Bayesian Nonparametric Models for Sparse Bipartite Networks

Candidate: Andrea Ferrero[1,2]

Supervisor: Saverio Ranciati[1]

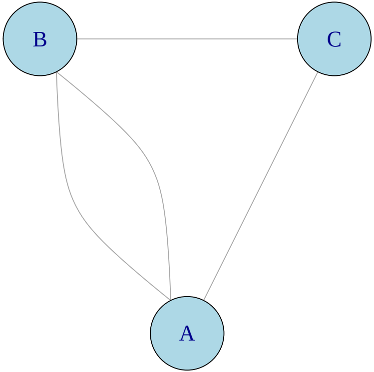
Co-supervisors: Julyan Arbel[2], Tâm Le Minh[2]

[1]Department of Statistical Sciences "Paolo Fortunati"
University of Bologna

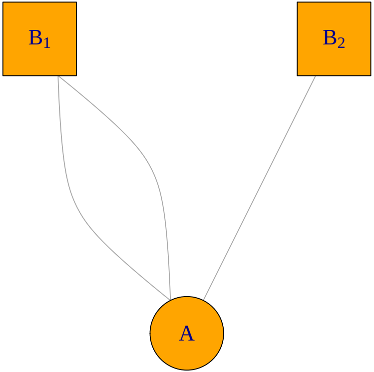
[2]Statify team
Inria Center of University of Grenoble Alpes

Overview

Networks Structures

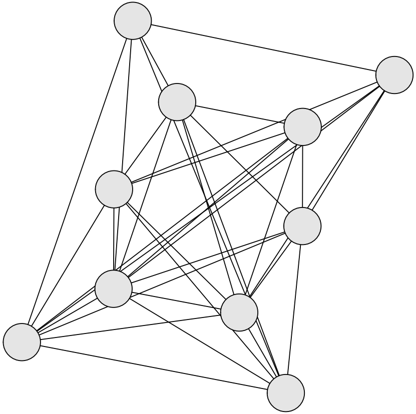


Unipartite

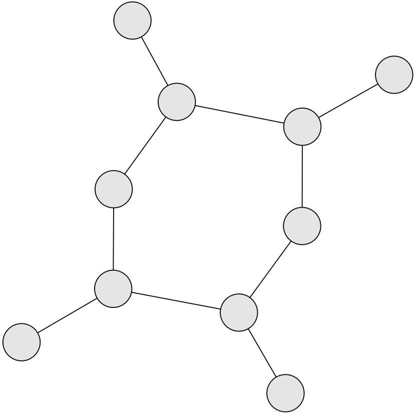


Bipartite

Sparsity



Dense



Sparse

Definitions

For a **growing graph sequence** $(\mathcal{Y}_n)_{n\in\mathbb{N}}$ (nodes and edges keep growing):

Dense

The number of edges grows **at least quadratically** with the nodes.

$$e(\mathcal{Y}_n) = \Omega(v(\mathcal{Y}_n)^2)$$

Sparse

The number of edges grows **sub-quadratically** with the nodes.

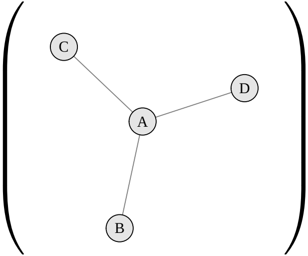
$$e(\mathcal{Y}_n) = o(v(\mathcal{Y}_n)^2)$$

Equivalently, the sequence is **sparse** if the density converges to zero:

$$d(\mathcal{Y}_n) = \frac{e(\mathcal{Y}_n)}{v(\mathcal{Y}_n)^2} \xrightarrow{n\to\infty} 0.$$

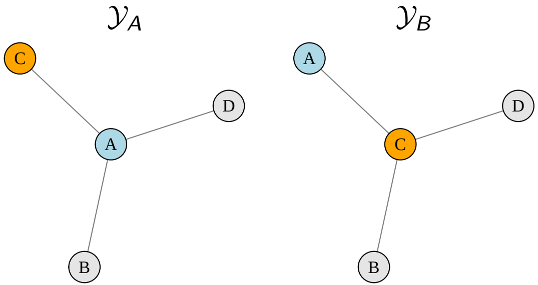
Network models represent probability distributions over the space of possible graphs.

$\mathsf{Pr}(\mathcal{Y}) = \mathsf{Pr}$



$= ?$

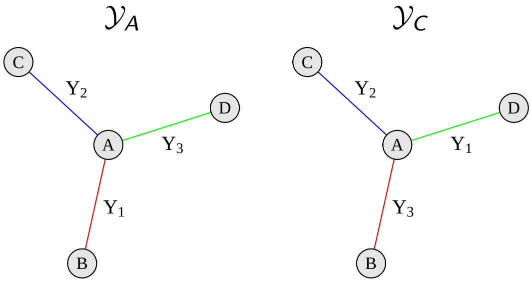
Node Exchangeable Models



Most network models in the literature assume **node exchangeability**, i.e. **invariance** of the distribution under permutation of nodes. Reasonable when **node labels** are uninformative.

$$\Pr(\mathcal{Y}_A) = \Pr(\mathcal{Y}_B)$$

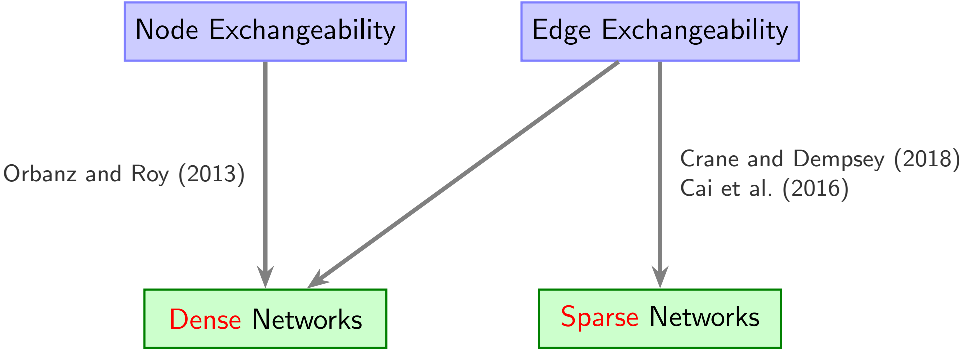
Edge Exchangeable Models



Edge exchangeability assumes **invariance** of the distribution under permutation of edges (the order of arrival).

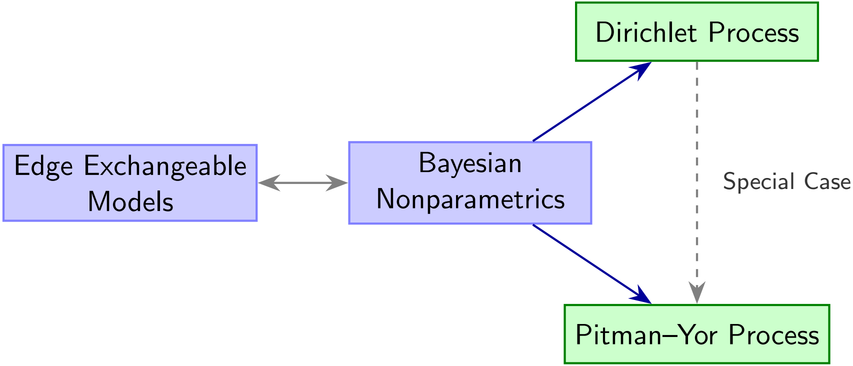
$$\Pr(\mathcal{Y}_A) = \Pr(\mathcal{Y}_C)$$

Implications of Exchangeability



Overview

Outline



Dirichlet Process
Edge Exchangeable Models
Bayesian Nonparametrics
Special Case
Pitman–Yor Process

The Dirichlet Process is a distribution over probability distributions (Ferguson, 1973).

$$G \sim \mathrm{DP}(\alpha, G_0)\,, \quad \alpha > 0$$

$$X_1, \ldots, X_n \mid G \overset{\text{i.i.d.}}{\sim} G,$$

where α is the **concentration** parameter, and G_0 is a **probability distribution** (on the same space Θ of the random variables X_i).

Property: G is discrete $\Rightarrow$ Samples from G contain **ties** with positive probability.

Chinese Restaurant Process

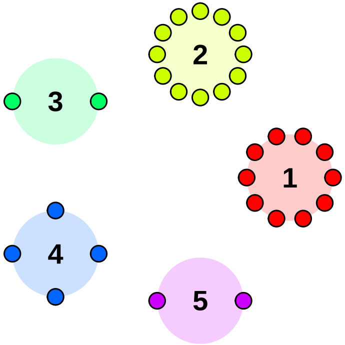
The probability of observing a new or an old value from a Dirichlet Process follows what is called the *Chinese Restaurant Process* (CRP).

Dirchlet Process CRP

Let n_k be the frequency of the k-th unique value in the sample $X_{1:n} = X_1, \ldots, X_n$.

$$\Pr(X_{n+1} = \text{Old k-th value} \mid X_{1:n}) = \frac{n_k}{\alpha + n}$$

$$\Pr(X_{n+1} = \text{New value} \mid X_{1:n}) = \frac{\alpha}{\alpha + n}$$



Pitman-Yor Process

The Pitman-Yor Process is an extension of the DP (Pitman and Yor, 1997).

$$G \sim \mathrm{PY}(\alpha, \sigma, G_0), \quad \sigma \in [0,1)$$
$$X_1, \ldots, X_n \mid G \overset{\text{i.i.d.}}{\sim} G.$$

Connection: $\sigma = 0 \Rightarrow$ PY and DP coincide.

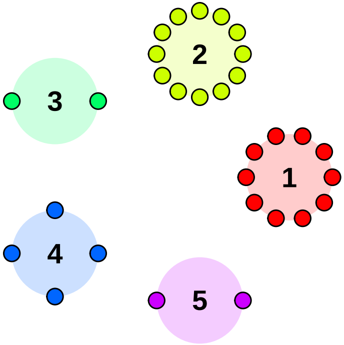
Pitman-Yor CRP Construction

Pitman-Yor Process CRP

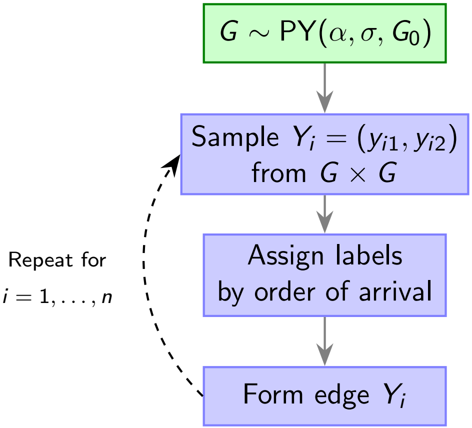
Let n_k be the frequency of the k-th unique value in the sample $X_{1:n} = X_1, \ldots, X_n$, and let K_n be the number of unique values.

$$\Pr(X_{n+1} = \text{Old k-th value} \mid X_{1:n}) = \frac{n_k - \sigma}{\alpha + n}$$

$$\Pr(X_{n+1} = \text{New value} \mid X_{1:n}) = \frac{\alpha + \sigma K_n}{\alpha + n}$$



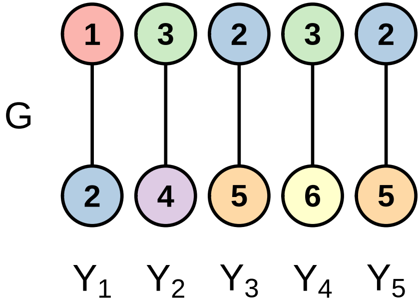
BNP Generative Model: Unipartite Case



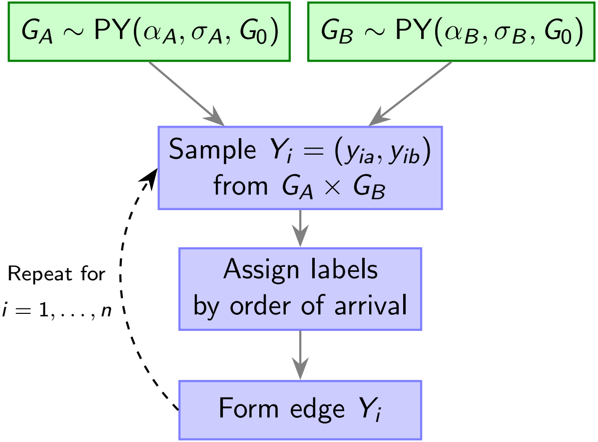
Hierarchical Model

$$G \sim \mathrm{PY}(\alpha, \sigma, G_0),$$

$$Y_1, \ldots, Y_n \mid G \overset{i.i.d.}{\sim} G \times G$$

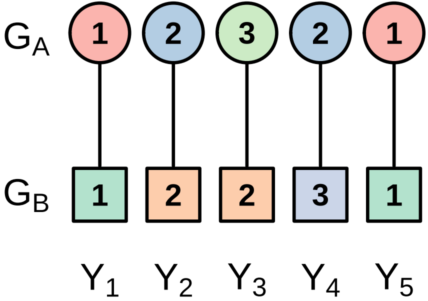


BNP Generative Model: Bipartite Case



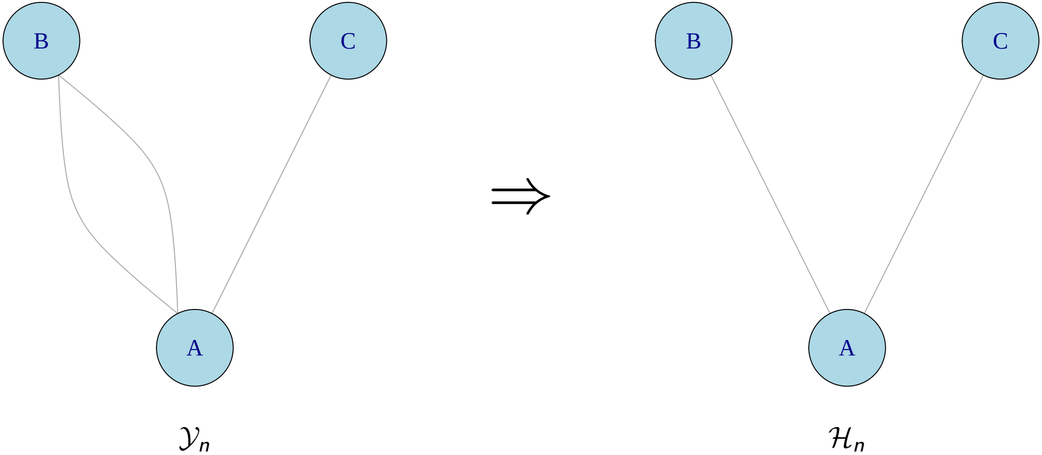
Hierarchical Model

$$G_A \sim \text{PY}(\alpha_A, \sigma_A, G_0)$$
$$G_B \sim \text{PY}(\alpha_B, \sigma_B, G_0)$$
$$Y_1, \ldots, Y_n \mid G_A, G_B \overset{i.i.d.}{\sim} G_A \times G_B$$



Overview

Binarisation



Sparsity in Binary Graphs with PY Model

Unipartite Model

$$d(\mathcal{H}_n) = \frac{e(\mathcal{H}_n)}{v(\mathcal{H}_n)^2}$$

Holds for $\sigma > 0$
Crane and Dempsey (2018)

Bipartite Model

$$d(\mathcal{H}_n) = \frac{e(\mathcal{H}_n)}{v_A(\mathcal{H}_n)v_B(\mathcal{H}_n)}$$

Does it hold?
For which range of σ_A and σ_B?

$$d(\mathcal{H}_n) \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

Proposition: Bipartite PY Model is Sparse

Proposition

The growing bipartite graph sequence $(\mathcal{H}_n)_{n\in\mathbb{N}}$ produced by the PY process is sparse:

$$d(\mathcal{H}_n) = \frac{e(\mathcal{H}_n)}{v_A(\mathcal{H}_n)v_B(\mathcal{H}_n)} \xrightarrow[n\to\infty]{\text{a.s.}} 0\,,$$

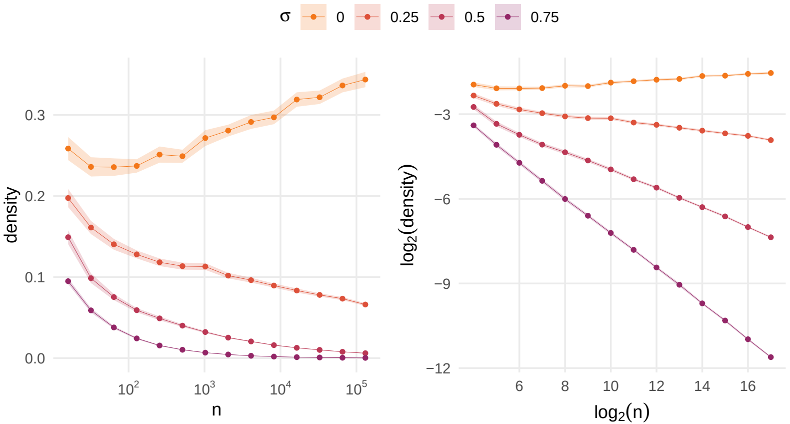
for either σ_A or σ_B in $(0,1)$.

The proof technique follows closely that of Crane and Dempsey (2018) for the unipartite case.

Remarks:

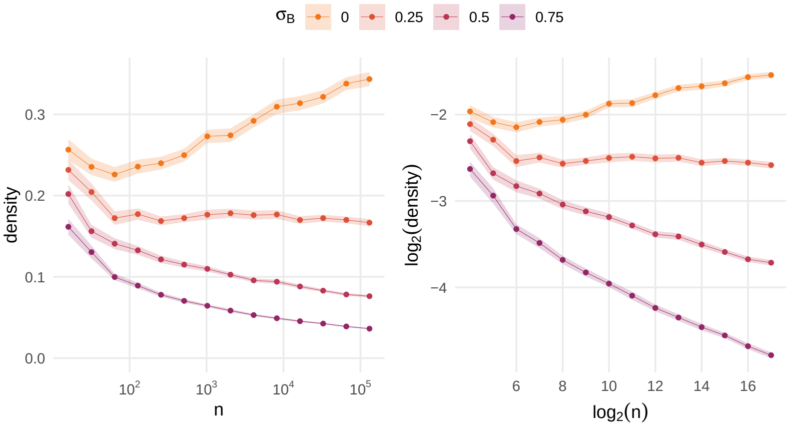
- If either $\sigma_A = 0$ or $\sigma_B = 0$, sparsity is preserved (Asymmetric DP $\times$ PY).
- If $\sigma_A = \sigma_B = 0$, sparsity is not guaranteed (Symmetric DP $\times$ DP)

Simulation Studies: Symmetric PY Model



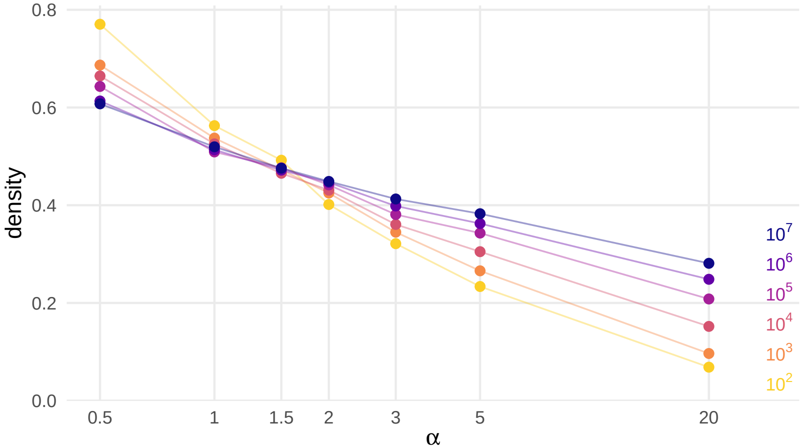
$$\mathrm{PY}(5, \sigma_A) \times \mathrm{PY}(5, \sigma_B)\,, \sigma_A = \sigma_B = \sigma$$

Simulation Studies: Asymmetric PY Model



$$\mathrm{DP}(5) \times \mathrm{PY}(5, \sigma_B)\,, \sigma_A = 0$$

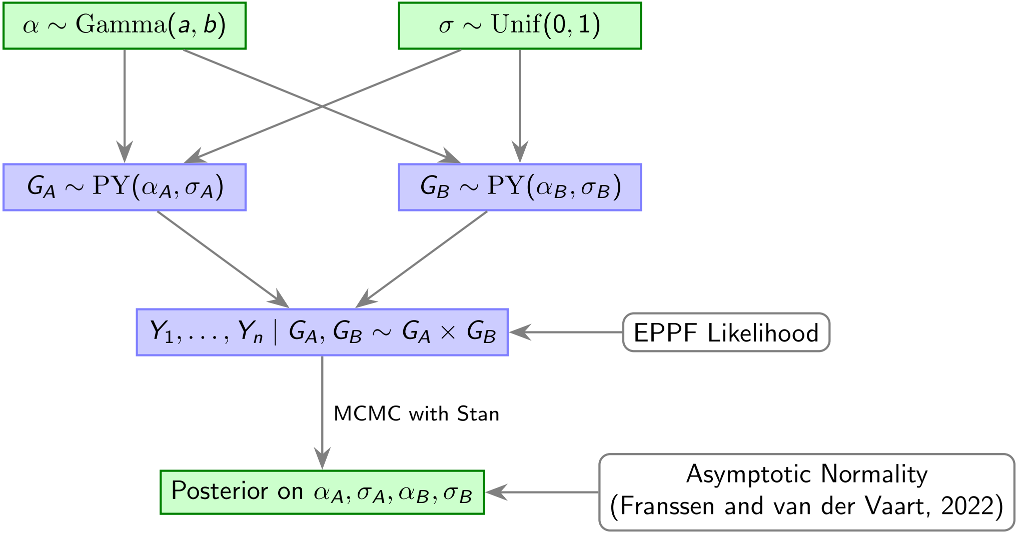
Simulation Studies: Symmetric DP Model



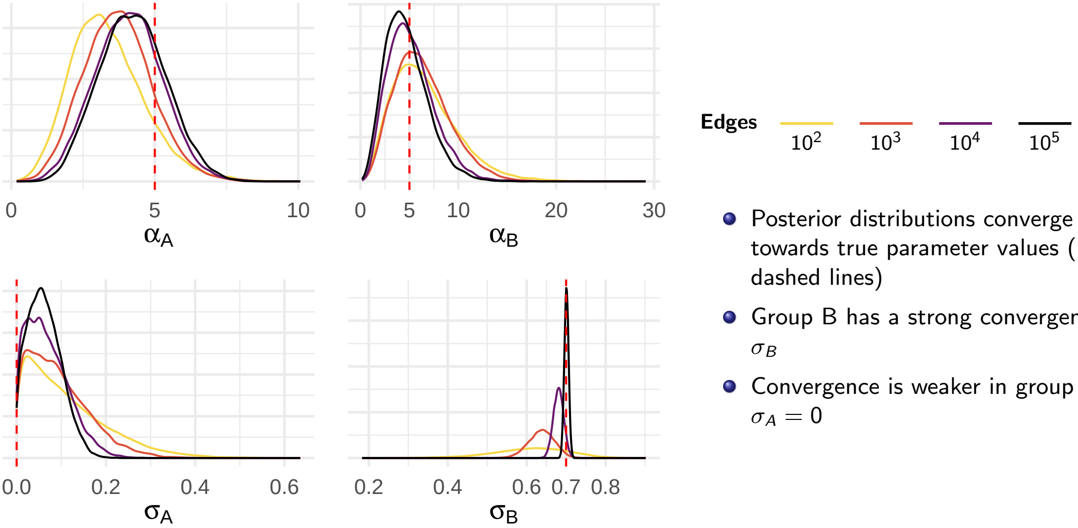
$$\mathrm{DP}(\alpha_A) \times \mathrm{DP}(\alpha_B), \alpha_A = \alpha_B = \alpha$$

Overview

Bayesian Model



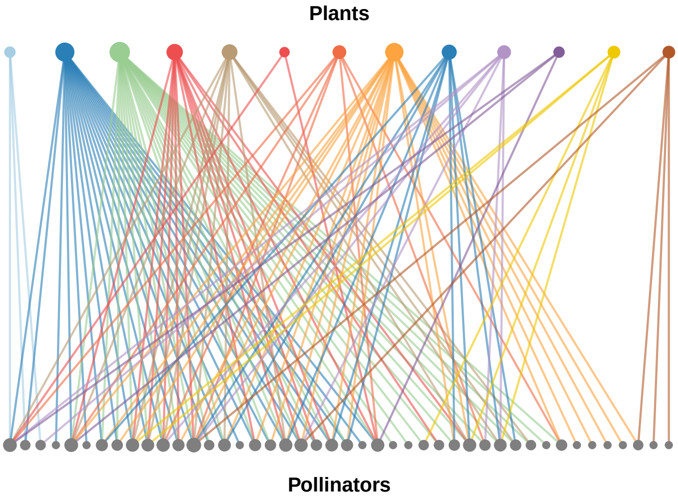
Posterior Convergence: DP(5) $\times$ PY(5, 0.7)



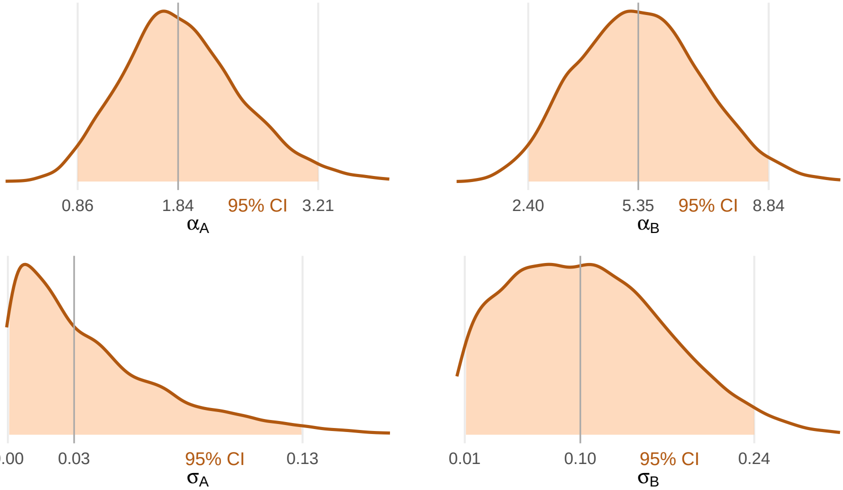
- Posterior distributions converge towards true parameter values (red dashed lines)
- Group B has a strong convergence for σ_B
- Convergence is weaker in group A with $\sigma_A = 0$

Real Data

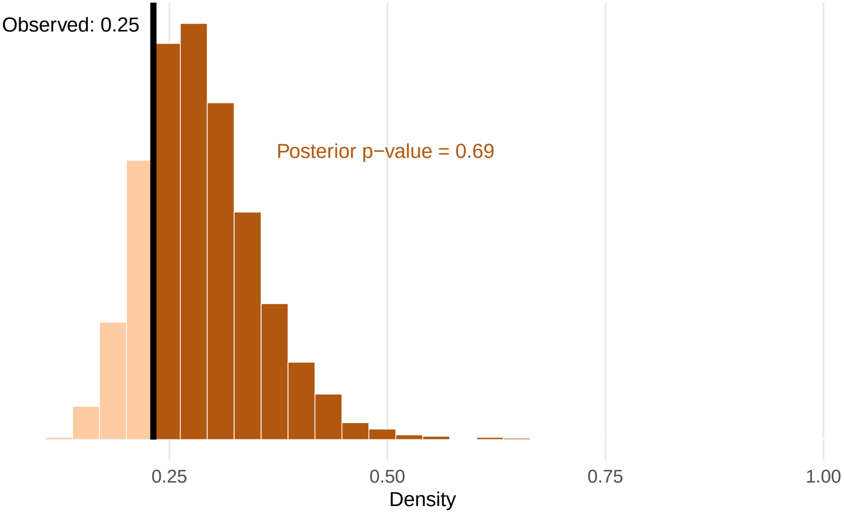
- Plants and Pollinators network data from forests in Piedmont North Carolina.
- Source: Web of Life. Original article: Motten (1986).
- 143 unique edges, 2225 observed interactions (sample size).
- 13 Plants (Group A)
  44 Pollinators (Group B).
- Density of 0.25.



Real Data: Posteriors



Real Data: Posterior Predictive Check



Test for Sparsity

We want to formally test if one group is inducing sparsity.

$$\mathrm{H}_0 : \sigma < \delta \quad \mathrm{H}_1 : \sigma > \delta \,, \quad \text{for } \delta \text{ close to } 0.$$

The **Bayes Factor** tells us what is the strength of the evidence in favour of H_0, as a ratio of the probabilities of observing the data under the null over the alternative hypothesis:

$$\underbrace{\frac{p(\sigma \mid Y_{1:n}, \mathrm{H}_0)}{p(\sigma \mid Y_{1:n}, \mathrm{H}_1)}}_{\text{Posterior Odds}} = \underbrace{\frac{p(Y_{1:n} \mid \sigma, \mathrm{H}_0)}{p(Y_{1:n} \mid \sigma, \mathrm{H}_1)}}_{\text{Bayes Factor}} \times \underbrace{\frac{p(\sigma \mid \mathrm{H}_0)}{p(\sigma \mid \mathrm{H}_1)}}_{\text{Prior Odds}} .$$

Bayes Factor on Real Data

H_0 :	$\sigma_A < \delta$		$\sigma_B < \delta$	
δ	BF	$\log_{10}$BF	BF	$\log_{10}$BF
0.01	26.69	1.43	3.55	0.55
0.05	47.54	1.68	5.33	0.73
0.10	117.90	2.07	8.85	0.95

Table: Bayes Factors for the discount parameters relative to threshold δ.

References

Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 4242–4250, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/1a0a283bfe7c549dee6c638a05200e32-Abstract.html.

Harry Crane and Walter Dempsey. Edge Exchangeable Models for Interaction Networks. *Journal of the American Statistical Association*, 113(523):1311–1326, 2018. doi: 10.1080/01621459.2017.1341413.

Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

S. E. M. P. Franssen and A. W. van der Vaart. Empirical and full bayes estimation of the type of a pitman-yor process. *arXiv preprint arXiv:2208.14255*, 2022.

Alexander F. Motten. Pollination Ecology of the Spring Wildflower Community of a Spring Wildflower Community. *Ecological Monographs*, 56(1):21–42, 1986. ISSN 00129615. URL http://www.jstor.org/stable/2937269.

Peter Orbanz and Daniel Roy. Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, 2013. doi: 10.1109/TPAMI.2014.2334607.

Jim Pitman and Marc Yor. The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25(2):855–900, 1997.