Supplementary Material: Bayesian inference on random simple graphs with power law degree distributions

Juho Lee [1] **Creighton Heaukulani** [2] **Zoubin Ghahramani** [2 3] **Lancelot F. James** [4] **Seungjin Choi** [1]

1. Proofs

We prove Theorem 3.1 and Theorem 5.1 in the paper. First consider the following redefinition of our model with slightly different notation; let W_n be a random variable constrained on $(0, C_n]$, with density

$$f_n(\mathrm{d}w) = \frac{1}{Z_n} w^{-\alpha-1}(1-e^{-w})\mathbb{1}_{\{0<w\leq C_n\}}\mathrm{d}w, \tag{1}$$

where $C_1, C_2, \dots,$ is a sequence of positive numbers satisfying

$$\lim_{n\to\infty} C_n = \infty, \qquad \lim_{n\to\infty} C_n^\alpha/n = 0. \tag{2}$$

Note that $Z_n \to \Gamma(1-\alpha)/\alpha$ as $n\to\infty$, and so the sequence of densities $f_n(dw)$ converges pointwise to the density of the BFRY distribution

$$f(w) = \frac{\alpha}{\Gamma(1-\alpha)} w^{-\alpha-1}(1-e^{-w})\mathbb{1}_{\{w>0\}}, \tag{3}$$

and W_n converges in distribution to a BFRY random variable. Let $W_{n,1},\dots,W_{n,n}$ be n i.i.d. copies of W_n. A random simple graph X is then defined to be a collection of Bernoulli random variables as follows:

$$\mathbb{P}\{X_{ij}=1 \mid r_{i,j}\} = \frac{r_{i,j}}{1+r_{i,j}}, \quad r_{i,j} = U_iU_j, \quad U_i = \frac{W_{n,i}}{\sqrt{L_n}}, \tag{4}$$

where $L_n := \sum_{i=1}^n W_{n,i}$. We will write $X \mid r \sim \mathrm{GRG}(n, r)$, where $r := (r_{i,j} : i<j\leq n)$.

We begin with a sequence of Lemmas. Define a sequence of random variables $V_{s,n}$, for every $s, n \geq 1$, by

$$V_{s,n} := \frac{W_n}{C_n^{s-\alpha}}. \tag{5}$$

Let $V_{s,n,1},\dots,V_{s,n,n}$ be n i.i.d. copies of $V_{s,n}$, and denote the empirical mean of these copies by

$$\bar{V}_{s,n} := \frac{1}{n}\sum_{i=1}^n V_{s,n,i}. \tag{6}$$

The expectation of $V_{s,n}$ is finite for all $s, n < \infty$, and is computed as

$$\begin{aligned}\mathbb{E}[V_{s,n}] &= \frac{1}{Z_n C_n^{s-\alpha}} \int_0^{C_n} w^{s-\alpha-1}(1-e^{-w})\mathrm{d}w \\ &= \frac{1}{Z_n}\left\{\frac{1}{s-\alpha} - \frac{\gamma(s-\alpha, C_n)}{C_n^{s-\alpha}}\right\},\end{aligned} \tag{7}$$

[1]Pohang University of Science and Technology (POSTECH), Pohang, South Korea [2]University of Cambridge, Cambridge, United Kingdom [3]Uber AI Labs, San Francisco, CA, USA [4]Hong Kong University of Science and Technology (HKUST), Clearwater Bay, Hong Kong. Correspondence to: Juho Lee <stonecold@postech.ac.kr>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, 2017. JMLR: W&CP.

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function.

Let $\xrightarrow{\mathbb{P}}$ denote convergence in probability. The following lemma is a standard mean convergence result:

Lemma 1.1. *$\bar{V}_{s,n} \xrightarrow{\mathbb{P}} \mathbb{E}[V_{s,n}]$, as $n \to \infty$.*

Proof. For all $\varepsilon > 0$, by Chebyshev's inequality and the condition in Eq. (2),

$$\mathbb{P}\{|\bar{V}_{s,n} - \mathbb{E}[V_{s,n}]| \geq \varepsilon\} \leq \frac{\mathrm{Var}(V_{s,n})}{n\varepsilon^2} \leq \frac{\mathbb{E}[V_{s,n}^2]}{n\varepsilon^2} = \frac{1}{Z_n\varepsilon^2}\left\{\frac{C_n^\alpha}{n(2s-\alpha)} - \frac{\gamma(2s-\alpha, C_n)}{nC_n^{2s-2\alpha}}\right\} \to 0, \tag{8}$$

as $n \to \infty$, as desired. □

The following lemma will be used to study various higher order moments in later results:

Lemma 1.2. *For $s \geq 2$,*

$$M_{s,n} := \frac{\sum_{i=1}^n W_{n,i}^s}{(\sum_{i=1}^n W_{n,i})^s} \xrightarrow{\mathbb{P}} 0, \qquad \textit{as } n \to \infty. \tag{9}$$

Proof. We have

$$M_{s,n} = \frac{nC_n^{s-\alpha}\bar{V}_{s,n}}{n^s C_n^{s-s\alpha}\bar{V}_{1,n}^s} = \left(\frac{C_n^\alpha}{n}\right)^{s-1}\frac{\bar{V}_{s,n}}{\bar{V}_{1,n}^s}. \tag{10}$$

As $n \to \infty$, the first factor on the right hand side clearly converges to zero (c.f. Eq. (2)), and, by Lemma 1.1, the second term converges to a constant in probability. □

Recall that $D_{n,i} := \sum_{j \neq i} X_{i,j}$ is the degree of the i-th node in the graph $X \mid r \sim \mathrm{GRG}(n, r)$, given by Eq. (4). The following result will show up in later calculations involving the probability generating function (PGF) of the degree random variables $D_{n,i}$:

Lemma 1.3. *For every collection $t_1, \ldots, t_n$ with $|t_i| \leq 1$, for $i \leq n$,*

$$\mathbb{E}\Big[\prod_{i=1}^n t_i^{D_{n,i}} \mid W_{n,1} = w_1, \ldots, W_{n,n} = w_n\Big] = \prod_{i<j\leq n} \frac{L_n + t_i t_j w_i w_j}{L_n + w_i w_j}, \tag{11}$$

for positive $w_1, \ldots, w_n$.

Proof. The proof is given by Britton et al. (2006). □

The following result studies a representation of the PGF of the degree random variables and their higher order moments:

Lemma 1.4. *Fix a node $k \leq n$. Define*

$$F_{n,k}(t; w_k) := \prod_{i \neq k} \frac{L_{n,-k} + w_k + tw_k W_{n,i}}{L_{n,-k} + w_k + w_k W_{n,i}}, \qquad \textit{for } |t| \leq 1, \textit{ and } w_k > 0, \tag{12}$$

where $L_{n,-k} := \sum_{i \neq k} W_{n,i}$. Note that the s-th derivative $F_{n,k}^{(s)}(t; w_k)$ exists for all $s \geq 0$. For all $s \geq 0$, the following hold:

1. *$F_{n,k}^{(s)}(t; w_k)$ is uniformly bounded, for all $n \geq 1$;*
2. *$F_{n,k}^{(s)}(t; w_k) \xrightarrow{\mathbb{P}} w_k^s \exp\{(t-1)w_k\}$, as $n \to \infty$.*

Proof. In the case $s = 0$, $F_{n,k}(t; w_k)$ is trivially bounded by 1 since $|t| \leq 1$. By the Taylor series expansion $\log(1 + x) = x + O(x^2)$, we have

$$F_{n,k}(t; w_k) = \exp\left\{(t-1)w_k \frac{L_{n,-k}}{L_{n,-k} + w_k} + O\left(w_k^2 \frac{\sum_{i \neq k} W_{n,i}^2}{(L_{n,-k} + w_k)^2}\right)\right\}. \tag{13}$$

By Lemma 1.1,

$$\frac{L_{n,-k}}{L_{n,-k} + w_k} = \frac{\bar{V}_{1,n,-k}}{\bar{V}_{1,n,-k} + w_k/(n-1)/C_n^{1-\alpha}} \xrightarrow{\mathbb{P}} 1, \tag{14}$$

where $\bar{V}_{1,n,-k}$ is the empirical mean in Eq. (6) excluding the element $V_{1,n,k}$. Furthermore, by Lemma 1.2,

$$O\left(w_k^2 \frac{\sum_{i \neq k} W_{n,i}^2}{(L_{n,-k} + w_k)^2}\right) \leq O(w_k^2 M_{2,n,-k}) \xrightarrow{\mathbb{P}} 0, \tag{15}$$

where $M_{s,n,-k}$ is $M_{s,n}$ computed without $V_{s,n,k}$. Combining, we have

$$F_{n,k}(t; w_k) \xrightarrow{\mathbb{P}} \exp\{(t-1)w_k\}. \tag{16}$$

Before proceeding for $s \geq 1$, we define

$$Q_{r,n,k}(t; w_k) := \sum_{i \neq k} \frac{W_{n,i}^r}{(L_{n,-k} + w_k + t w_k W_{n,i})^r}, \tag{17}$$

for all $r, n \geq 1$. One can easily see that $Q_{r,n,k}(t; w_k) \leq 1$ for all $r, n \geq 1$. For $r = 1$, we have

$$\sum_{i \neq k} \frac{W_{n,i}}{L_{n,-k} + w_k + t w_k C_n} \leq Q_{1,n,k}(t; w_k) \leq 1, \tag{18}$$

and

$$\begin{aligned}
\sum_{i \neq k} \frac{W_{n,i}}{L_{n,-k} + w_k + t w_k C_n} &= \frac{1}{1 + w_k/L_{n,-k} + t w_k C_n/L_{n,-k}} \\
&= \left\{1 + \frac{w_k}{(n-1)C_n^{1-\alpha}} \bar{V}_{s,n,-k}^{-1} + t w_k \frac{C_n^\alpha}{n} \frac{n}{n-1} \bar{V}_{s,n,-k}^{-1}\right\}^{-1} \xrightarrow{\mathbb{P}} 1.
\end{aligned} \tag{19}$$

Hence, by the squeeze theorem, $Q_{1,n,k}(t; w_k) \xrightarrow{\mathbb{P}} 1$. For $r \geq 2$, we have

$$0 \leq Q_{r,n,k}(t; w_k) \leq M_{r,n,-k} \xrightarrow{\mathbb{P}} 0, \tag{20}$$

by Lemma 1.2. Hence, we have $Q_{r,n,k}(t; w_k) \xrightarrow{\mathbb{P}} 0$ for $r \geq 2$.

Now we show that

$$F_{n,k}^{(s)}(t; w_k) = w_k F_{n,k}^{(s-1)}(t; w_k) Q_{1,n,k}(t; w_k) + \sum_{r=2}^{s} a_{s,r} F_{n,k}^{(s-r)}(t; w_k) Q_{r,n,k}(t; w_k), \tag{21}$$

for some constants $\{a_{s,r}\}$ for all $s \geq 1$ and $r \geq 2$. We proceed by the mathematical induction. For $s = 1$,

$$\begin{aligned}
F_{n,k}^{(1)}(t; w_k) &= \sum_{i \neq k} \frac{w_k W_{n,i}}{L_{n,-k} + w_k + w_k W_{n,i}} \prod_{j \neq i,k} \frac{L_{n,-k} + w_k + t w_k W_{n,j}}{L_{n,-k} + w_k + w_k W_{n,j}} \\
&= w_k F_{n,k}(t; w_k) Q_{1,n,k}(t; w_k).
\end{aligned} \tag{22}$$

Now by the inductive hypothesis,

$$\begin{aligned}F^{(s+1)}_{n,k}(t;w_k) &= w_k F^{(s)}_{n,k}(t;w_k)Q_{1,n,k}(t;w_k) - w_k^2 F^{(s-1)}_{n,k}(t;w_k)Q_{2,n,k}(t;w_k) \\ &\quad + \sum_{r=2}^{s} a_{s,r}(F^{(s+1-r)}_{n,k}(t;w_k)Q_{r,n,k}(t;w_k) - r w_k F^{(s-r)}_{n,k}Q_{r+1,n,k}(t;w_k)) \\ &= w_k F^{(s)}_{n,k}(t;w_k)Q_{1,n,k}(t;w_k) + \sum_{r=2}^{s+1} a_{s+1,r}F^{(s+1-r)}_{n,k}(t;w_k)Q_{r,n,k}(t;w_k),\end{aligned} \tag{23}$$

where

$$a_{s+1,2} = a_{s,2} - w_k^2, \quad a_{s+1,r} = a_{s,r} - a_{s,r-1}(r-1)w_k \quad \text{for } r \geq 2, \tag{24}$$

so the inductive argument holds.

Having (21), by mathematical induction, we can easily show that $F^{(s)}_{n,k}(t;w_k)$ is uniformly bounded for all $s, n \geq 1$. Moreover,

$$F^{(1)}_{n,k}(t;w_k) = w_k F_{n,k}(t;w_k)Q_{1,n,k}(t;w_k) \xrightarrow{\mathbb{P}} w_k \exp\{(t-1)w_k\}, \tag{25}$$

by (16) and (19). Combining this with (20), by mathematical induction, we can show that for all $s \geq 1$,

$$F^{(s)}_{n,k}(t;w_k) \xrightarrow{\mathbb{P}} w_k^s \exp\{(t-1)w_k\}. \tag{26}$$

□

We will now use our collected results to analyze the asymptotic distribution of the degree random variables; the following result characterizes this distribution:

Lemma 1.5. *Fix a node k. Given $\{W_{n,k} = w_k\}$, for some $w_k > 0$, the degree $D_{n,k}$ of node k converges in distribution to a Poisson random variable with rate w_k, as $n \to \infty$.*

Proof. The PGF of $D_{n,k}$ is given by

$$\mathbb{E}[t^{D_{n,k}} \mid W_{n,k} = w_k] = \mathbb{E}[F_{n,k}(t;w_k)], \qquad \text{for } |t| \leq 1. \tag{27}$$

Note that these expectations are under the σ-field generated by $\{W_k = w_k\}$. For all $s \geq 0$, we will derive the limit of $\mathbb{P}\{D_{n,k} = s \mid w_k\}$, as $n \to \infty$, which we note is given by the s-th order derivatives of the PGF in Eq. (27), evaluated at the argument $t = 0$. It therefore suffices to show that $\mathbb{E}[F^{(s)}_{n,k}(t;w_k)] \to w_k^s \exp\{(t-1)w_k\}$, as $n \to \infty$, for all $s \geq 0$. By Lemma 1.4, we know that $F^{(s)}_{n,k}(t;w_k)$ is uniformly bounded and that $F^{(s)}_{n,k}(t;w_k) \xrightarrow{\mathbb{P}} w_k^s \exp\{(t-1)w_k\}$, as $n \to \infty$. Therefore, by uniform integrability,

$$\lim_{n\to\infty} \mathbb{E}[F^{(s)}_{n,k}(t;w_k)] = \mathbb{E}\Big[\lim_{n\to\infty} F^{(s)}_{n,k}(t;w_k)\Big] = w_k^s \exp\{(t-1)w_k\}. \tag{28}$$

□

We are now ready to prove the main theorems in the paper.

Proof of Theorem 3.1. We will first verify that, for $y \gg 1$, $\mathbb{P}\{D_{n,k} = y\} \to cy^{-1-\alpha}$ for every node k and for some constant $c > 0$ as $n \to \infty$. By Lemma 1.5, conditioned on $\{W_k = w_k\}$, the degree $D_{n,k}$ converges in distribution to a Poisson random variable with rate w_k. Then by dominated convergence,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\{D_{n,k} = y\} &= \lim_{n\to\infty} \int_0^\infty \mathbb{P}\{D_k = y | w_k\} p_n(\mathrm{d}w_k) \\ &= \int_0^\infty \frac{w_k^y e^{-w_k}}{y!} p(\mathrm{d}w_k) \\ &= \frac{\alpha\Gamma(y-\alpha)}{y!\Gamma(1-\alpha)}(1 - 2^{\alpha-y}).\end{aligned} \tag{29}$$

By the asymptotics of the Gamma function, for $y \gg 1$, we have

$$\lim_{n\to\infty} \mathbb{P}\{D_{n,k} = y\} = cy^{-1-\alpha}, \tag{30}$$

for some constant c.

Next we show that, for any finite m, the collection of random variables $D_{n,1}, \dots, D_{n,m}$ are asymptotically independent, as $n \to \infty$. We compute the (joint) probability generating function of $(D_{n,1}, \dots, D_{n,m})$, with $|t_i| \le 1$ for $i = 1, \dots, m$. By Lemma 1.3,

$$\begin{aligned}
\mathbb{E}\Big[\prod_{i=1}^{m} t_i^{D_{n,i}}\Big] &= \mathbb{E}\Big[\prod_{i=1}^{m}\prod_{j=i+1}^{m} \frac{L_n + t_i t_j W_{n,i}W_{n,j}}{L_n + W_{n,i}W_{n,j}} \prod_{j=m+1}^{n} \frac{L_n + t_i W_{n,i}W_{n,j}}{L_n + W_{n,i}W_{n,j}}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i=1}^{m}\prod_{j=i+1}^{m} \frac{L_{n,m+1:n} + \ell_{n,1:m} + t_i t_j w_i W_{n,j}}{L_{n,m+1:n} + \ell_{n,1:m} + w_i W_{n,j}} \\
&\qquad \times \prod_{j=m+1}^{n} \frac{L_{n,m+1:n} + \ell_{n,1:m} + t_i w_i W_{n,j}}{L_{n,m+1:n} + \ell_{n,1:m} + w_i W_{n,j}} \mid W_{n,1:m} = w_{1:m}\Big]\Big].
\end{aligned} \tag{31}$$

Given $w_{1:m}$, by a similar argument as in the proof of Lemma 1.4, one can easily show that

$$\prod_{j=i+1}^{m} \frac{L_{n,m+1:n} + \ell_{n,1:m} + t_i t_j w_i W_{n,j}}{L_{n,m+1:n} + \ell_{n,1:m} + w_i W_{n,j}} \xrightarrow{\mathbb{P}} 1, \quad \text{as } n \to \infty, \tag{32}$$

and

$$\prod_{j=m+1}^{n} \frac{L_{n,m+1:n} + \ell_{n,1:m} + t_i w_i W_{n,j}}{L_{n,m+1:n} + \ell_{n,1:m} + w_i W_{n,j}} \xrightarrow{\mathbb{P}} \exp\{(t_i - 1)w_i\}, \quad \text{as } n \to \infty. \tag{33}$$

Hence, again by a similar argument as in the proof of Lemma 1.4, we have

$$\lim_{n\to\infty} \mathbb{E}\Big[\prod_{i=1}^{m} t_i^{D_{n,i}}\Big] = \prod_{i=1}^{m} \mathbb{E}[\exp\{(t_i - 1)W_i\}], \tag{34}$$

that is, the joint PGF asymptotically factorizes into the product of the PGFs for i.i.d. random variables, and the result follows. □

Proof of Theorem 3.2. Using the fact that the expected number of nodes $E_n := \sum_{i=1}^{n} D_{n,i}/2$, we may take $t_1 = \dots = t_n = \sqrt{t}$ and obtain

$$\mathbb{E}[t^{E_n}] = \mathbb{E}\Big[\prod_{i<j\le n} \frac{L_n + tW_{n,i}W_{n,j}}{L_n + W_{n,i}W_{n,j}}\Big]. \tag{35}$$

We evaluate the derivative of the PGF to obtain the first moment

$$\mathbb{E}[E_n] = \frac{\partial \mathbb{E}[t^{E_n}]}{\partial t}\Big|_{t=1} = \mathbb{E}\Big[\sum_{i<j\le n} \frac{W_{n,i}W_{n,j}}{L_n + W_{n,i}W_{n,j}}\Big] \le \frac{1}{2}\mathbb{E}\Big[\sum_{i\le j\le n} \frac{W_{n,i}W_{n,j}}{L_n}\Big] = \frac{n}{2}\mathbb{E}[W_n]. \tag{36}$$

Since

$$\mathbb{E}[W_n] = \frac{1}{Z_n}\Big\{\frac{C_n^{1-\alpha}}{1-\alpha} - \gamma(1-\alpha, C_n)\Big\}, \tag{37}$$

we have

$$\mathbb{E}[E_n] = O(nC_n^{1-\alpha}). \tag{38}$$

□

Proof of Theorem 5.1. Recall that

$$\mathbb{P}\{X = x \mid r\} = \prod_{i<j\leq n} \frac{r_{i,j}}{1 + r_{i,j}} = G^{-1}(r) \prod_{i<j\leq n} A_{i,j}^{x_{i,j}} \prod_{i=1}^{n} U_i^{D_{n,i}}, \tag{39}$$

where $A := (A_{i,j})_{i<j\leq n}$ and

$$G(r) := \prod_{i<j\leq n} (1 + A_{i,j} U_i U_j). \tag{40}$$

Since $\sum_x \mathbb{P}\{X = x \mid r\} = 1$, we have

$$G(r) = \sum_x \prod_{i<j\leq n} A_{i,j}^{x_{i,j}} \prod_{i=1}^{n} u_i^{D_{n,i}}. \tag{41}$$

The joint PGF of $(D_{n,1}, \ldots, D_{n,n})$ is then

$$\begin{aligned}
\mathbb{E}\Big[\prod_{i=1}^{n} t_i^{D_{n,i}} \mid A, W_{n,1:n}\Big] &= \sum_x \mathbb{P}\{X = x \mid r\} \prod_{i=1}^{n} t_i^{D_{n,i}(x)} \\
&= G^{-1}(r) \sum_x \prod_{i<j\leq n} A_{i,j}^{x_{i,j}} \prod_{i=1}^{n} (t_i U_i)^{D_{n,i}} \\
&= \prod_{i<j\leq n} \frac{1 + A_{i,j} t_i t_j U_i U_j}{1 + A_{i,j} U_i U_j}.
\end{aligned} \tag{42}$$

The remainder of the proof follows analogously to the proof of Theorem 3.1 above. □

References

Britton, T., Deijfen, M., and Martin-Löf, A. Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.

Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. In *ICLR*, 2014.

Knowles, D. A. Stochastic gradient variational Bayes for gamma approximating distributions. *arXiv preprint arXiv:1509.01631*, 2015.

Salimans, T. and Knowles, D. A. Fixed-form variational posterior approximation through stochastic linear regression. *Bayesian Analysis*, 8(4):837–882, 2013.