

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



Contents

Background

Sampling and inference

Experiments

Temporal networks

Example

- ▶ Messages sent between people over time
- ▶ Protein-protein interactions

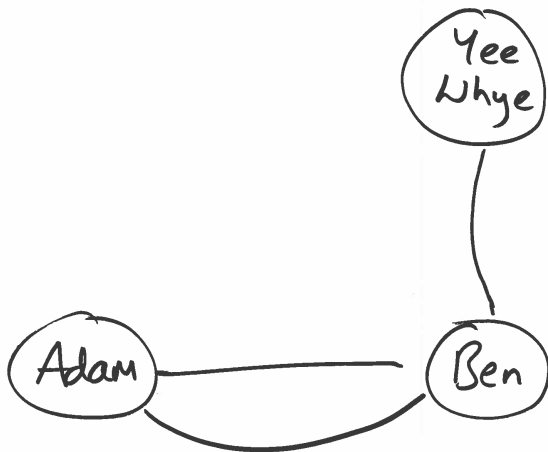
Temporal networks



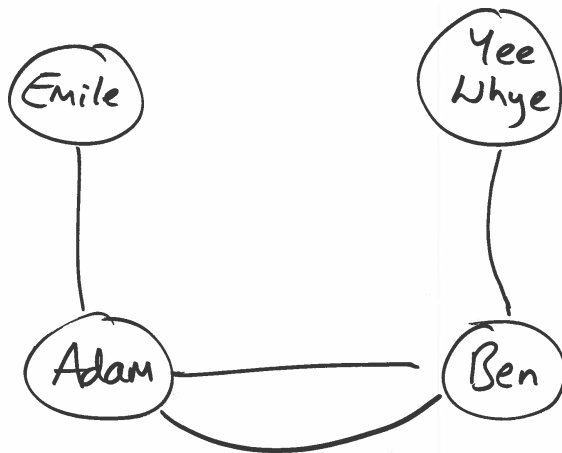
Temporal networks



Temporal networks

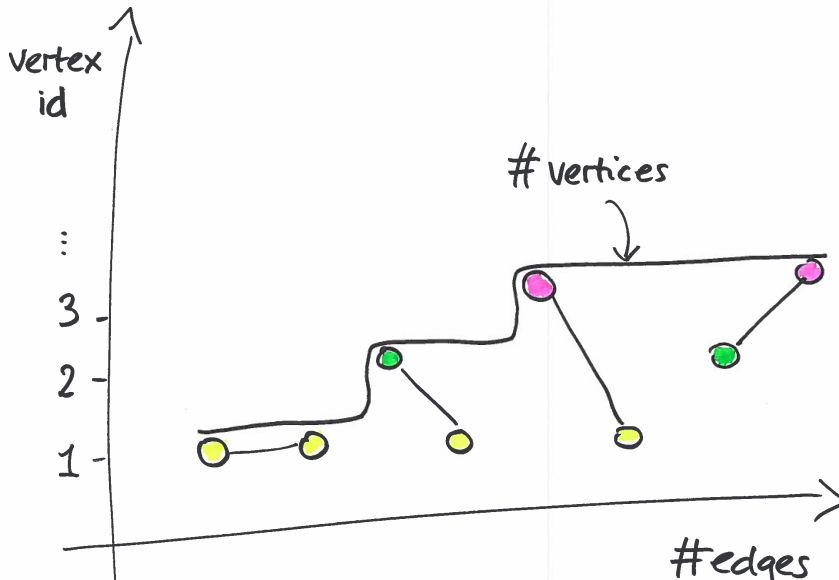


Temporal networks

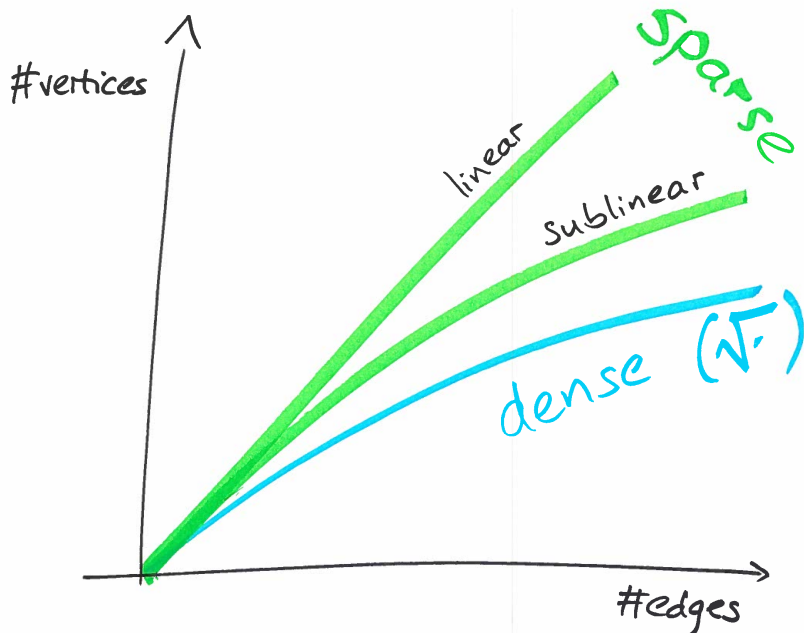


Edges and vertices

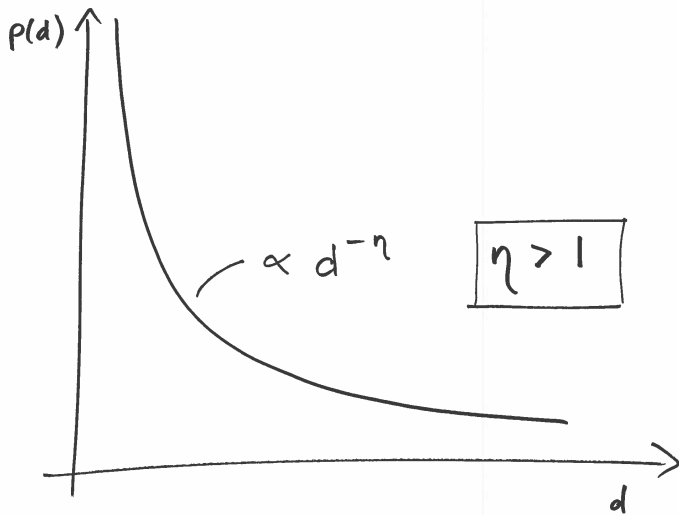
TODO: break picture into 4



Sparsity



Power law degree distribution



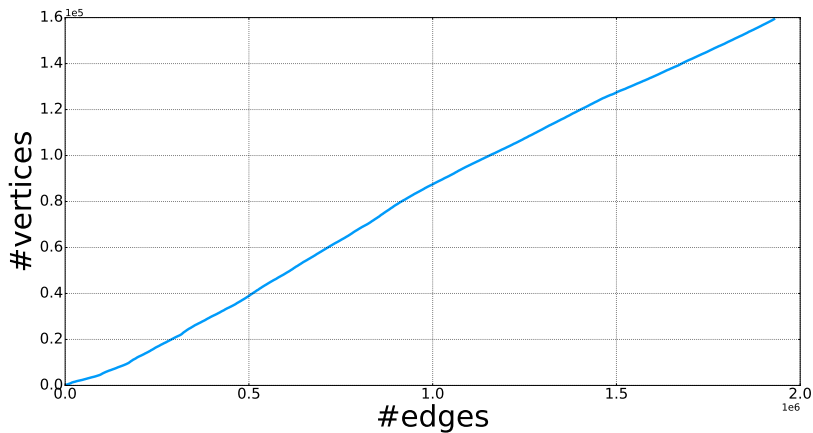
Sparsity and power law

Sublinear sparsity $\iff \eta \in (1, 2)$

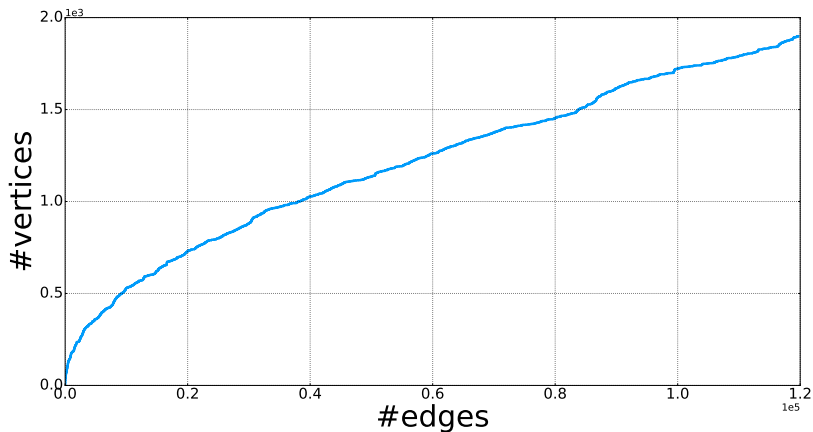
Linear sparsity $\iff \eta > 2$

Empirical study

SNAP dataset [1]	# of vertices	# of edges
Ask Ubuntu	159,316	964,437
UCI social network	1,899	20,296
⋮	⋮	⋮



UCI social network



Models



Vertex
exchangeable

Models

dense

sparse

Vertex
exchangeable

Models

dense

Vertex
exchangeable

sparse

Preferential
attachment

Models

dense

sparse

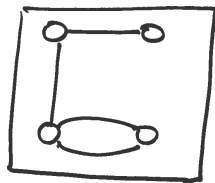
Vertex
exchangeable

Edge
exchangeable

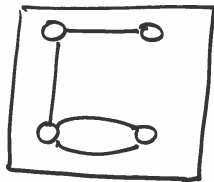
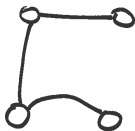
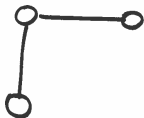
Pitman Yor

Preferential
attachment

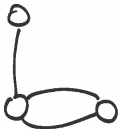
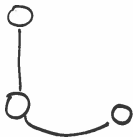
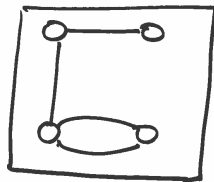
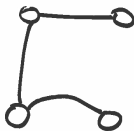
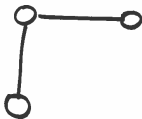
Edge exchangeable models [2], [3]



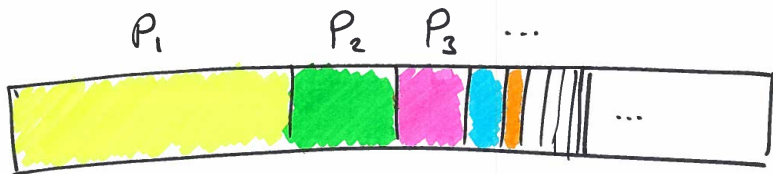
Edge exchangeable models [2], [3]



Edge exchangeable models [2], [3]



Paintbox representation

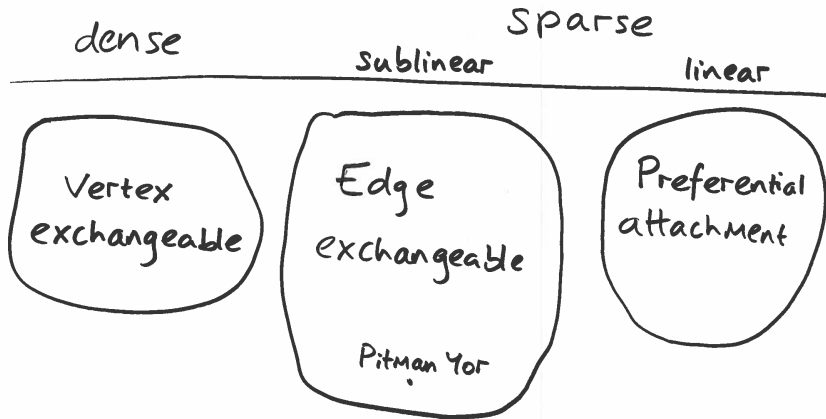


Paintbox representation

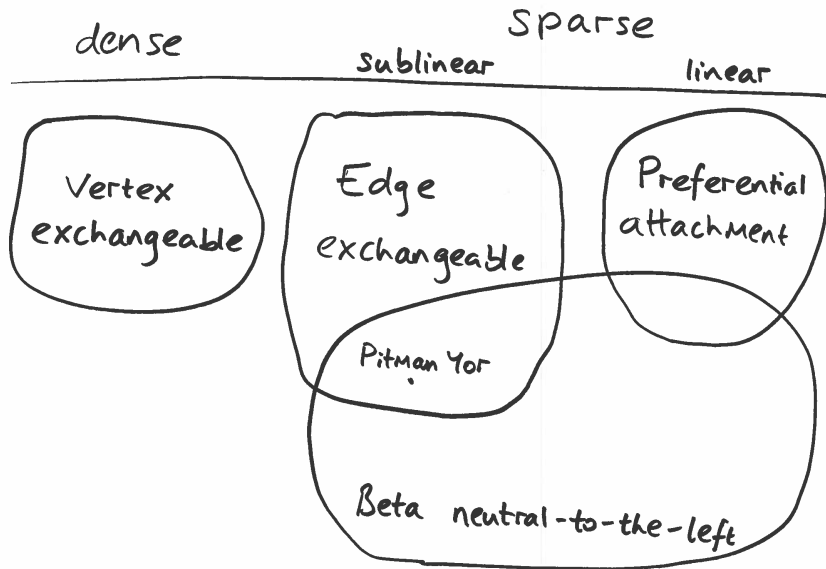
Consequence

- ▶ Edge exchangeable models have sublinear sparsity

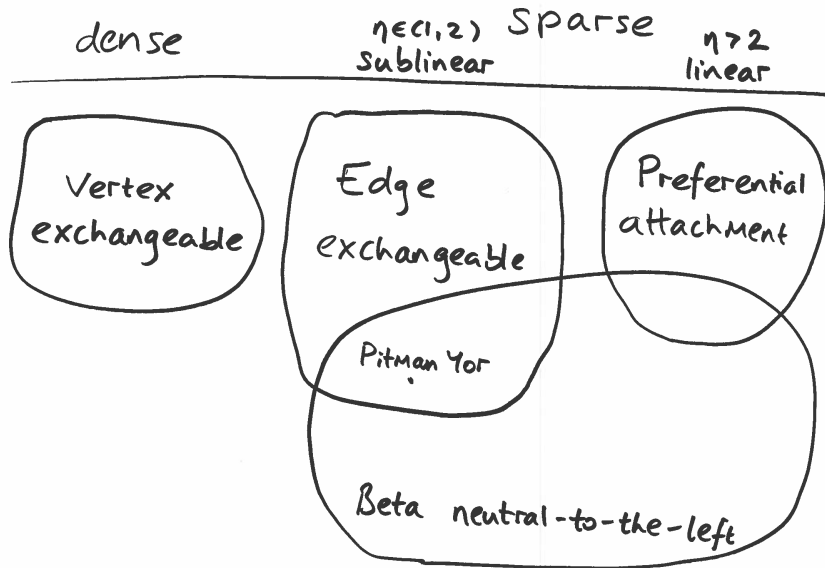
Models



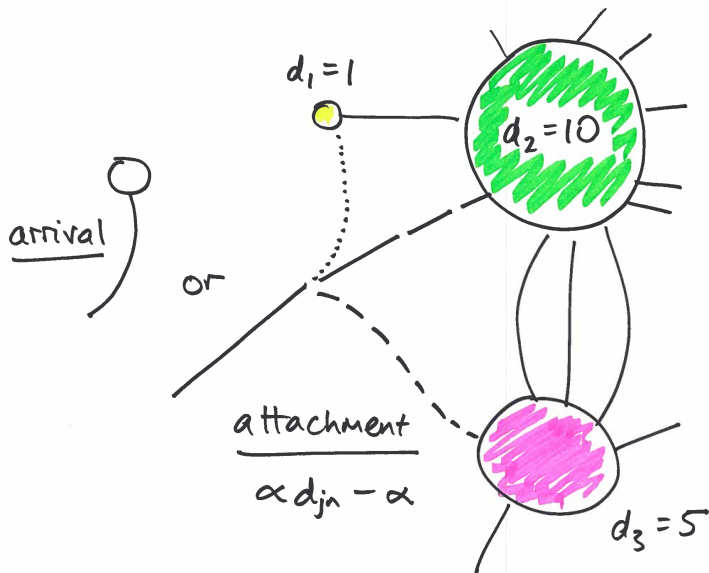
Models



Models



Beta Neutral-to-the-left Model [4]



Latent representation

$k = 1$ 

$k = 2$ 

$k = 3$ 

$k = 4$ 

Sampling and inference

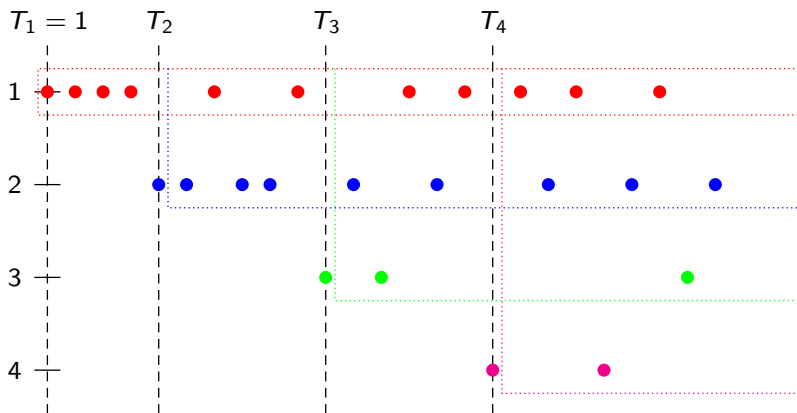
Why a paper on sampling and inference for BNTL models?

Sampling and inference

Why a paper on sampling and inference for BNTL models?

- ▶ Inference is notoriously difficult for *non-exchangeable* structures
- ▶ Need to identify *exchangeable substructures*

Exchangeable substructure



Gibbs structure

The joint density has **Gibbs structure**

$$p(\text{graph}|\mathbf{T}) = \prod_{j=1}^K p(\text{choose } j \text{ } d_j \text{ times out of } n - T_j)$$

- ▶ $K = \#\text{vertices}$
- ▶ $n = \#\text{edges}$
- ▶ $d_j = \text{degree of vertex } j$
- ▶ $T_j = \text{arrival time of vertex } j$

Gibbs structure

Explicitly

$$p(\text{graph}|\mathbf{T}) = \frac{\Gamma(d_1 - \alpha)}{\Gamma(n - K\alpha)} \prod_{j=2}^K \frac{\Gamma(T_j - j\alpha)\Gamma(d_j - \alpha)}{\Gamma(T_j - 1 - (j-1)\alpha)\Gamma(1 - \alpha)}$$

- ▶ $K = \#\text{vertices}$
- ▶ $n = \#\text{edges}$
- ▶ $d_j = \text{degree of vertex } j$
- ▶ $T_j = \text{arrival time of vertex } j$

Available data

Observation	Unobserved variables
Entire history	α, ϕ, Ψ
Degrees in arrival order	$\alpha, \phi, \Psi, \mathbf{T}$
Snapshot	$\alpha, \phi, \Psi, \mathbf{T}, \sigma$

- ▶ α = BTNL parameter $\in (-\infty, 1)$
- ▶ ϕ = arrival distribution parameters
- ▶ Ψ = latent sociabilities
- ▶ \mathbf{T} = arrival times
- ▶ σ = arrival order

Gibbs sampler

Variable	Gibbs sampling scheme
α	MCMC, e.g. slice sampling
ϕ	Depends on arrival dist. family Λ_ϕ
Ψ	$\Psi_j \text{graph}, \Psi_{\setminus j} \sim \text{Beta}(d_j - \alpha, \bar{d}_{j-1} - (j-1)\alpha)$ where $\bar{d}_j = \sum_{i=1}^j d_i$ can marginalise out Ψ
\mathbf{T}	Simple update for T_j , can't move past neighbours
σ	Initialise in descending degree order use M-H with adjacent swap proposal $\sigma_j \leftrightarrow \sigma_{j+1}$

Point estimation

If entire history observed, **maximum a posterior** (or **maximum likelihood**) estimates for α, ϕ computable

Experiments

- ▶ Parameter recovery
- ▶ Scaling in n
- ▶ Point estimation with massive graphs

Parameter recovery

- ▶ Simulate 500 edges with fixed α
- ▶ Arrivals either \mathcal{PYP} or Geom
- ▶ Observe final snapshot of the graph

Gen. arrival distn.	Inference model	$ \hat{\alpha} - \alpha^* $	Pred. log-lik.
$\mathcal{PYP}(1.0, 0.75)$	$(\tau, \mathcal{PYP}(\theta, \tau))$	0.046 ± 0.002	-2637.0 ± 0.1
$\mathcal{PYP}(1.0, 0.75)$	$(\alpha, \text{Geom}(\beta))$	0.049 ± 0.004	-2660.5 ± 0.7
Geom(0.25)	$(\tau, \mathcal{PYP}(\theta, \tau))$	0.086 ± 0.002	-2386.8 ± 0.1
Geom(0.25)	$(\alpha, \text{Geom}(\beta))$	0.043 ± 0.003	-2382.6 ± 0.2

Scalability of Gibbs sampler

- Simulate with fixed α and $\text{Geom}(\beta)$ arrivals

	100 edges	10000 edges
$ \hat{\alpha} - \alpha^* $	0.12 ± 0.01	0.01 ± 0.00
$ \hat{\beta} - \beta^* $	0.02 ± 0.00	0.00 ± 0.00
Effective Sample Size	0.90 ± 0.04	0.75 ± 0.08
Runtime (s)	21 ± 0	2267 ± 2

- Most expensive Gibbs update is for \mathbf{T}

MLEs for SNAP datasets

Fitted point estimates

Dataset	Coupled $\mathcal{PYP}(\theta, \alpha)$			$\hat{\alpha}$	Uncoupled $\mathcal{PYP}(\theta, \tau)$		Geom(β)		
	$(\hat{\theta}, \hat{\alpha})$	$\hat{\eta}$	Pred. I-I.		$(\hat{\theta}, \hat{\tau})$	Pred. I-I.	$\hat{\beta}$	$\hat{\eta}$	Pred. I-I.
Ask Ubuntu	(18080, 0.25)	1.25	-3.707e6	-2.54	(-0.99, 0.99)	-3.678e6	0.083	2.32	-3.678e6
UCI social network	(320.4, 4.4e-11)	—	-1.600e5	-4.98	(5.50, 0.52)	-1.595e6	0.016	2.10	-1.596e5
EU email	(113.6, 2.5e-14)	—	-8.06e5	-1.86	(113.6, 9.2e-10)	-8.06e5	0.001	2.00	-8.07e5
Math Overflow	(2575, 0.15)	1.15	-1.685e6	-6.62	(-0.97, 0.997)	-1.670e6	0.025	2.19	-1.670e6
Stack Overflow	(297600, 0.11)	1.11	-3.358e8	-8.94	(-1.0, 1.0)	-3.333e8	0.020	2.21	-3.333e8
Super User	(20640, 0.24)	1.24	-5.855e6	-4.19	(-0.996, 1.0)	-5.775e6	0.067	2.37	-5.775e6
Wikipedia talk pages	(14870, 0.54)	1.54	-3.074e7	-0.25	(-1.0, 1.0)	-3.066e7	0.073	2.10	-3.066e7

MLEs for SNAP datasets

\mathcal{PYP} parameter estimates vary coupled and uncoupled

Dataset	Coupled $\mathcal{PYP}(\theta, \alpha)$			$\hat{\alpha}$	Uncoupled $\mathcal{PYP}(\theta, \tau)$		Geom(β)		
	$(\hat{\theta}, \hat{\alpha})$	$\hat{\eta}$	Pred. I-I.		$(\hat{\theta}, \hat{\tau})$	Pred. I-I.	$\hat{\beta}$	$\hat{\eta}$	Pred. I-I.
Ask Ubuntu	(18080, 0.25)	1.25	-3.707e6	-2.54	(-0.99, 0.99)	-3.678e6	0.083	2.32	-3.678e6
UCI social network	(320.4, 4.4e-11)	–	-1.600e5	-4.98	(5.50, 0.52)	-1.595e6	0.016	2.10	-1.596e5
EU email	(113.6, 2.5e-14)	–	-8.06e5	-1.86	(113.6, 9.2e-10)	-8.06e5	0.001	2.00	-8.07e5
Math Overflow	(2575, 0.15)	1.15	-1.685e6	-6.62	(-0.97, 0.997)	-1.670e6	0.025	2.19	-1.670e6
Stack Overflow	(297600, 0.11)	1.11	-3.358e8	-8.94	(-1.0, 1.0)	-3.333e8	0.020	2.21	-3.333e8
Super User	(20640, 0.24)	1.24	-5.855e6	-4.19	(-0.996, 1.0)	-5.775e6	0.067	2.37	-5.775e6
Wikipedia talk pages	(14870, 0.54)	1.54	-3.074e7	-0.25	(-1.0, 1.0)	-3.066e7	0.073	2.10	-3.066e7

MLEs for SNAP datasets

Edge exchangeable models likely misspecified

Dataset	Coupled $\mathcal{PYP}(\theta, \alpha)$			$\hat{\alpha}$	Uncoupled $\mathcal{PYP}(\theta, \tau)$		Geom(β)		
	$(\hat{\theta}, \hat{\alpha})$	$\hat{\eta}$	Pred. I-I.		$(\hat{\theta}, \hat{\tau})$	Pred. I-I.	$\hat{\beta}$	$\hat{\eta}$	Pred. I-I.
Ask Ubuntu	(18080, 0.25)	1.25	-3.707e6	-2.54	(-0.99, 0.99)	-3.678e6	0.083	2.32	-3.678e6
UCI social network	(320.4, 4.4e-11)	–	-1.600e5	-4.98	(5.50, 0.52)	-1.595e6	0.016	2.10	-1.596e5
EU email	(113.6, 2.5e-14)	–	-8.06e5	-1.86	(113.6, 9.2e-10)	-8.06e5	0.001	2.00	-8.07e5
Math Overflow	(2575, 0.15)	1.15	-1.685e6	-6.62	(-0.97, 0.997)	-1.670e6	0.025	2.19	-1.670e6
Stack Overflow	(297600, 0.11)	1.11	-3.358e8	-8.94	(-1.0, 1.0)	-3.333e8	0.020	2.21	-3.333e8
Super User	(20640, 0.24)	1.24	-5.855e6	-4.19	(-0.996, 1.0)	-5.775e6	0.067	2.37	-5.775e6
Wikipedia talk pages	(14870, 0.54)	1.54	-3.074e7	-0.25	(-1.0, 1.0)	-3.066e7	0.073	2.10	-3.066e7

MLEs for SNAP datasets

Though better than Geom for some datasets

Dataset	Coupled $\mathcal{PYP}(\theta, \alpha)$			$\hat{\alpha}$	Uncoupled $\mathcal{PYP}(\theta, \tau)$		Geom(β)		
	$(\hat{\theta}, \hat{\alpha})$	$\hat{\eta}$	Pred. I-I.		$(\hat{\theta}, \hat{\tau})$	Pred. I-I.	$\hat{\beta}$	$\hat{\eta}$	Pred. I-I.
Ask Ubuntu	(18080, 0.25)	1.25	-3.707e6	-2.54	(-0.99, 0.99)	-3.678e6	0.083	2.32	-3.678e6
UCI social network	(320.4, 4.4e-11)	–	-1.600e5	-4.98	(5.50, 0.52)	-1.595e6	0.016	2.10	-1.596e5
EU email	(113.6, 2.5e-14)	–	-8.06e5	-1.86	(113.6, 9.2e-10)	-8.06e5	0.001	2.00	-8.07e5
Math Overflow	(2575, 0.15)	1.15	-1.685e6	-6.62	(-0.97, 0.997)	-1.670e6	0.025	2.19	-1.670e6
Stack Overflow	(297600, 0.11)	1.11	-3.358e8	-8.94	(-1.0, 1.0)	-3.333e8	0.020	2.21	-3.333e8
Super User	(20640, 0.24)	1.24	-5.855e6	-4.19	(-0.996, 1.0)	-5.775e6	0.067	2.37	-5.775e6
Wikipedia talk pages	(14870, 0.54)	1.54	-3.074e7	-0.25	(-1.0, 1.0)	-3.066e7	0.073	2.10	-3.066e7

Conclusion

- ▶ BNTL models are *flexible*
- ▶ BNTL models are *tractable*

Future work

- ▶ Scalability of inference
 - ▷ Metropolis-Hastings to update \mathbf{T} altogether
 - ▷ Variational inference for σ

References

- [1] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- [2] Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems*, pages 4249–4257, 2016.
- [3] Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, (just-accepted), 2017.
- [4] Benjamin Bloem-Reddy and Peter Orbanz. Preferential attachment and vertex arrival times. *arXiv preprint arXiv:1710.02159*, 2017.