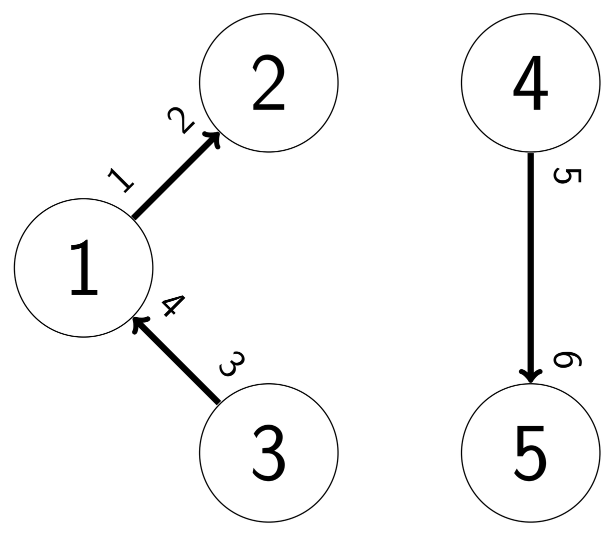
Sampling and Inference for Beta Neutral-to-the-Left Models of Sparse Networks

Benjamin Bloem-Reddy, Adam Foster, Emile Mathieu, Yee Whye Teh
Department of Statistics, University of Oxford

Random graphs

- Ends of edges $\mathbf{Z}_n = Z_1, ..., Z_n$
- Number of vertices K_n
- Arrival time of jth vertex $T_j := \inf\{n : Z_n = j\}$
- Degree of jth vertex $d_{j,n}$
- Degree counts $m_n(d) := |\{j : d_{j,n} = d\}|$

$$\mathbf{Z}_6 = \underline{1,2}, \underline{3,1}, \underline{4,5}$$



Sparsity means $K_n = O(n^{1/(1+\sigma)})$ for $0 \leq \sigma < 1$.

The asymptotic degree distribution has **power law tail with exponent** $\eta > 1$ if

$$\frac{m_n(d)}{K_n} \xrightarrow[n\to\infty]{p} L(d)d^{-\eta} \,, \tag{1}$$

for slowly varying function $L(d)$.

For sparse graphs, $\sigma = 0 \leftrightarrow \eta > 2$ and $\sigma > 0 \leftrightarrow \eta \in (1, 2)$.

Empirical properties of temporal networks

SNAP [1] includes real-world temporal networks. Ask Ubuntu dataset has 159,316 and 964,437 edges. Empirically $K_n = O(n)$ and $\hat{\eta} \approx 2.14$ (estimated using [2]).

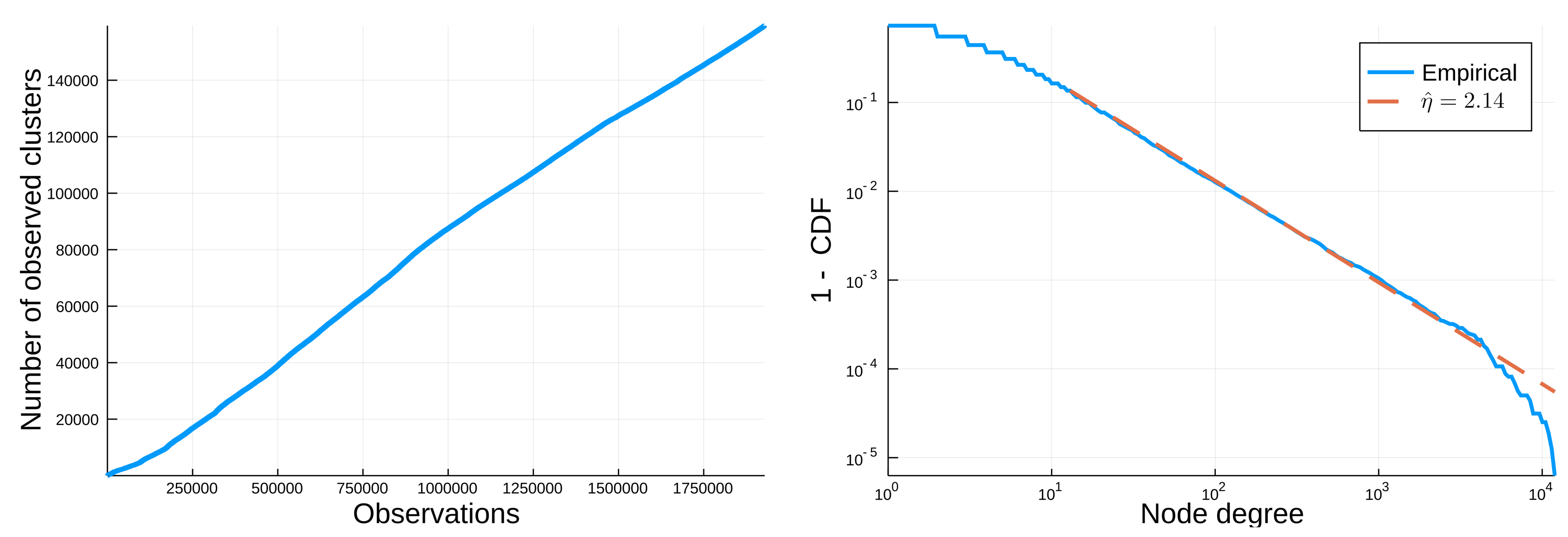


Figure 1: Ask Ubuntu arrival process (left) and node degree distribution (right)

Models

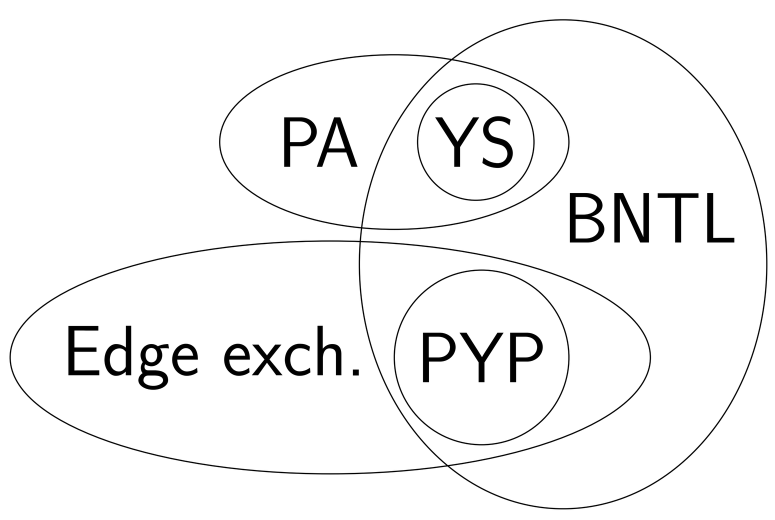
Edge exchangeable models [3] [4] include **Exchangeable Gibbs partitions** and **Pitman–Yor process** (PYP).

Preferential attachment (PA) models include **Yule–Simon** (YS) model.

A **Beta Neutral-to-the-left model** (BNTL) [5] is parameterized by $\alpha \in (-\infty, 1)$ and arrival distribution Λ_ϕ on $\mathbb{N}^\infty$. Distribution on $\mathbf{Z}$ is

$$(T_1, T_2, ...) \sim \Lambda_\phi \tag{2}$$

$$Z_{n+1}|\mathbf{Z}_n, \mathbf{T} = \begin{cases} K_{n+1} \text{ w.p. } 1 & \text{if } n+1 = T_{K_{n+1}} \\ j \text{ w.p.} \propto (d_{j,n} - \alpha) & \text{otherwise} \end{cases} \tag{3}$$



Network properties

	Growth rate	Degree exponent, η
Ask Ubuntu	Linear.	$\hat{\eta} = 2.14$
All SNAP datasets	Linear and sublinear.	$\hat{\eta} \in (1.5, 3)$
Edge exchangeable models	Sublinear. $K_n = o(n)$	$\eta \in (1, 2)$
Yule–Simon model	Linear. $\Delta_j \overset{\text{i.i.d.}}{\sim} \text{Geom}(\beta)$	$\eta \in (2, \infty)$
BNTL models	**T has law** Λ_ϕ	$\eta \in (1, \infty)$

Inference [6]

BNTL models have **tractable inference** due to the factorisations

$$\mathbb{P}_{\alpha,\phi}(\mathbf{Z}_n) = \mathbb{P}_\alpha(\mathbf{Z}_n|\mathbf{T}_{K_n})\Lambda_\phi(\mathbf{T}_{K_n}) \,, \tag{4}$$

$$\mathbb{P}_\alpha[G(\mathbf{Z}_n)|\mathbf{T}_{K_n}] = \frac{\Gamma(d_{1,n} - \alpha)}{\Gamma(n - K_n\alpha)} \prod_{j=2}^{K_n} \frac{\Gamma(T_j - j\alpha)\Gamma(d_{j,n} - \alpha)}{\Gamma(T_j - 1 - (j-1)\alpha)\Gamma(1 - \alpha)} \,, \tag{5}$$

in particular the degree sequence $\mathbf{d}_n := (d_{1,n}, \ldots, d_{K_n,n})$ is a sufficient statistic for α conditional on $\mathbf{T}_{K_n}$.

Observation	Unobserved variables
End of edge sequence $\mathbf{Z}_n$	$\alpha, \phi, \mathbf{\Psi}_{K_n}$
Vertex arrival-ordered graph	$\alpha, \phi, \mathbf{\Psi}_{K_n}, \mathbf{T}_{K_n}$
Unlabeled graph	$\alpha, \phi, \mathbf{\Psi}_{K_n}, \mathbf{T}_{K_n}, \sigma[K_n]$

Variable	Gibbs sampling scheme
α	Slice sampling
ϕ	Depends on family Λ_ϕ
	Conjugate updates possible e.g. $\Delta_j \sim \text{Geom}(\beta)$
$\mathbf{\Psi}_{K_n}$	$\Psi_j|\mathbf{Z}_n, \mathbf{\Psi}_{\backslash j} \sim \text{Beta}(d_{n,j} - \alpha, \bar{d}_{n,j-1} - (j-1)\alpha)$
	where $\bar{d}_{n,j} = \sum_{i=1}^{j} d_{j,n}$, marginalise if $\mathbf{Z}_n$ not observed
$\mathbf{T}_{K_n}$	Assume Markov structure
	Simple update for T_j – can't move past neighbours
$\sigma[K_n]$	Swap proposal probability is cheap

For massive graphs with $\mathbf{Z}_n$ observed, **maximum a posterior** (or **maximum likelihood**) estimates for α, ϕ computable from (4).

Experiments

Gibbs sampler accuracy on synthetic data (500 edges)

Gen. arrival distn.	Inference model	$\lvert\hat{\alpha} - \alpha^*\rvert$	$\lvert\hat{\mathbf{S}} - \mathbf{S}^*\rvert$	Pred. log-lik.
$\mathcal{PYP}(1.0, 0.75)$	$(\tau, \mathcal{PYP}(\theta, \tau))$	**0.046** $\pm$ **0.002**	**28.5** $\pm$ **0.7**	**-2637.0** $\pm$ **0.1**
$\mathcal{PYP}(1.0, 0.75)$	$(\alpha, \text{Geom}(\beta))$	0.049 $\pm$ 0.004	66.8 $\pm$ 1.2	-2660.5 $\pm$ 0.7
$\text{Geom}(0.25)$	$(\tau, \mathcal{PYP}(\theta, \tau))$	0.086 $\pm$ 0.002	56.6 $\pm$ 1.3	-2386.8 $\pm$ 0.1
$\text{Geom}(0.25)$	$(\alpha, \text{Geom}(\beta))$	**0.043** $\pm$ **0.003**	**24.8** $\pm$ **0.8**	**-2382.6** $\pm$ **0.2**

where $\mathbf{S} := \frac{1}{K_n - 1}\sum_{j>1}(\bar{d}_{j-1} - T_j)$

Scalability of Gibbs sampler (Geom(0.25) arrivals)

	$n = 200$	$n = 20000$
$\lvert\hat{\alpha} - \alpha^*\rvert$	0.12 $\pm$ 0.01	0.01 $\pm$ 0.00
$\lvert\hat{\beta} - \beta^*\rvert$	0.02 $\pm$ 0.00	0.00 $\pm$ 0.00
ESS	0.90 $\pm$ 0.04	0.75 $\pm$ 0.08
Runtime (s)	21 $\pm$ 0	2267 $\pm$ 2

Maximum likelihood parameter estimation for Ask Ubuntu

Coupled $\mathcal{PYP}(\theta, \alpha)$			Uncoupled $\mathcal{PYP}(\theta, \tau)$			Geom(β)		
$(\hat{\theta}, \hat{\alpha})$	$\hat{\eta}$	Pred. l-l.	$\hat{\alpha}$	$(\hat{\theta}, \hat{\tau})$	Pred. l-l.	$\hat{\beta}$	$\hat{\eta}$	Pred. l-l.
(18080, 0.25)	1.25	-3.707e6	-2.54	(-0.99, 0.99)	-3.678e6	0.083	2.32	**-3.678e6**

Future work

- Scale MCMC inference to larger networks
- Variational inference

Acknowledgements

YWT, BBR, EM's research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071. EM and YWT gratefully acknowledge Microsoft Research and EPSRC for partially funding EM's studentship. AF gratefully acknowledges funding from EPSRC grant no. EP/N509711/1.

References

[1] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

[2] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[3] Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, (just-accepted), 2017.

[4] Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems*, pages 4249–4257, 2016.

[5] Benjamin Bloem-Reddy and Peter Orbanz. Preferential attachment and vertex arrival times. *arXiv preprint arXiv:1710.02159*, 2017.

[6] Benjamin Bloem-Reddy, Adam Foster, Emile Mathieu, and Yee Whye Teh. Sampling and inference for beta neutral-to-the-left models of sparse networks. In *UAI (to appear)*, 2018.