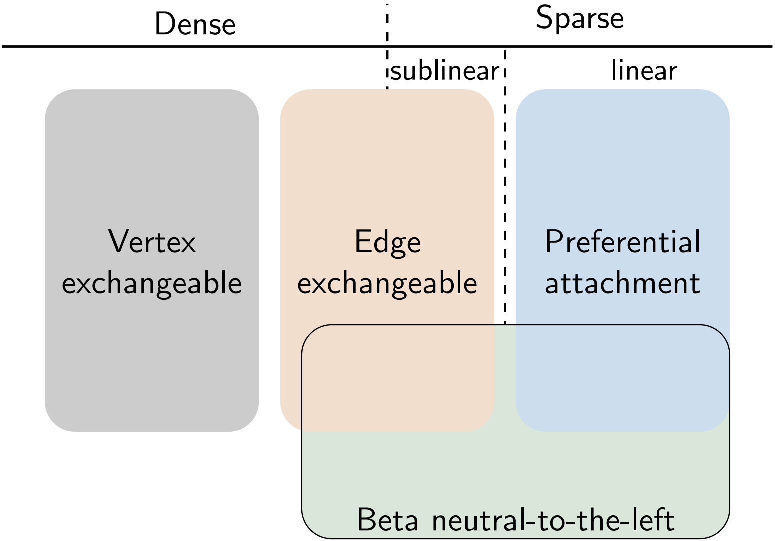
Sampling and Inference for Beta Neutral-to-the-Left Models of Sparse Networks

Benjamin Bloem-Reddy, Adam Foster, Emile Mathieu, Yee Whye Teh

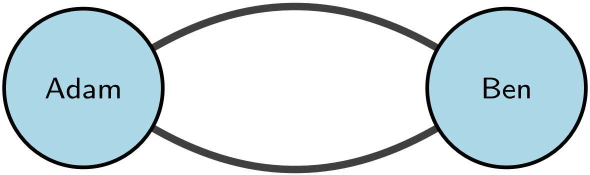
Department of Statistics, University of Oxford

Models for networks

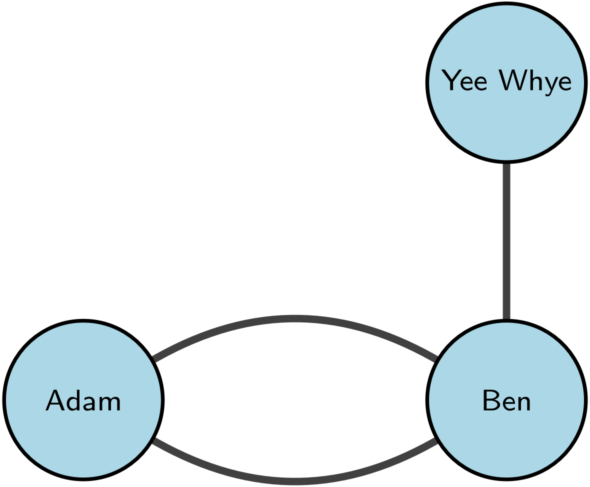


Temporal networks

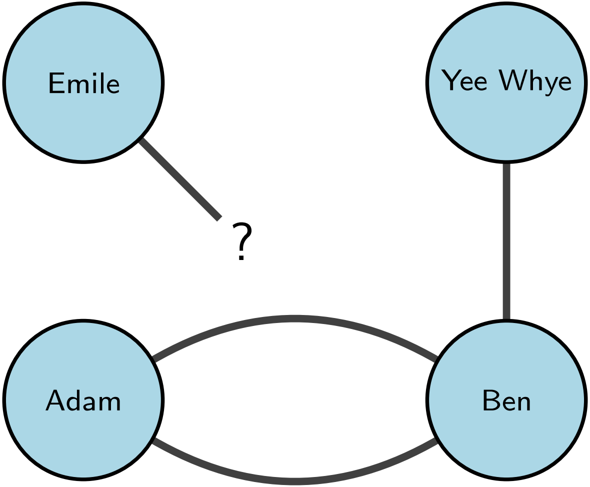
Temporal networks



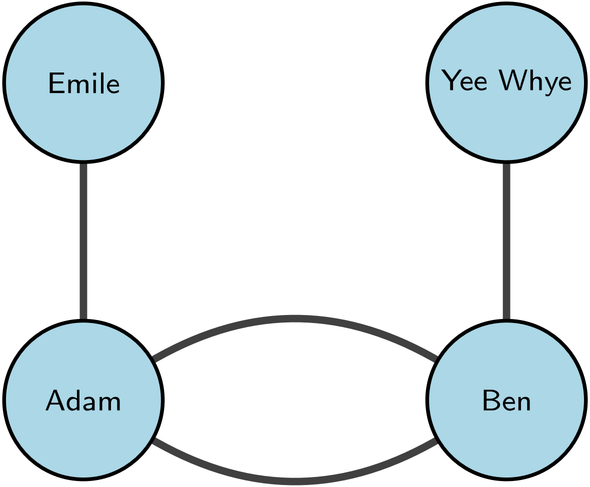
Temporal networks



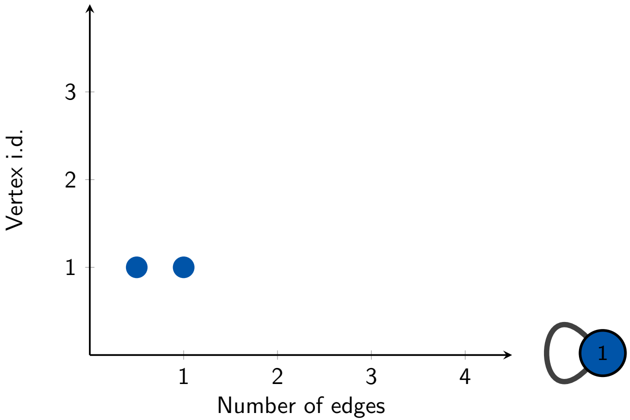
Temporal networks



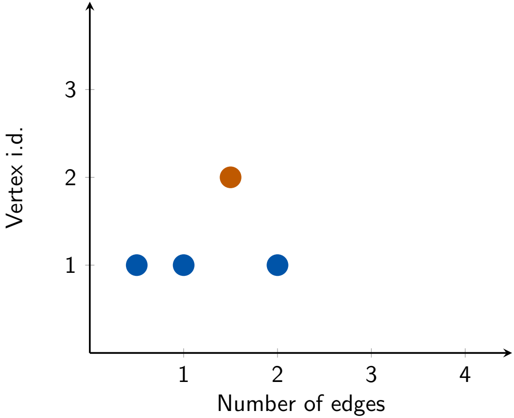
Temporal networks



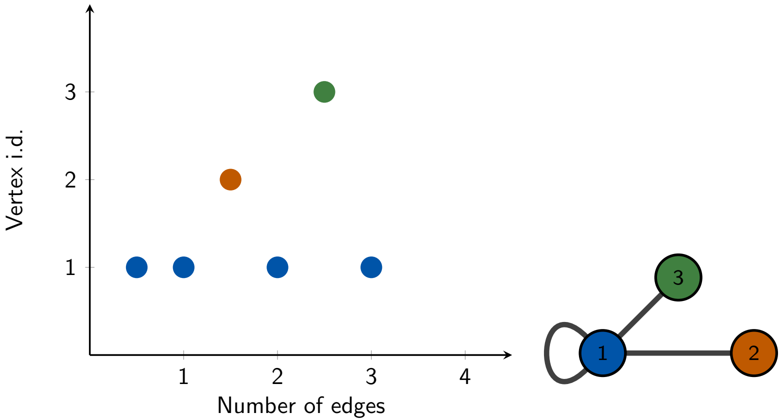
Edges and vertices



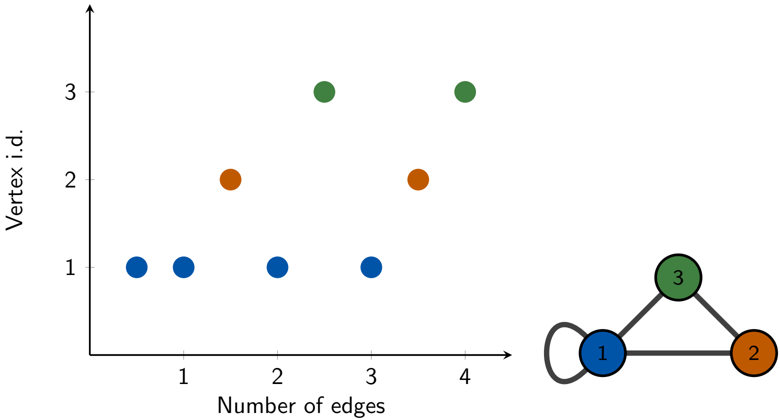
Edges and vertices



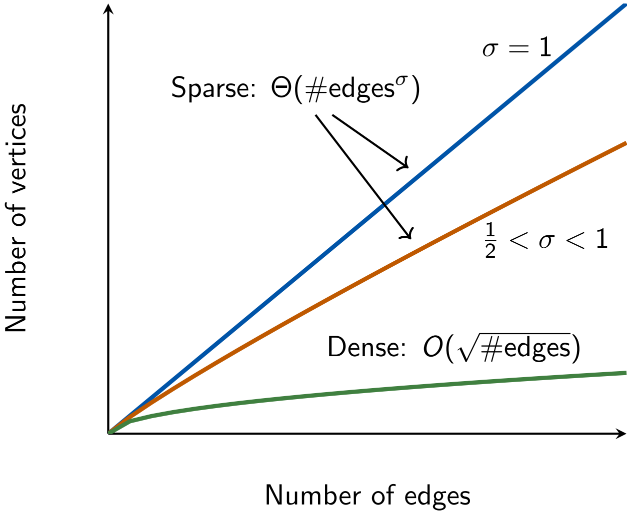
Edges and vertices



Edges and vertices

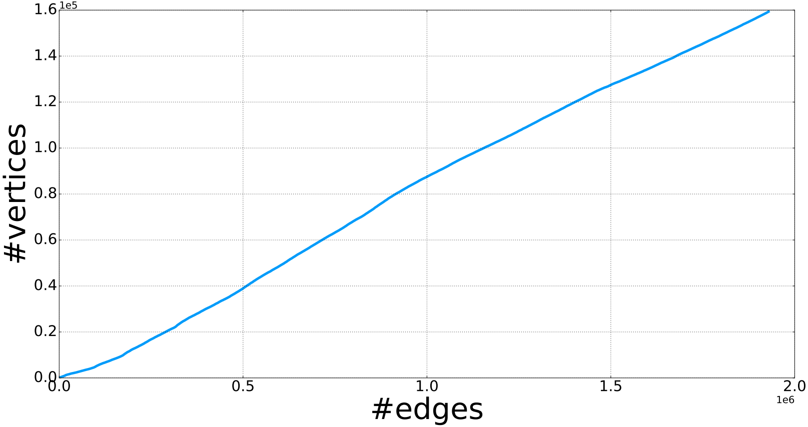


Sparsity

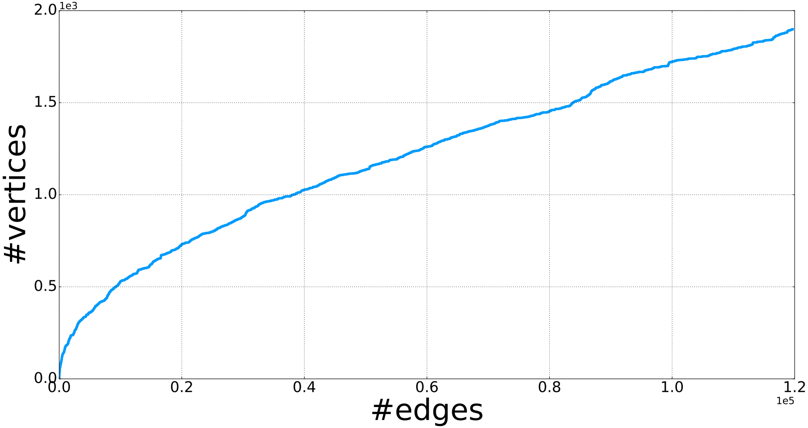


$\sigma = 1$
Sparse: $\Theta(\#\text{edges}^{\sigma})$
$\frac{1}{2} < \sigma < 1$
Dense: $O(\sqrt{\#\text{edges}})$
Number of vertices
Number of edges

Empirical study: Ask Ubuntu [1]



Empirical study: UCI social network [1]



Exchangeable models

A sequence of random variables $X_1, X_2, \ldots$ is *exchangeable* if for a finite permutation σ,

$$X_{\sigma(1)}, X_{\sigma(2)}, \ldots$$

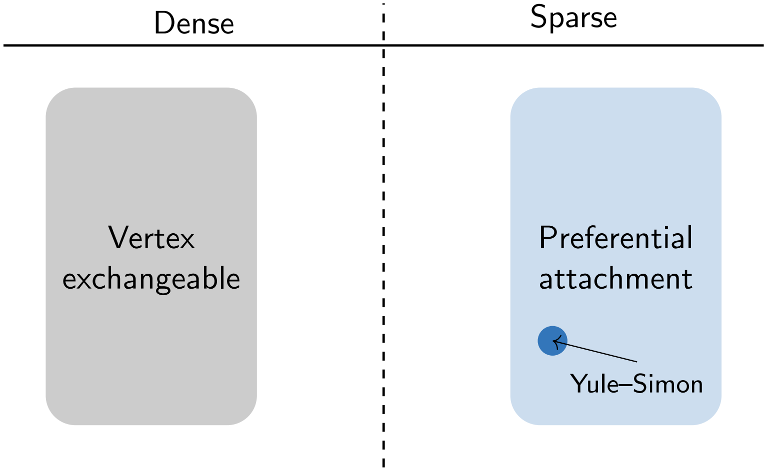
has the same distribution as the original sequence.

Exchangeable models tend to lead to tractable inference.

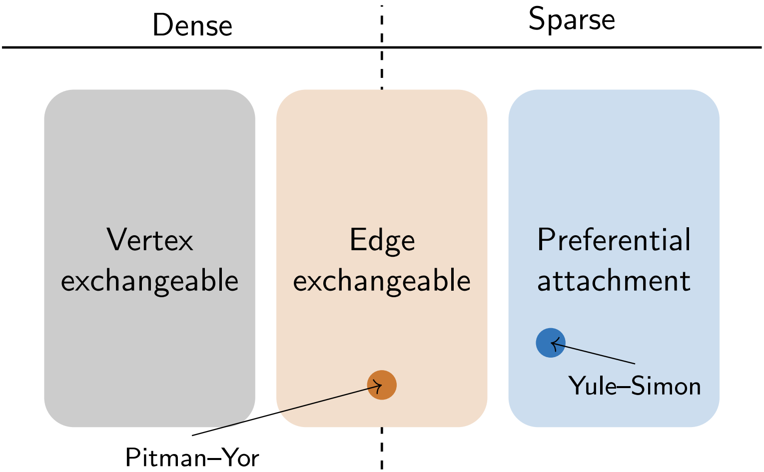
Models

Dense	Sparse
Vertex exchangeable	

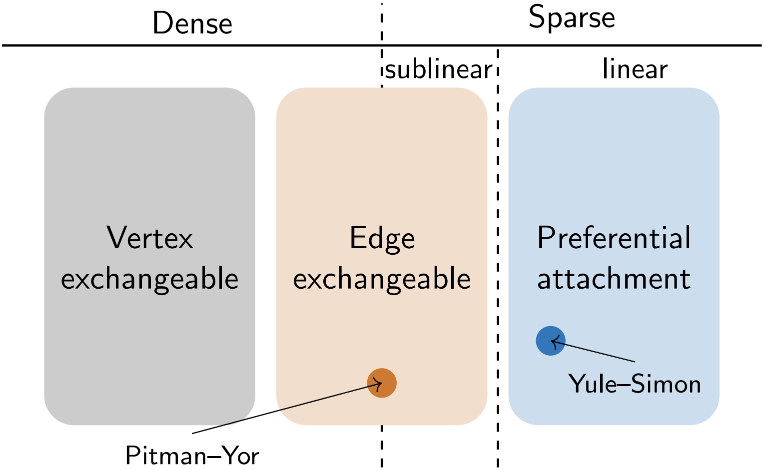
Models


Dense
Sparse
Vertex exchangeable
Preferential attachment
Yule–Simon

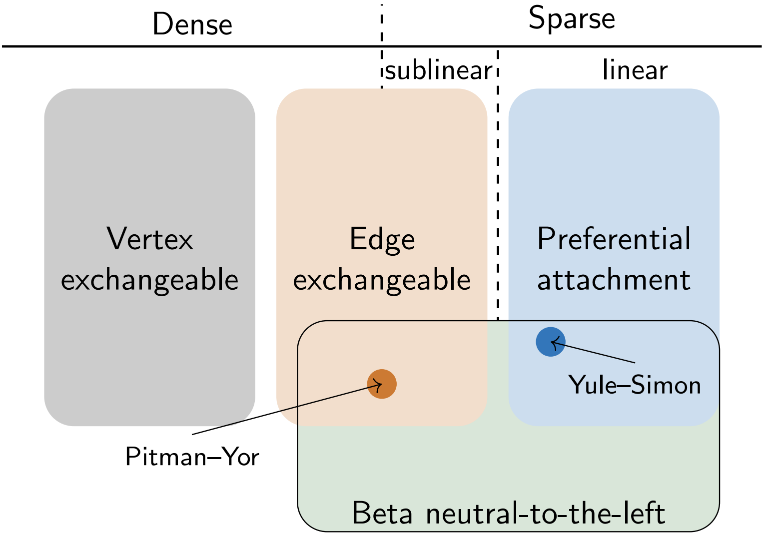
Models


Dense
Sparse
Vertex exchangeable
Edge exchangeable
Preferential attachment
Yule–Simon
Pitman–Yor

Models

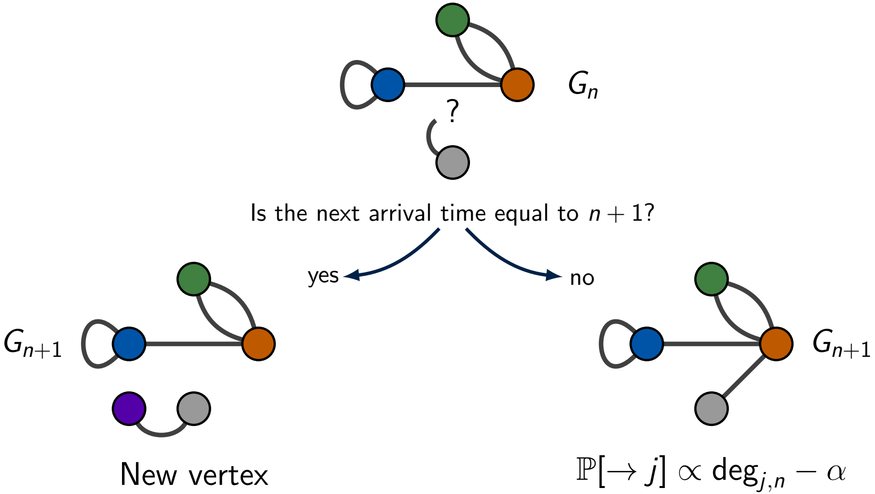

Dense
Sparse
sublinear
linear
Vertex exchangeable
Edge exchangeable
Preferential attachment
Yule–Simon
Pitman–Yor

Models


Dense
Sparse
sublinear
linear
Vertex exchangeable
Edge exchangeable
Preferential attachment
Yule–Simon
Pitman–Yor
Beta neutral-to-the-left

Beta Neutral-to-the-left Model [2]

1. Generate arrival times $1 = T_1 < T_2 < T_3 < \dots$ in any way.
2. Conditioned on arrival times, generate ends of edges sequentially:



Sampling and inference

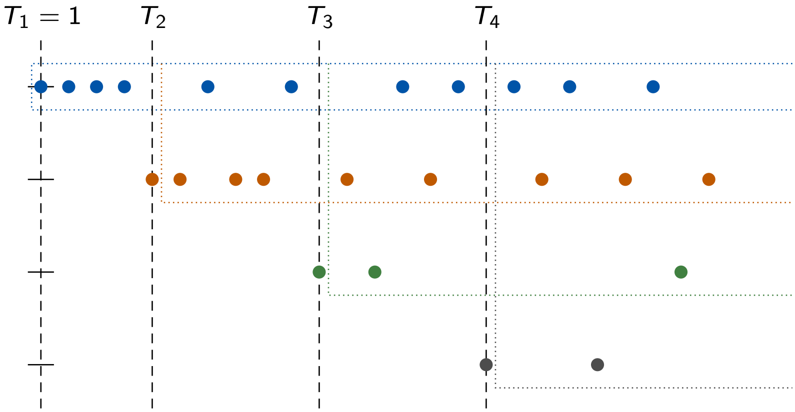
Why a paper on sampling and inference for BNTL models?

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* for tractable inference

Exchangeable substructure



Neutrality

At each step $n+1$ not corresponding to an arrival time, sample end of edge from discrete distribution

$$P_{n+1} = \left(\frac{d_{1,n} - \alpha}{n - K\alpha}, \ldots, \frac{d_{K,n} - \alpha}{n - K\alpha} \right) .$$

Equivalently (see paper and [2]), can sample from

$$\tilde{P}_K = (\tilde{P}_{1,K}, \ldots, \tilde{P}_{K,K})$$

where $$\frac{\tilde{P}_{j,K}}{\sum_{i=1}^{j} \tilde{P}_{i,K}} = \Psi_j \sim \mathrm{Beta}(1-\alpha, T_j - 1 - (j-1)\alpha) .$$

$\Psi_i \perp\!\!\!\perp \Psi_j$ for all $i \neq j$. These independent increments are called *neutral-to-the-left* and yield a factorized joint probability.

Gibbs structure

The joint probability has **Gibbs structure**

$$P(\text{graph}|T_1, T_2, ...) = \prod_{j \in \text{vertices}} P(\text{choose } j \ d_j - 1 \text{ times out of } T - T_j \text{ trials })$$

- d_j = degree of vertex j
- T = final time
- T_j = arrival time of vertex j

Gibbs sampler

Variable	**Gibbs sampling scheme**
Model variables	Analytic using Gibbs structure
Arrival times	Update each arrival time separately
Arrival order	Initialise in descending degree order use M-H with adjacent swap proposal

Experiments

- Gibbs: parameter recovery
- Gibbs: scalability
- Point estimation with massive graphs

Parameter recovery

- Synthetic data

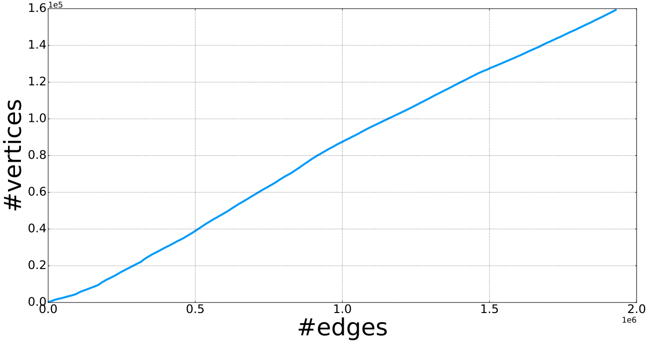
Gen. arrival distn.	Inference model	$\lvert\hat{\alpha} - \alpha^*\rvert$	Pred. log-lik.
$\mathcal{PYP}(1.0, 0.75)$	$(\tau, \mathcal{PYP}(\theta, \tau))$	**0.046** $\pm$ **0.002**	-**2637.0** $\pm$ **0.1**
$\mathcal{PYP}(1.0, 0.75)$	$(\alpha, \text{Geom}(\beta))$	0.049 $\pm$ 0.004	-2660.5 $\pm$ 0.7
Geom(0.25)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

Scalability

- Runtime linear in #edges
- Most expensive Gibbs update is for arrival times

MLEs for real data

Ask Ubuntu



- Pitman–Yor (edge exchangeable) misspecified
- Non-exchangeable BNTL provide better fit

Conclusion

- BNTL models are *flexible*
- Inference was challenging due to non-exchangeability

Conclusion

- BNTL models are *flexible*
- Inference was challenging due to non-exchangeability
- BNTL models are *tractable* due to exchangeable substructure

Thank you

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

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