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

Temporal networks

Examples

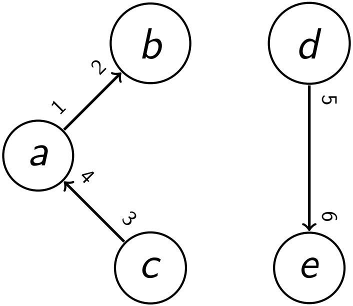
- Messages on WhatsApp
- Posts + replies on StackOverflow

Abstraction

- Graph grows adding one edge (Z_i, Z_{i+1}) at a time
- Vertices enter the graph when connected to

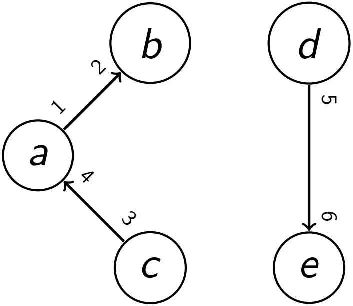
Temporal networks

Ends of edges $\mathbf{Z}_n = Z_1, ..., Z_n$
E.g. $\mathbf{Z}_6 = \underline{a, b}, \underline{c, a}, \underline{d, e}$



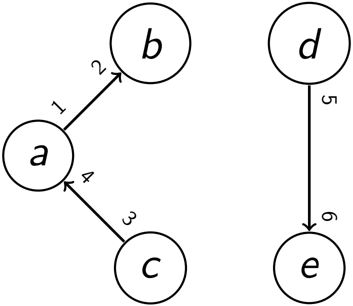
Temporal networks

Number of vertices K_n
E.g. $K_6 = 5$



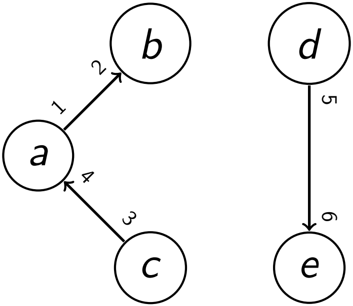
Temporal networks

Arrival time of vertex j is $T_j := \inf\{n : Z_n = j\}$
E.g. $T_e = 6$



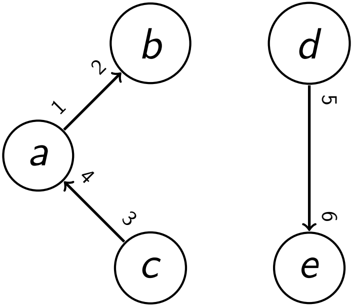
Temporal networks

Degree of vertex j is $d_{j,n}$
E.g. $d_{e,6} = 1$



Temporal networks

Degree counts $m_n(d) := |\{j : d_{j,n} = d\}|$
E.g. $m_6(1) = 4, m_6(2) = 1$



Sparsity

- For a dense graph, $K_n = O(n^{1/2})$
- For a sparse graph,

$$K_n = O(n^{1/(1+\sigma)})$$

  for $0 \leq \sigma < 1$
- Stack Overflow network likely sparse

Power law degree distribution

A power law distribution of exponent η on $\{1, 2, ...\}$ has

$$p(d) \propto d^{-\eta}$$

where $\eta > 1$.

Power law degree distribution

A power law distribution of exponent η on $\{1, 2, ...\}$ has

$$p(d) \propto d^{-\eta}$$

where $\eta > 1$.
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 $L(d)$.

Power laws and sparsity

We have

$$K_n = \sum_{d=1}^{\infty} m_n(d),$$
$$n = \sum_{d=1}^{\infty} d\, m_n(d).$$

Suppose (somewhat informally) that

$$K_n = C_{n,\eta} \sum_{d=1}^{n} d^{-\eta}$$

Power laws and sparsity

We have

$$K_n = \sum_{d=1}^{\infty} m_n(d),$$
$$n = \sum_{d=1}^{\infty} d\, m_n(d).$$

Suppose (somewhat informally) that

$$K_n = C_{n,\eta} \sum_{d=1}^{n} d^{-\eta}$$

then

$$n = C_{n,\eta} \sum_{d=1}^{n} d^{-\eta+1} = K_n \frac{\sum_{d=1}^{n} d^{-\eta+1}}{\sum_{d=1}^{n} d^{-\eta}}$$

Power laws and sparsity

Letting $n \to \infty$ in

$$\frac{K_n}{n} = \frac{\sum_{d=1}^{n} d^{-\eta}}{\sum_{d=1}^{n} d^{-\eta+1}}$$

we see $K_n = O(n)$ if $\eta > 2$, $K_n = o(n)$ if $\eta \in (1, 2]$.

Power laws and sparsity

Letting $n \to \infty$ in

$$\frac{K_n}{n} = \frac{\sum_{d=1}^{n} d^{-\eta}}{\sum_{d=1}^{n} d^{-\eta+1}}$$

we see $K_n = O(n)$ if $\eta > 2$, $K_n = o(n)$ if $\eta \in (1, 2]$.

Summary

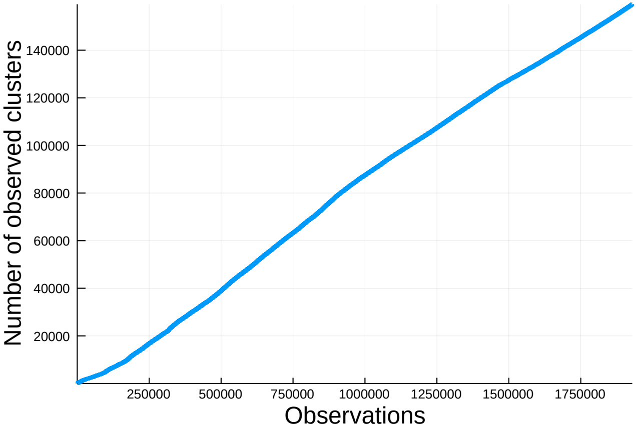
For sparse graphs, $\sigma = 0 \leftrightarrow \eta > 2$ and $\sigma > 0 \leftrightarrow \eta \in (1, 2]$.

Empirical study

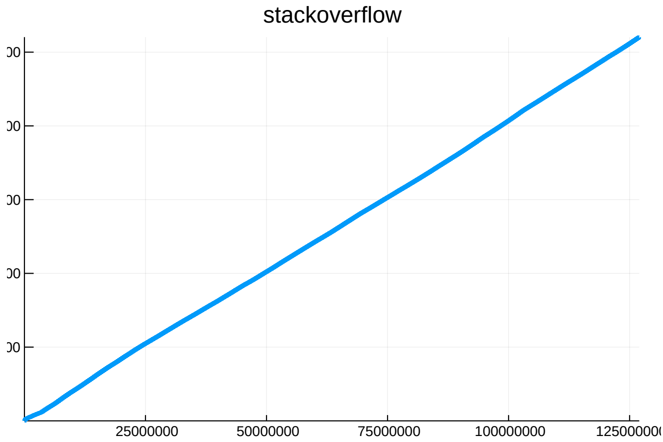
SNAP datasets [1]

Dataset	# of vertices	# of edges
Ask Ubuntu	159,316	964,437
UCI social network	1,899	20,296
EU email	986	332,334
Math Overflow	24,818	506,550
Stack Overflow	2,601,977	63,497,050
Super User	194,085	1,443,339
Wikipedia talk pages	1,140,149	7,833,140

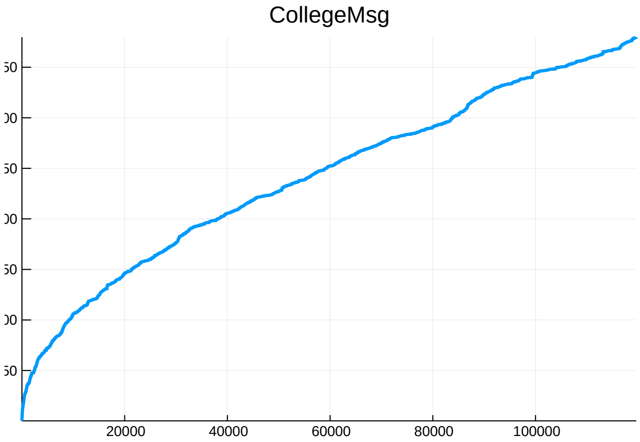
Ask Ubuntu arrival process



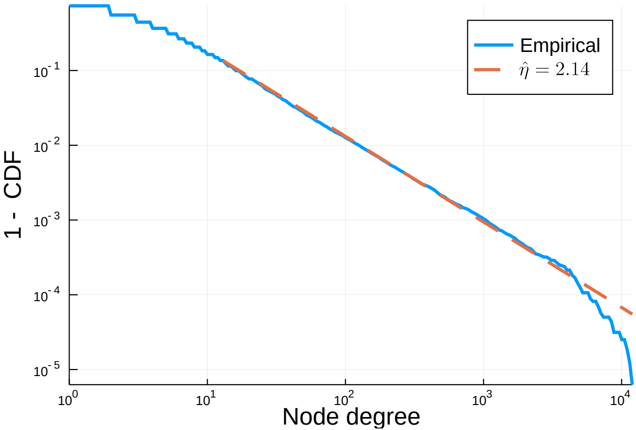
Stack Overflow arrival process



UCI social network arrival process



Ask Ubuntu degree distribution



Estimation used technique of [2]

Todo

- Recompile the images with better labels on the axes
- Estimate σ by linear regression

Models

- Vertex exchangeable models do not give sparsity [3] [4]
- Exchangeable point process models [5] have an independent notion of time
- **Preferential attachment models** [6]
- **Edge exchangeable models** [7] [8]

Yule-Simon Process

Parameter $\beta \in (0, 1)$.

Arrivals

$$T_{j+1} - T_j \overset{\text{i.i.d.}}{\sim} \text{Geom}(\beta) \tag{2}$$

Size-biased reinforcement

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

Yule-Simon Process

Asymptotic power law degree distribution with

$$\eta = 1 + \frac{1}{1-\beta} > 2$$

and $K_n = O(n)$

Pitman-Yor Process

Parameters $\tau \in (0,1), \theta > -\tau$.

Urn process

$$Z_{n+1} | \mathbf{Z}_n = \begin{cases} K_{n+1} \text{ w.p. } \dfrac{\theta + K_n \tau}{n + \theta} \\ \\ j \text{ w.p. } \dfrac{d_{j,n} - \tau}{\theta + n} \end{cases} \tag{4}$$

Pitman-Yor Process

Asymptotic power law degree distribution with

$$\eta = 1 + \tau \in (1, 2)$$

and $K_n = o(n)$

Edge exchangeable models [8], [7]

"The probability of all orderings of edge arrivals is the same"

$\eta \in (1, 2)$
$\exists$ a class of models that includes (some) edge exchangeable models, but also YS and admits all the ηs

Rewriting the Pitman-Yor Process

Parameters $\tau \in (0,1), \theta > -\tau$.

Arrivals

$$\mathbb{P}(T_{j+1} - T_j > t \mid T_j) = \prod_{i=1}^{t} \frac{T_j + t - j\tau}{T_j + t + \theta} \tag{5}$$

Size-biased reinforcement

$$Z_{n+1} | \mathbf{Z}_n, \mathbf{T} = \begin{cases} K_{n+1} \text{ w.p. } 1 & \text{if } n+1 = T_{K_{n+1}} \\ j \text{ w.p. } \propto (d_{j,n} - \tau) & \text{otherwise} \end{cases} \tag{6}$$

Beta Neutral-to-the-left Process [9]

Parameters $\alpha \in (-\infty, 1)$ and Λ_ϕ a law on $\mathbb{N}^\infty$.

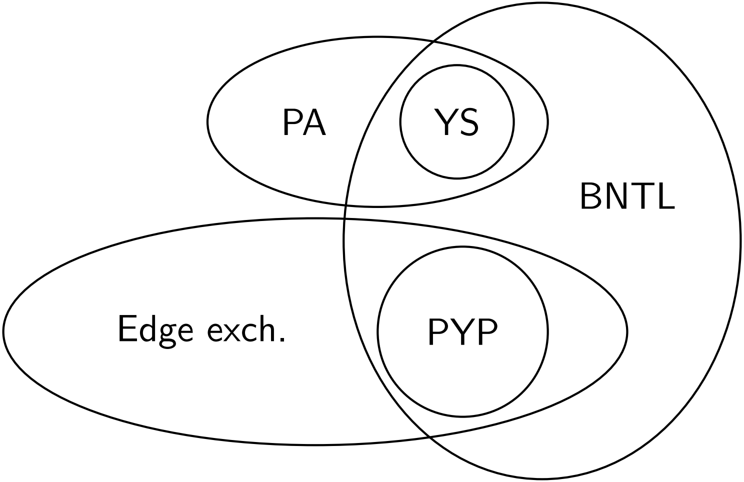
Arrivals

$$\mathbf{T} \sim \Lambda_\phi \tag{7}$$

Size-biased reinforcement

$$Z_{n+1} | \mathbf{Z}_n, \mathbf{T} = \begin{cases} K_{n+1} \text{ w.p. } 1 & \text{if } n+1 = T_{K_{n+1}} \\ j \text{ w.p. } \propto (d_{j,n} - \alpha) & \text{otherwise} \end{cases} \tag{8}$$

Relationship with other model classes



Hierarchical representation of BNTL process

Arrivals

$$\mathbf{T} \sim \Lambda_\phi$$

Latent sociabilities

$$\Psi_j | T_j \sim \text{Beta}(1-\alpha, T_j - 1 - (j-1)\alpha) \text{ for } j \geq 1$$

Left-neutral resampling probabilities

$$P_{j,k+1} = \begin{cases} P_{j,k}(1-\Psi_{k+1}), & j \in \{1,\ldots,k\} \\ \Psi_{k+1}, & j = k+1 \end{cases}$$

Sampling rule

$$Z_{n+1} | \mathbf{P}_{K_n}, \mathbf{T} = \begin{cases} K_{n+1} \text{ w.p. } 1 & \text{if } n+1 = T_{K_{n+1}} \\ j \text{ w.p. } P_{j,K_n} & \text{otherwise} \end{cases}$$

Observation cases

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	$\alpha, \phi, \mathbf{\Psi}_{K_n}, \mathbf{T}_{K_n}, \sigma[K_n]$

Sampling $\boldsymbol{\Psi}$

If $\mathbf{Z}_n$ is observed we have a closed form for $\boldsymbol{\Psi}$

$$p_{\alpha,\phi}(\boldsymbol{\Psi}_{K_n}|\mathbf{Z}_n) = \frac{p_{\alpha,\phi}(\boldsymbol{\Psi}_{K_n}, \mathbf{Z}_n)}{p_{\alpha,\phi}(\mathbf{Z}_n)},$$

furthermore

$$\Psi_j \mid \mathbf{Z}_n, \boldsymbol{\Psi}_{\setminus j} \sim \mathrm{Beta}(d_{j,n} - \alpha, \bar{d}_{j-1,n} - (j-1)\alpha) \ ,$$

where

$$\bar{d}_{j,n} = \sum_{i=1}^{j} d_{j,n}.$$

Talk more about degree being suff stats – insert $p(d, \Psi)$

- For fixed α, we have a closed form posterior
- Learning other variables, we have a closed form Gibbs update

Sampling α, ϕ

- Place priors on α, ϕ
- Left with one-dimensional unnormalized density for α and MCMC is applicable
- For ϕ, depends on Λ_ϕ. Our experiments used conjugacy or slice sampling.

Sampling $\mathbf{T}$

Assume

$$\boldsymbol{\Lambda}^{\phi}(\mathbf{T}_k) = \delta_{T_1}(1) \prod_{j=2}^{k} \Lambda_j^{\phi}(\Delta_j | T_{j-1}) \ ,$$

Support of $T_j - T_{j-1} | T_{\setminus j}$ is

$$\{1, ..., \min(T_{j+1} - T_{j-1} - 1, \bar{d}_{j-1} - T_{j-1} + 1)\}$$

Using our assumptions we can compute each conditional probability

Sampling $\sigma[K_n]$

Use Metropolis-Hastings with swap proposal $\sigma_j \leftrightarrow \sigma j + 1$
Ratio of joints from ...

Point estimation

Factorization $p_{\alpha,\phi}(\mathbf{Z}_n) = p_\alpha(\mathbf{Z}_n|\mathbf{T}_{K_n})\Lambda_\phi(\mathbf{T}_{K_n})$ aids MLE/MAP for α, ϕ

We have explicit formulae for MLE/MAP estimates for $\boldsymbol{\Psi}$

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

[4] Douglas N Hoover. Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2, 1979.

[5] François Caron and Emily B Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366, 2017.

[6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

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

[8] Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable