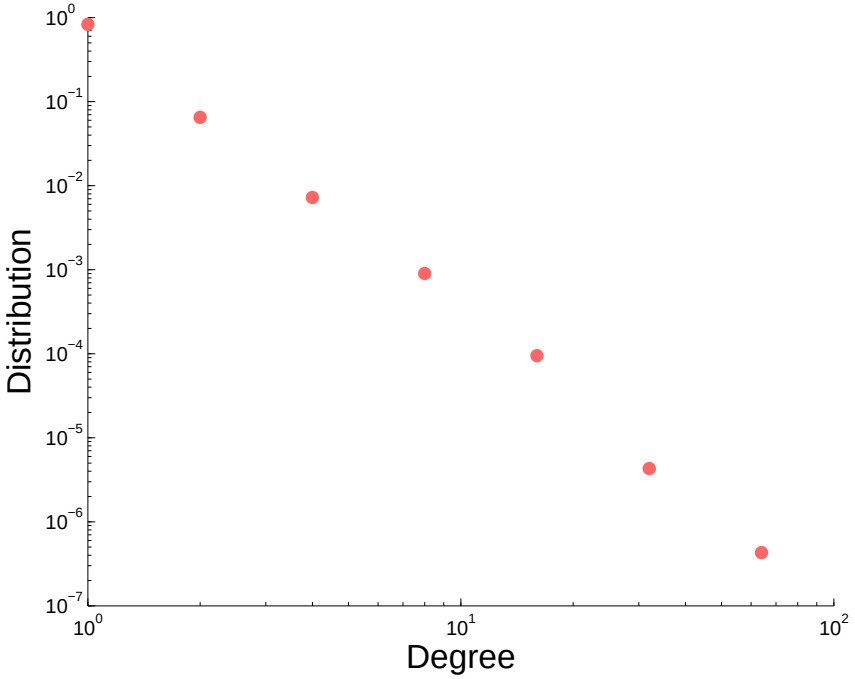


Book-crossing community network

Degree distributions on log-log scale



(a) Readers



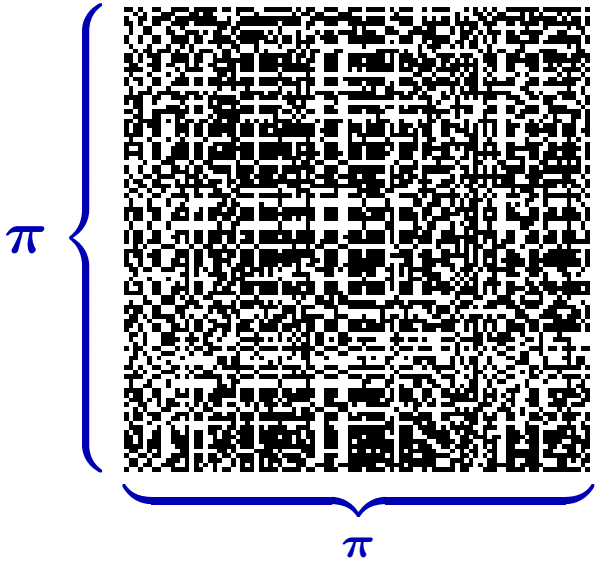
(b) Books

Introduction

- Simple graphs
- 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 $\mathbb{N}$



Introduction

- Aldous-Hoover representation theorem for exchangeable binary matrices
$$\boldsymbol{X_{ij}|U_i, U_j, W \sim \mathrm{Ber}(W(U_i, U_j))}$$
with $\boldsymbol{U_i} \overset{\text{iid}}{\sim} \mathbf{Unif(0,1)}$ and $\boldsymbol{W} : \mathbf{[0,1]^2 \to [0,1]}$ a random function
- Several network models fit in this framework
  - Erdös-Rényi, (mixed-membership) stochastic block-models, infinite relational models, etc

[Hoover, 1979, Aldous, 1981, Lloyd et al., 2012, Orbanz and Roy, 2015]

Introduction

- Corollary of A-H theorem

  Graphs represented by an exchangeable matrix are either trivially 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"

[Hoover, 1979, Aldous, 1981, Lloyd et al., 2012, Orbanz and Roy, 2015]

Introduction

How to handle sparse graphs?

- Give up infinite exchangeability?
  - Non-exchangeable generative models
    - Preferential attachment model
  - Sequence of finitely exchangeable models $(X_{ij}^{(n)})_{1 \leq i,j \leq n}$
    - Chung-Lu

$$X_{ij}^{(n)} \sim \mathrm{Ber}\left(\frac{w_i w_j}{\sum_{k=1}^{n} w_k}\right)$$

    - Sparsification of the graphon

$$X_{ij}^{(n)} \sim \mathrm{Ber}(\rho_n W(U_i, U_j))$$

  with $\rho_n \to 0$

[Barabási and Albert, 1999, Chung and Lu, 2002, Bickel and Chen, 2009, Wolfe and Olhede, 2013]

Point process representation

- Representation of a graph as a (marked) point process over $\mathbb{R}_+^2$
- Representation theorem by Kallenberg for jointly exchangeable point processes on the plane
- Construction based on completely random measures
- Properties of the model
  - Exchangeable point process
  - Sparsity
  - Heavy-tailed degree distributions
- Scalable inference

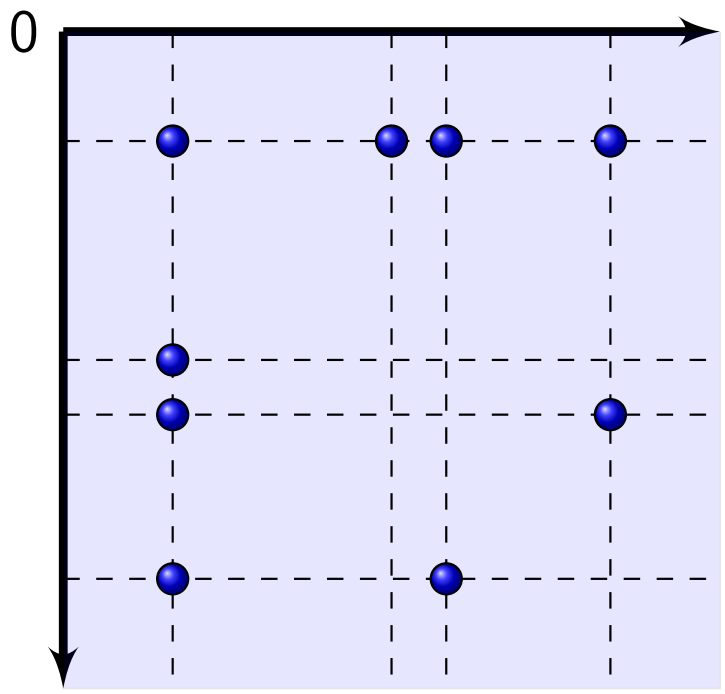
[Kallenberg, 2005, Caron and Fox, 2014]

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 $z_{ij} = z_{ji}$



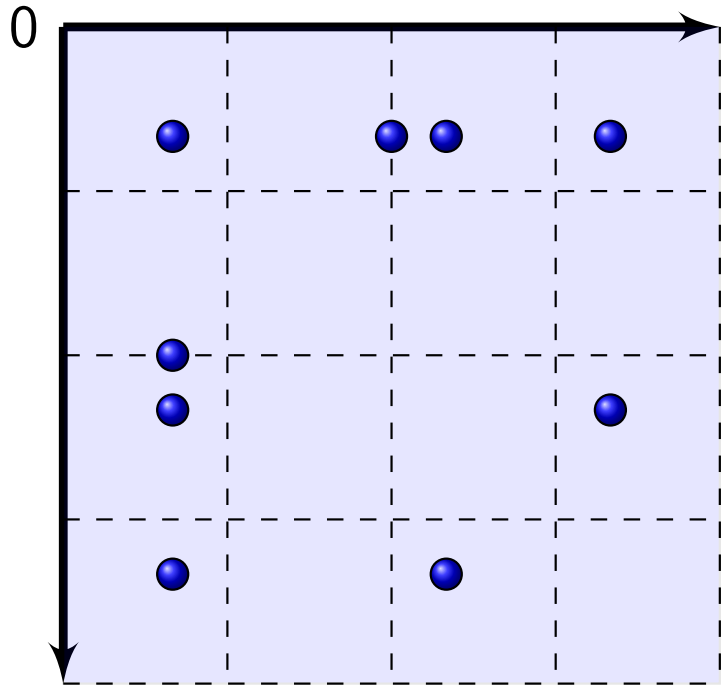
Point process representation

Joint exchangeability

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

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

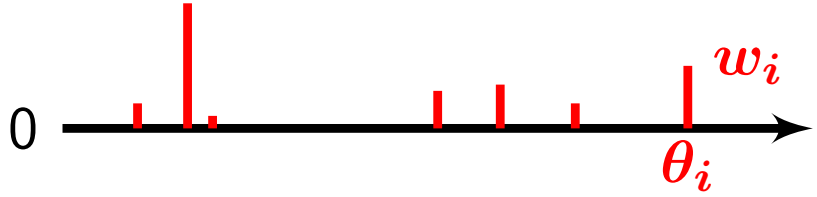
for any permutation π of $\mathbb{N}$ and any $h > 0$



Completely random measures

- Nodes are embedded at some location $\theta_i \in \mathbb{R}_+$
- Each node has a sociability parameter w_i
- Homogeneous completely random measure on $\mathbb{R}_+$

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



- Lévy measure $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$

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

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

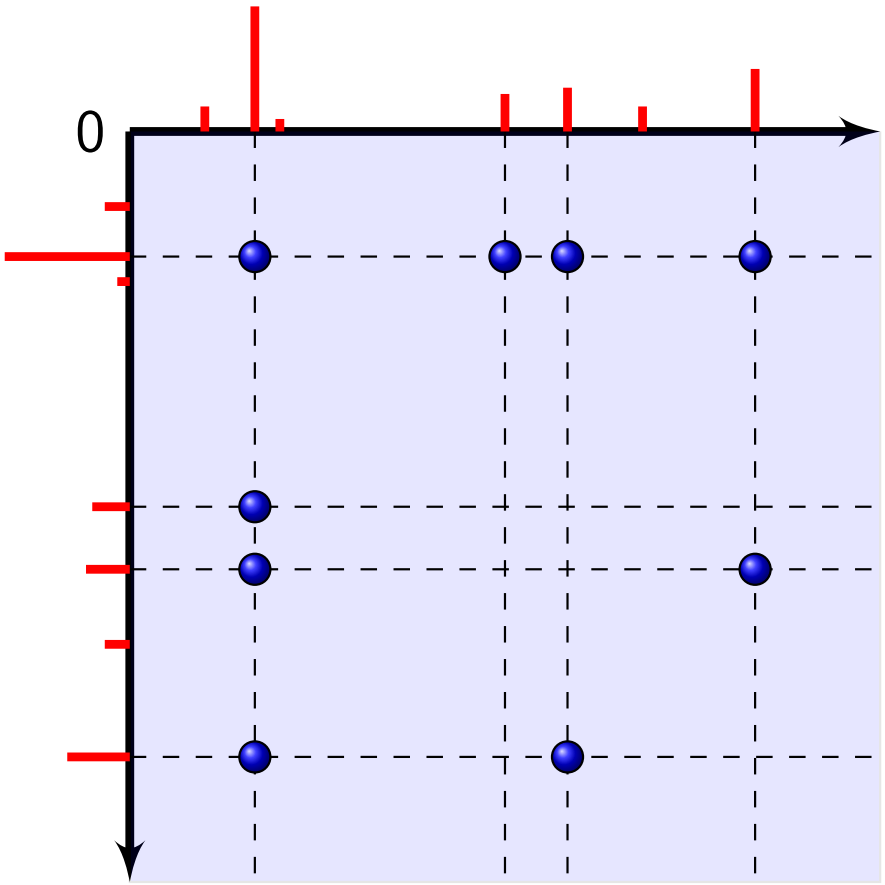
[Kingman, 1967]

Model for undirected graphs

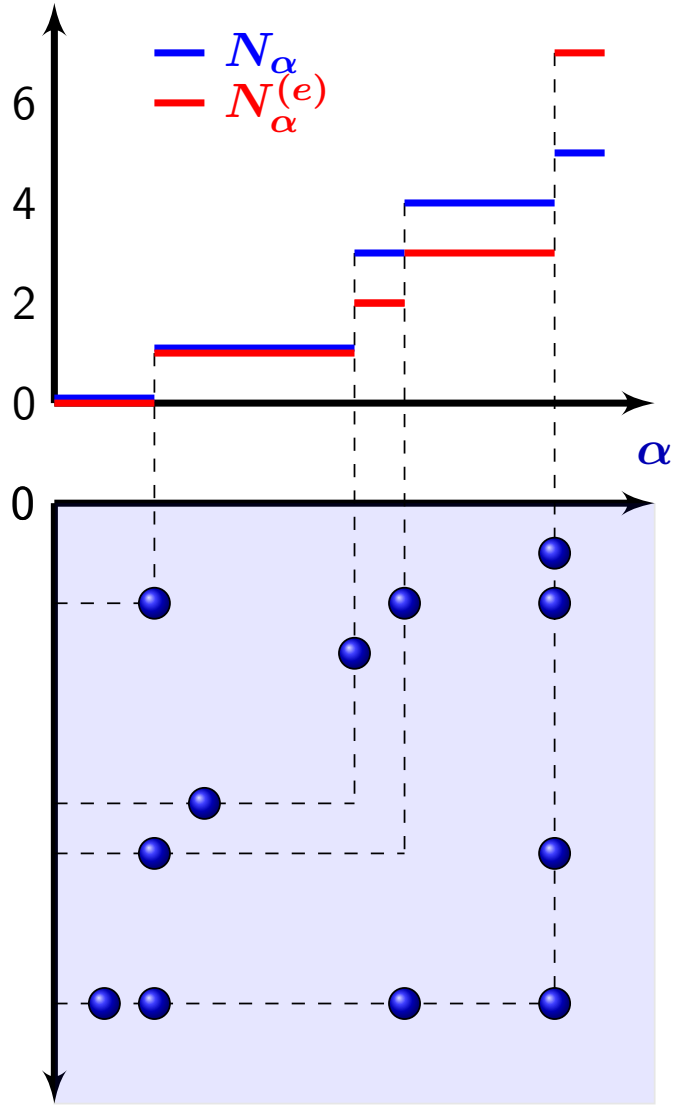
- For $i \leq j$

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

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



Properties: Sparsity



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

$$N_\alpha^{(e)} = \begin{cases} \Theta\left(N_\alpha^2\right) & \text{if } W \text{ is finite-activity} \\ o\left(N_\alpha^2\right) & \text{if } W \text{ is infinite-activity} \end{cases}$$

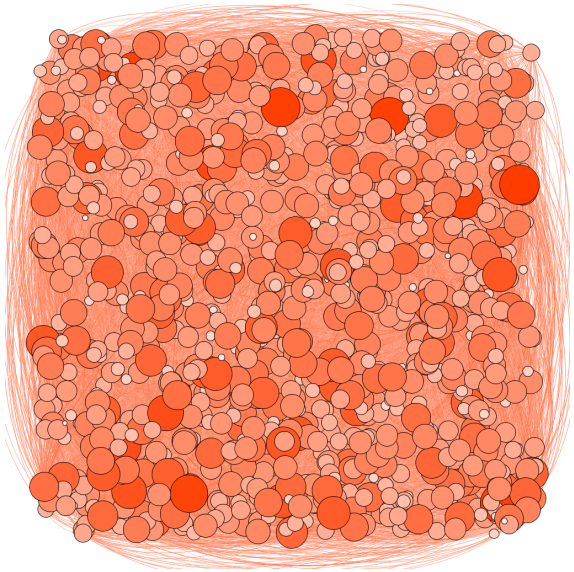
almost surely as $\alpha \to \infty$.

Particular case: Generalized Gamma Process

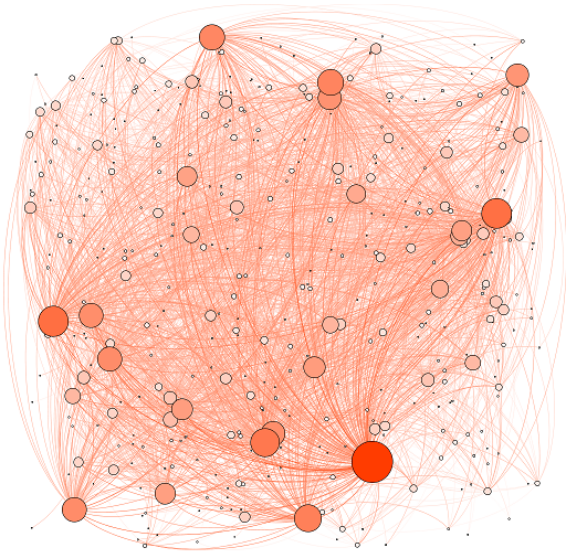
- Lévy intensity
$$\frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\tau w}$$
with $\sigma \in (-\infty, 0]$ and $\tau > 0$
or $\sigma \in (0, 1)$ and $\tau \geq 0$
- Infinite activity for $\sigma \geq 0$
- Exact sampling of the graph via an urn process
- Power-law degree distribution

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

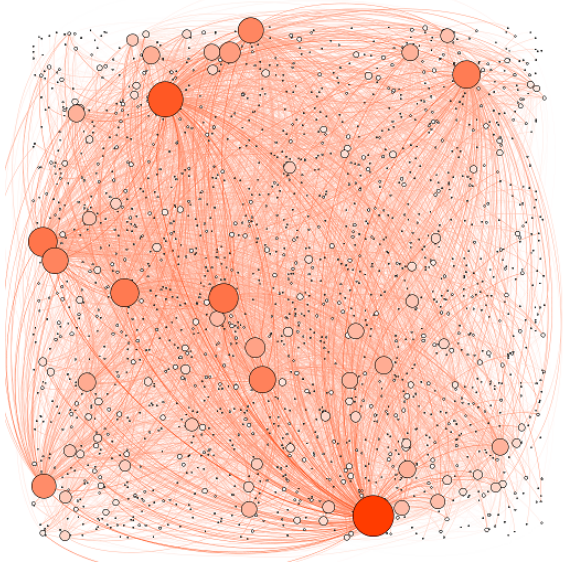
Particular case: Generalized Gamma Process



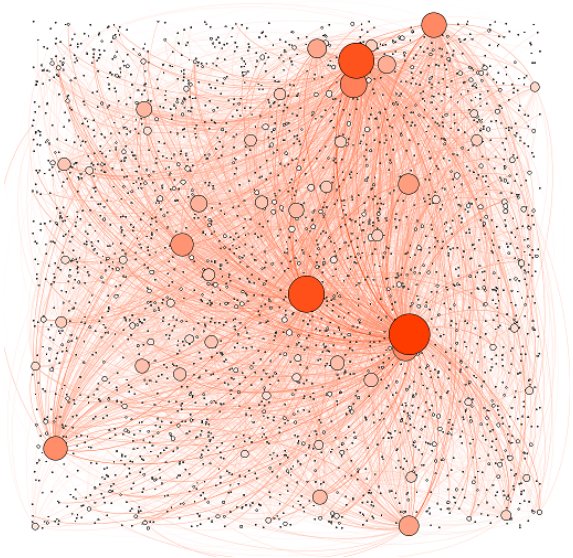
Erdös-Rényi $G(1000, 0.05)$



Gamma Process



GGP ($\sigma = 0.5$)

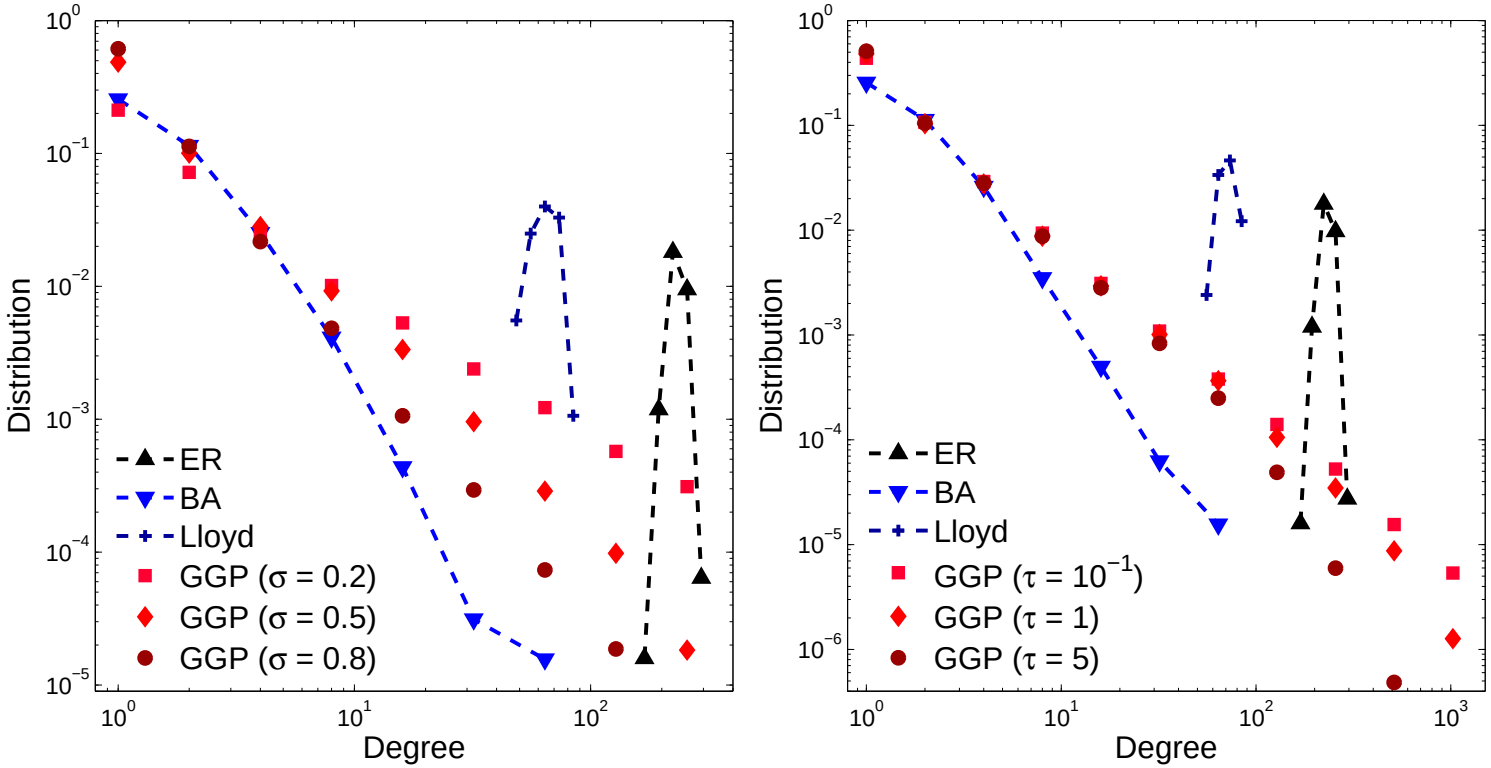


GGP ($\sigma = 0.8$)

Particular case: 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 case: Generalized Gamma Process

Posterior inference

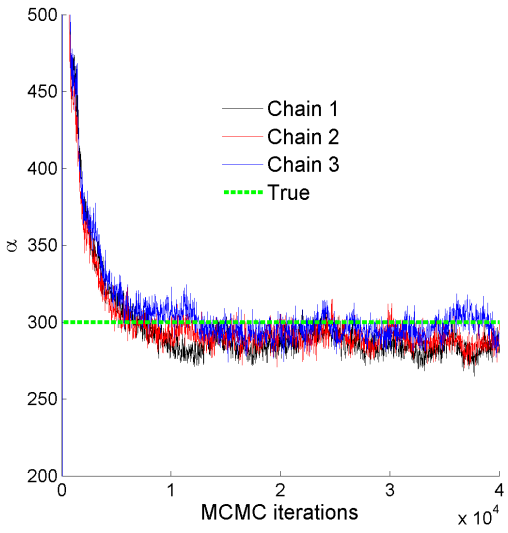
- Let $\phi = (\alpha, \sigma, \tau)$ with improper priors
- We want to approximate

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

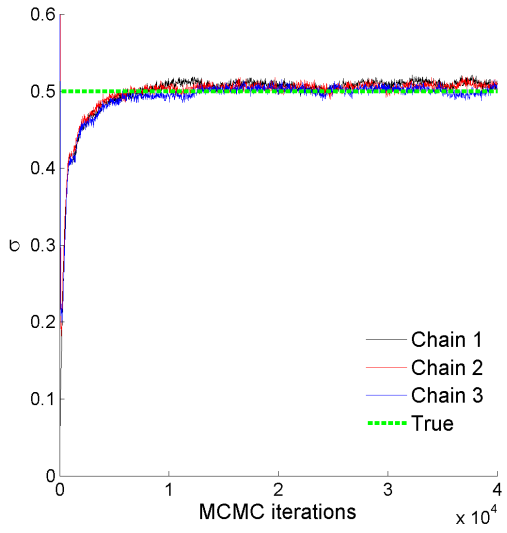
- Latent count variables $\overline{n}_{ij} = n_{ij} + n_{ji}$
- Markov chain Monte Carlo sampler
  1. Update the weights $(w_1, \ldots, w_{N_\alpha})$ given the rest using an Hamiltonian Monte Carlo update
  2. Update the total mass w_* and hyperparameters $\phi = (\alpha, \sigma, \tau)$ given the rest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

Simulated data

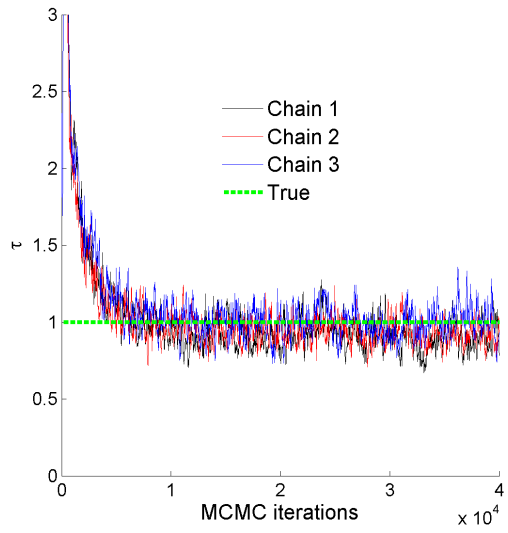
- Simulation of a GGP graph with $\boldsymbol{\alpha = 300, \sigma = 1/2, \tau = 1}$
- 13,995 nodes and 76,605 edges
- MCMC sampler with 3 chains and 40,000 iterations
- Takes 10min on a standard desktop with Matlab



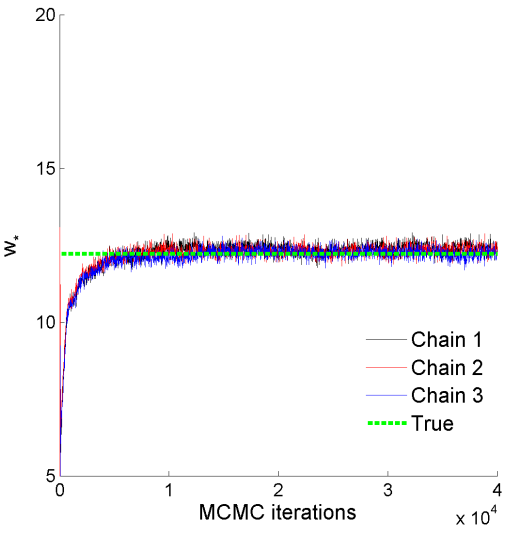
(a) α



(b) σ

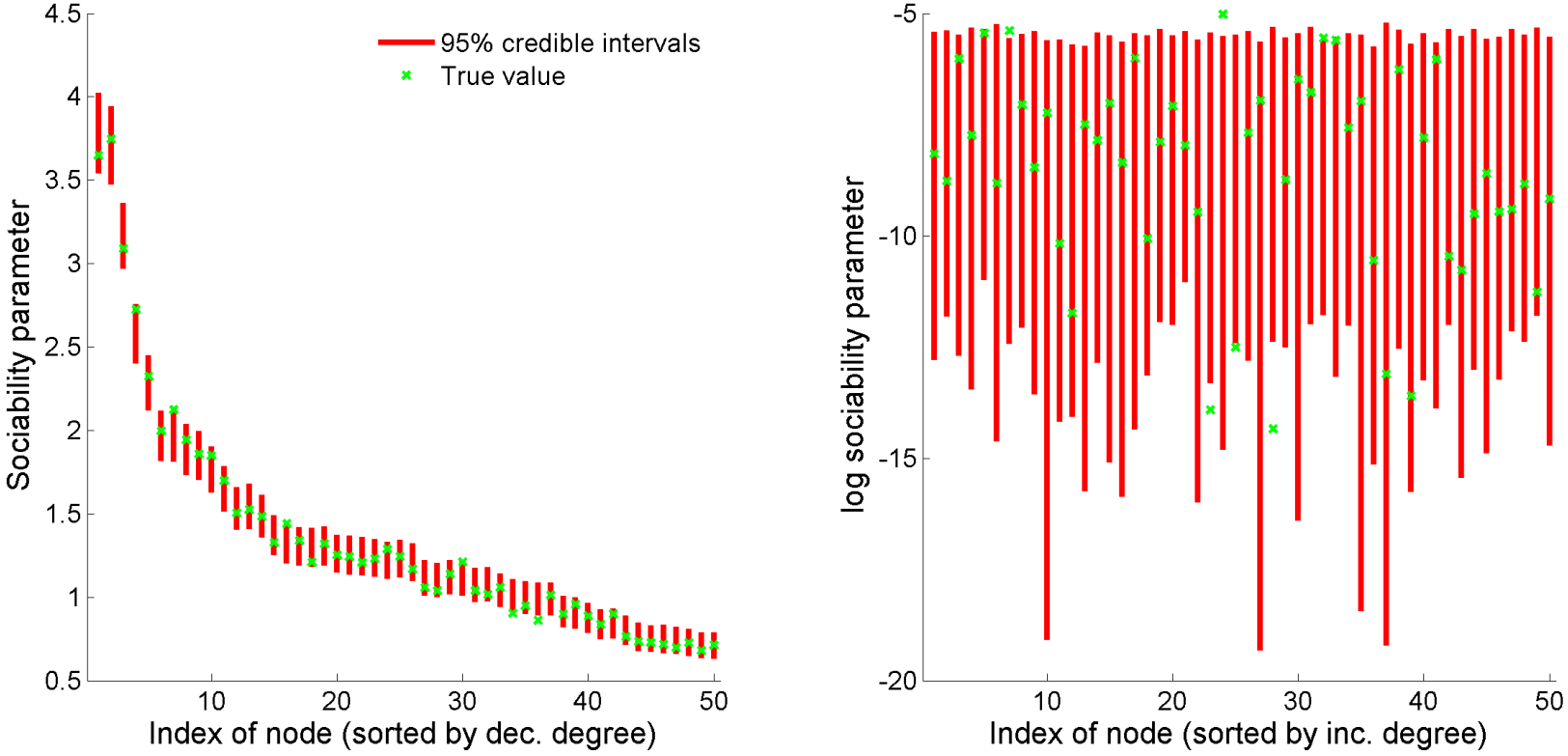


(c) τ



(d) w_*

Simulated data



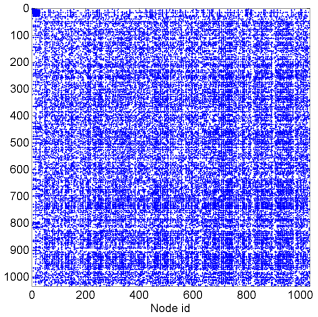
(a) 50 nodes with highest degree

(b) 50 nodes with lowest degree

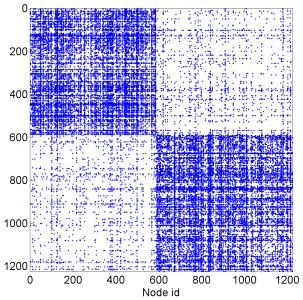
Figure: 95 % posterior intervals of (a) the sociability parameters w_i of the 50 nodes with highest degree and (b) the log-sociability parameter $\log w_i$ of the 50 nodes with lowest degree. True values are represented by a green star.

Real network data

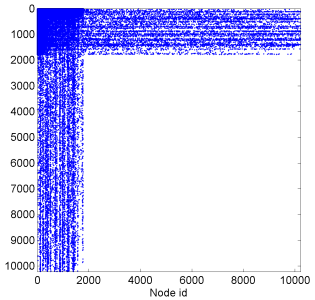
- Assessing the sparsity of the network
- We aim at reporting $\mathbf{Pr}(\boldsymbol{\sigma} \geq \mathbf{0}|\boldsymbol{z})$ based on a set of observed connections $(\boldsymbol{z})$
- 12 different networks
- $\sim \mathbf{1,000 - 300,000}$ nodes and $\mathbf{10,000 - 1,000,000}$ edges



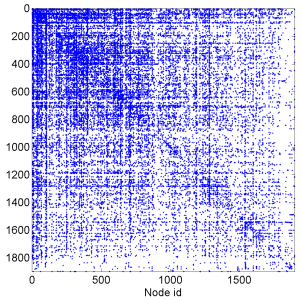
(a) facebook107



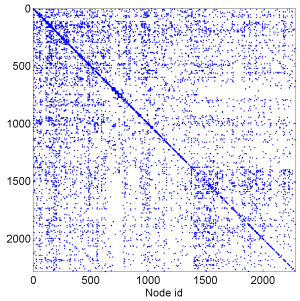
(b) polblogs



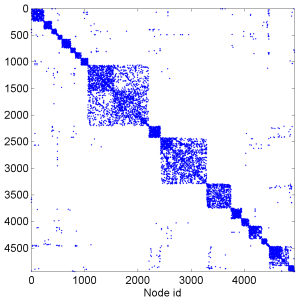
(c) USairport



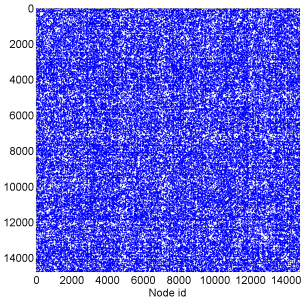
(d) UCirvine



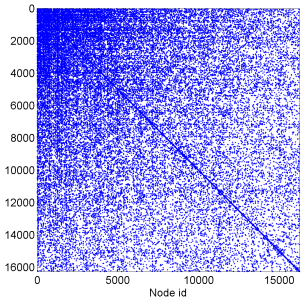
(e) yeast



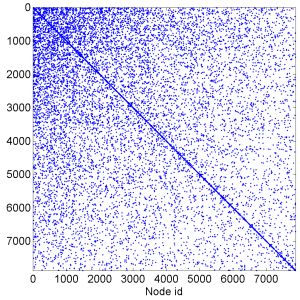
(f) USpower



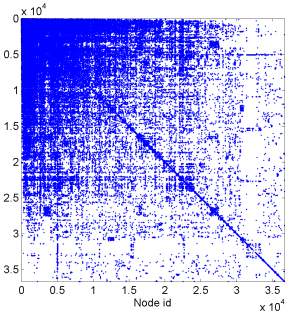
(g) IMDB



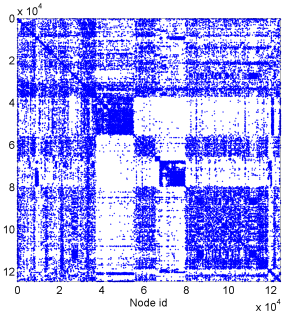
(h) cond-mat1



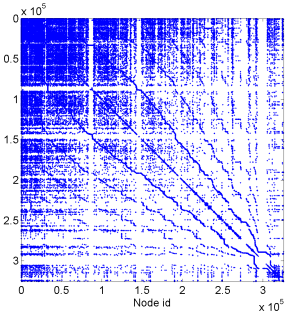
(i) cond-mat2



(j) enron



(k) internet



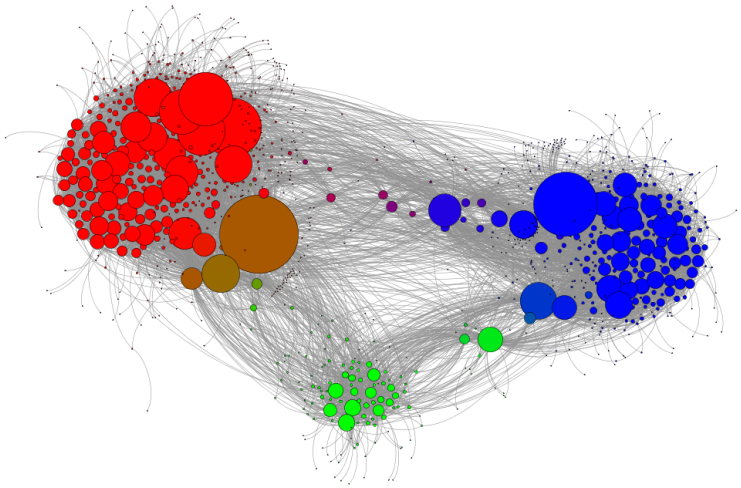
(l) www

Real network data

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

Conclusion

- Statistical network models
- Build on exchangeable random measures
- Sparsity and power-law properties
- Scalable inference
- Extensions to more structured models: non-negative factorization, block-model, covariates, dynamic networks,etc



- Matlab code available

  http://www.stats.ox.ac.uk/~caron/code/bnpgraph/

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.

Bickel, P. J. and Chen, A. (2009).
A nonparametric view of network models and Newman–Girvan and other modularities.
Proceedings of the National Academy of Sciences, 106(50):21068–21073.

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.

Bibliography II

Chung, F. and Lu, L. (2002).
The average distances in random graphs with given expected degrees.
Proceedings of the National Academy of Sciences, 99(25):15879–15882.

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

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

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

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

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.

Bibliography III

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.

Wolfe, P. J. and Olhede, S. C. (2013).
Nonparametric graphon estimation.
arXiv preprint arXiv:1309.5936.